Mainland Canadian English in Newfoundland

The Canadian Shift in Urban Middle-Class St. John’s

Dissertation

zur Erlangung des akademischen Grades

doctor philosophiae (Dr. phil.)

vorgelegt der Philosophischen Fakultät der

Technischen Universität Chemnitz

von Herrn Matthias Hofmann, geboren am 23. 07. 1980 in Schlema

Chemnitz, den 05. 11. 2014
3 Newfoundland and its Englishes

3.1 The Settlement History and Recent Migration Patterns .................. 79
3.2 The Sociolinguistic Situation .................................................. 85
   3.2.1 The Low-back Merger ................................................... 88
   3.2.2 TRAP and BATH ......................................................... 88
   3.2.3 DRESS and Kit Raising ................................................ 90
   3.2.4 DRESS and Kit Lowering ............................................. 90
   3.2.5 FOOT and STRUT ...................................................... 91
   3.2.6 The St. John’s Speech Community .................................... 92
3.3 Recent Developments ............................................................ 95
   3.3.1 Language Attitudes and Linguistic Identity ......................... 96
   3.3.2 Linguistic Change ...................................................... 100
3.4 Deriving the Research Questions ............................................ 102

4 Data and Methodology ............................................................... 107

4.1 Comparability with Earlier Studies ......................................... 111
4.2 Study Design ........................................................................... 116
   4.2.1 The Sociolinguistic Interview .......................................... 120
      4.2.1.1 The Word List ...................................................... 121
      4.2.1.2 The Reading Passage ............................................ 126
      4.2.1.3 The Interview ..................................................... 127
   4.2.2 Sampling and Data Collection ......................................... 133
      4.2.2.1 Socioeconomic Class Membership .............................. 137
      4.2.2.2 Local-ness to St. John’s ....................................... 140
   4.3 Acoustic Theory .................................................................... 143
      4.3.1 The Source-Filter Theory ............................................ 143
      4.3.2 Tube Models ............................................................. 147
      4.3.3 Linear Predictive Coding (LPC) ..................................... 155
   4.4 Measurements ..................................................................... 167
      4.4.1 Single-point versus Multiple-point Measurements ............... 168
      4.4.2 Quantifying the Maximal Displacement Approach ............... 172
      4.4.3 Phrasal Accent ............................................................ 177
         4.4.3.1 Pitch ................................................................. 177
         4.4.3.2 Vowel Duration .................................................. 181
      4.4.4 Tracking Formants Automatically .................................... 182
         4.4.4.1 Spectrogram Settings ........................................... 184
         4.4.4.2 Formant Tracker Settings .................................... 186
      4.4.5 Linguistic Context ........................................................ 192
      4.4.6 Tokens for Analysis ..................................................... 195
4.5 Data Preparation .................................................. 196
  4.5.1 Data Cleaning .............................................. 196
  4.5.2 Normalization ............................................... 199
4.6 Data Analysis and Statistical Modeling ........................ 208
  4.6.1 Assumptions of Data Exploration Methods .................. 209
  4.6.2 Assumptions of Statistical Tests ........................... 210
    4.6.2.1 Normality and Outliers ................................ 211
    4.6.2.2 Homoscedasticity ...................................... 213
    4.6.2.3 Independence of Observations ......................... 215
    4.6.2.4 Additional Assumptions of General Linear Models .... 216
    4.6.2.5 Assumptions of Generalized Linear Mixed Models ..... 218
  4.6.3 Selection of Statistical Tests ............................... 223
    4.6.3.1 Merged Vowels ......................................... 227
    4.6.3.2 Shifted Vowels ......................................... 231
    4.6.3.3 Triangulation via Generalized Linear Mixed-Modeling 233
    4.6.3.4 Fitting a Regression Model ............................ 239

5 Analysis and Discussion ........................................ 243
  5.1 First Results ................................................ 247
    5.1.1 The Stability of STRUT in Apparent Time ................ 249
    5.1.2 The Status of TRAP and BATH ............................ 253
  5.2 The Low-back Merger .......................................... 255
    5.2.1 The Merger within Age Groups ............................ 256
    5.2.2 The Merger across Age Groups ............................. 263
    5.2.3 Non-pruned Decision Tree for LOT ........................ 271
    5.2.4 Generalized Mixed-effects Modeling of the Merger ....... 273
      5.2.4.1 Linear Regression .................................... 274
      5.2.4.2 Logistic Regression .................................. 281
  5.3 TRAP Retraction .............................................. 284
    5.3.1 Retraction of TRAP across Age Groups .................... 285
      5.3.1.1 Following Phonological Environment .................. 285
      5.3.1.2 Euclidean Distances .................................. 287
      5.3.1.3 TRAP Retraction versus Lowering ..................... 290
    5.3.2 Generalized Mixed-effects Modeling of TRAP Retraction .. 292
      5.3.2.1 Linear Regression .................................... 293
      5.3.2.2 Logistic Regression .................................. 297
  5.4 The Shift of DRESS ........................................... 300
    5.4.1 DRESS Movement across Age Groups ...................... 302
      5.4.1.1 Correlation of Age and Formant Values ................ 304
      5.4.1.2 DRESS Retraction versus Lowering .................... 306
H.5 Ergebnisse .............................................................. 467
H.6 Schlussbetrachtung .................................................. 468

I Eidestattliche Erklärung zur Eigenständigkeit 469
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Phonetic realizations of the NLE lax vowels</td>
<td>87</td>
</tr>
<tr>
<td>4.1</td>
<td>Social matrix of age, sex and social class of individual participants in the planning stage of the fieldwork.</td>
<td>117</td>
</tr>
<tr>
<td>4.2</td>
<td>Stratification of middle-class participants according to age and sex</td>
<td>136</td>
</tr>
<tr>
<td>4.3</td>
<td>Birth year, sex, education and occupation of the individual respondents</td>
<td>139</td>
</tr>
<tr>
<td>4.4</td>
<td>Ligen, Lischool, LIabroad and LItotal of the individual respondents</td>
<td>142</td>
</tr>
<tr>
<td>4.5</td>
<td>Twelve points indicating the arbitrary amplitude of a sound wave</td>
<td>159</td>
</tr>
<tr>
<td>4.6</td>
<td>Comparison of formant measurements differences (Delta F1 and Delta F2) in Hertz of FOOT via LPC and FFT spectra</td>
<td>165</td>
</tr>
<tr>
<td>4.7</td>
<td>Summary of automatic vowel analysis procedures under comparison</td>
<td>174</td>
</tr>
<tr>
<td>4.8</td>
<td>Mean differences between manual and automatic formant measurements for five different measurement points</td>
<td>175</td>
</tr>
<tr>
<td>4.9</td>
<td>Raw formant and bandwidth values in Hertz for DRESS in veterinary</td>
<td>192</td>
</tr>
<tr>
<td>4.10</td>
<td>Generalized effects of places of articulation on formants</td>
<td>193</td>
</tr>
<tr>
<td>4.11</td>
<td>Illustration of variables and variants in the initial analysis</td>
<td>240</td>
</tr>
<tr>
<td>4.12</td>
<td>Cross-tabulation of place and manner of articulation for all KIT tokens</td>
<td>240</td>
</tr>
<tr>
<td>4.13</td>
<td>Cross-tabulation of following voice and crossed place and manner of articulation for all KIT tokens</td>
<td>241</td>
</tr>
<tr>
<td>4.14</td>
<td>Results for crossed place and manner of articulation with interactions</td>
<td>242</td>
</tr>
<tr>
<td>5.1</td>
<td>MANOVA Assumptions of STRUTs mean formant values</td>
<td>249</td>
</tr>
<tr>
<td>5.2</td>
<td>Overview of the t-test results for TRAP and BATH</td>
<td>254</td>
</tr>
<tr>
<td>5.3</td>
<td>T-test results for LOT and THOUGHT excluding pre-nasal tokens</td>
<td>259</td>
</tr>
<tr>
<td>5.4</td>
<td>ANOVA Assumptions for age groups and EDs between LOT and THOUGHT</td>
<td>266</td>
</tr>
<tr>
<td>5.5</td>
<td>Results of linear regression analysis of merged LOT-THOUGHT</td>
<td>276</td>
</tr>
<tr>
<td>5.6</td>
<td>Cross-tabulation of phoneme label and sex</td>
<td>277</td>
</tr>
<tr>
<td>5.7</td>
<td>Results from logistic regression analysis of merged LOT-THOUGHT in SPSS</td>
<td>282</td>
</tr>
</tbody>
</table>
5.8 Assumptions of an ANOVA regarding the distribution of age groups and EDs between TRAP and STRUT ................................................................. 288
5.9 Results of linear regression analysis of TRAP .................................................. 294
5.10 Results of logistic regression analysis of TRAP .............................................. 297
5.11 ANOVA Assumptions of age groups and EDs between DRESS and STRUT ...... 304
5.12 Results of linear regression analysis of DRESS ............................................. 310
5.13 Results of logistic regression analysis of DRESS .......................................... 314
5.14 ANOVA Assumptions of age groups and EDs between KIT and STRUT ........ 319
5.15 Results of linear regression analysis of KIT ................................................. 322
5.16 Results of linear regression analysis of KIT in monosyllabic lexical items .... 327
5.17 Results of logistic regression analysis of KIT .............................................. 329
5.18 Results of logistic regression analysis of KIT in monosyllabic lexical items .... 333
5.19 Formant means of the three lax vowels and merged LOT-THOUGHT per age 
   group ........................................................................................................ 337
5.20 Tabular comparison of the regression results for the front lax vowels ........ 339
5.21 Summary and juxtaposition of $R^2_{GLMM(m)}$ and other $R^2$ values of the front lax
   vowels ....................................................................................................... 343
List of Figures

2.1 Map of Canada including provincial and territorial boundaries . . . . . . . . . 34
2.2 Peripheral and non-peripheral tracks . . . . . . . . . . . . . . . . . . . . . . . 42
2.3 Typical directions of nuclei in chain shifts . . . . . . . . . . . . . . . . . . . . . 42
2.4 The Canadian Shift according to Clarke, Elms and Youssef (1995) . . . . . . . 48
2.5 The Canadian Shift according to Boberg (2005) . . . . . . . . . . . . . . . . . 53
2.6 The Canadian Shift according to Labov, Ash and Boberg (2006b) . . . . . . . 58
3.1 An old map of Newfoundland’s railway . . . . . . . . . . . . . . . . . . . . . . 82
4.1 Schematized power spectra of the source pulses with no filter in (a) and with
the filter function superimposed in (b) . . . . . . . . . . . . . . . . . . . . . . . 146
4.2 A schematic diagram of a simplified version of a neutral vocal tract in the
position for the vowel /ə/ as a tube closed at one end . . . . . . . . . . . . . . . . . 148
4.3 A diagram of the wavelength of a sound wave at an F0 of 200 Hz . . . . . . . 148
4.4 A graph of a sound wave with $\frac{3}{4}$ of its wavelength within the tube . . . . . 150
4.5 A schematized graph of a Helmholtz resonator for the vocal tract shape in the
vowel FLEECE . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 153
4.6 Two digitally stored waveforms . . . . . . . . . . . . . . . . . . . . . . . . . . 155
4.7 Schematic representation of an input, a filter and its output . . . . . . . . . . 157
4.8 A damped sinusoidal wave sampled at a number of points . . . . . . . . . . . 158
4.9 LPC spectrum superimposed on a narrowband FFT spectrum of FOOT . . . . 163
4.10 Extracted waveform of FOOT using a Hanning window (top) and a rectangular
window (bottom) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 165
4.11 Histogram of differences between automatic and manual formant measure-
ments using the Third method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 176
4.12 An expanded view of the waveform of something. . . . . . . . . . . . . . . . . 178
4.13 FFT to spectrogram I . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 183
4.14 FFT to spectrogram II . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 183
4.15 Spectrograms at two bandwidths compared . . . . . . . . . . . . . . . . . . . . . 185
4.16 Sound editor Praat 5.3 for veterinary of speaker 11CDMF at a setting of 5500 Hz and 5 formants ....................................................... 189
4.17 Sound editor Praat 5.3 for veterinary of speaker 11CDMF at a setting of 6600 Hz and 6 formants and 8500 Hz and 7 formants ........................................ 191
4.18 A complex waveform with additional noise ........................................ 197
4.19 LPC analysis of a THOUGHT vowel token ........................................ 198
4.20 Centroid S in Watt & Fabricius’ (2002) normalization technique ........... 204
4.21 Flow chart of the statistical tests chosen for the analyses of this study ........ 226

5.1 ANAE normalized means of the first and second formants ....................... 248
5.2 Boxplots of STRUT per age group .................................................. 251
5.3 Boxplots of TRAP and BATH for middle-aged speakers ....................... 253
5.4 Boxplots of LOT and THOUGHT for young speakers .............................. 257
5.5 Mean formant values of KIT, DRESS, TRAP, LOT-THOUGHT and STRUT .... 263
5.6 Boxplots of LOT and THOUGHT per age group .................................. 264
5.7 F1xF2 scatterplot of LOT and THOUGHT for speaker 07HPOM ............... 267
5.8 Scatterplot with regression line for all speakers’ EDs and Pillai’s traces of LOT and THOUGHT .............................................................. 268
5.9 Non-pruned Decision Tree of merged LOT-THOUGHT ......................... 271
5.10 Interaction plots from the linear regression of merged LOT-THOUGHT ....... 278
5.11 Application rates for merged LOT-THOUGHT in a back position ............. 283
5.12 Boxplots of TRAP and STRUT per age group .................................... 287
5.13 Correlation of F1 mean values of the individual speakers per age ............ 291
5.14 Application rates for TRAP retraction ............................................. 298
5.15 Boxplots of DRESS and STRUT per age groups .................................. 303
5.16 F1 means per age group and glottal state ........................................ 311
5.17 Boxplots of KIT and STRUT per age groups ...................................... 318
5.18 Application rates of KIT lowering .................................................... 330
5.19 Application rates of KIT lowering in monosyllabic words ..................... 334
5.20 The Canadian Shift in St. John’s, Newfoundland .................................. 336
5.21 Individual vowel measurements of a young female ................................ 349
5.22 Individual vowel measurements of a middle-aged male ....................... 355
5.23 F1xF2 plots of a young and middle-aged speaker compared ................... 357
5.24 Individual vowel measurements of an old male .................................. 359
5.25 F1xF2 plots of a middle-aged and old speaker compared ..................... 361
The aim of the prologue is two-fold. First, I wish to familiarize linguists and those sociolinguists who find the pleasure of their work in the manifold other areas sociolinguistics covers with the variationist approach to sociolinguistics. Second, I wish to critically place my thesis within variationist sociolinguistics. The terms variation(ist) tradition, Labovian (Socio-)linguistics, Sociolinguistics as Language Variation and Change, Variation Theory, empirical linguistics, variationist paradigm and Variationist Sociolinguistics are used interchangeably in the literature. Throughout this thesis, I use the latter to refer to all of them.

As will be detailed in the sections to follow, I deem a critical evaluation necessary: Many scholars who adopt a variationist approach, for instance, do not question its theories and methods, especially with regard to the object of investigation and the statistical evaluation of their data. At the same time, many variationists reject valuable findings and theories from other subdisciplines of sociolinguistics and other disciplines of the social sciences for various reasons. The critical perspective of this thesis is not limited to the prologue, I attempt to maintain it throughout.

0.1 The Beginnings

‘Sound change proceeds according to laws that are without exceptions’. [...The] sound laws that analogy [(the exception to the ‘exceptionless operation of sound laws’)] can disrupt are psychologically conditioned. [But] we are forced from every side to recognize that regularity is inherent in the psychological as well as the physiological [(the laws of sound change)] linguistic principle. This is confirmed by the fact that between the occurrences of the two categories no gap is to be found, only a gradation. [...] The existence of even [...] small [variations in the pronunciation of the individual] contradicts the notion of the impossibility of differences. [...] Old
and new forms are distributed [...] within a single dialect not only according to age, but also according to sex, education, temperament [...] Every stage of language is a transitional stage. [...] There is always talk about the principle of least effort whenever the causes of sound change are being debated. [...] I find it [...] remarkable that the psychological bases of sound change, the social character of a language, the fluid borders of its spatial and temporal variations can be perceived with such lucidity [of a mathematical system (postulated by the neogrammarian dogma).] It is rooted in [an] earlier point of view that separated speech from human beings, that attributed to it an independent life. [...] Linguistics as humanities] does not view language as a natural organism, but as a social product. Most people have adopted [the neogrammarian doctrine] on account of its already noted methodological convenience, which [...] reduces the demands upon independent thinking to a minimum and thus makes possible the participation of an extraordinary number of actually incompetent people in ‘scientific’ work. [...] Sound laws] serve in part, and only in an auxiliary function, for the clarification of the migration of peoples and cultural relationships. But first they must be assimilated within the science itself. We must learn to find the general rule in the specific detail. (Schuchardt 1972 [1885]: 42-66)

The field of sociolinguistics\textsuperscript{1} was launched at the Linguistic Institute in Bloomington, Indiana, USA, in the summer of 1964 (Spolsky 2010: 3). Most of the later founders of several different subfields of sociolinguistics attended the institute that year: William Labov, who established the variationist tradition; Dell Hymes, who shaped the anthropological tradition\textsuperscript{2} and educational linguistics; John Gumperz, who developed interactional sociolinguistics; Basil Bernstein, who oriented his work on class-related ‘codes’ towards American sociolinguistics for a short period of time; and Joshua Fishman and Charles Ferguson, who founded the sociology of language (Bayley, Cameron and Lucas 2013a: 1; Fishman 1997; Paulston and Tucker 1997; Shuy 1990; Spolsky 2010: 3; also cf. Bucholtz and Hall 2008: 401; and Fishman 1991).

Among those founders who did not attend the 1964 meeting, Uriel Weinreich is additionally worth mentioning. Not only did he greatly influence and contribute to the work of his student Labov, which culminated in the (1968) paper (Weinreich, Labov and Herzog 1968), he had also made way for sociolinguistic research and publications along with his friend Fishman roughly a decade before 1964. His (1953) seminal work is regularly cited even today when it comes to understanding language contact (Spolsky 2010: 3). He was the first to adopt the term sociolinguistics from Cuerie’s (1952: 37) work and edited the journal \textit{Word} in which Labov’s (1963) study on Martha’s Vineyard and Ferguson’s (1959) classic paper on diglossia were first published (cf. Spolsky 2010: 3).

An earlier study that might be considered as sociolinguistic is Fischer’s (1958) of New England school children and their use of velar nasal fronting in -\textit{ing} clusters, which he found to be conditioned by sex (gender; cf. Bayley 2013b: 12) and social class (Chambers

\textsuperscript{1} Although such use has been controversial throughout the past five decades, I am using the term sociolinguistics in its broadest sense, to include all subdisciplines in which the interactional nature of social structure and language use has been subject of theoretical and/or practical study (cf. Section 0.3).

\textsuperscript{2} This subfield of sociolinguistics is more commonly known as ethnography of communication.
2009: 14). The early social aspirations included in studies in traditional dialectology (e.g. Gauchat 1905; McDavid 1948) have an “oblique rather than direct” relationship to sociolinguistics, although both disciplines are concerned with language variation, and although it is “plausible to view sociolinguistics as a refocusing of traditional dialectology” (and thus as quite similar to modern dialectology; Chambers 2002b: 6, 2013b: 3).³

The pioneering American sociolinguists of the 1960s had to face the increasing popularity of the Chomskyan framework. In this framework, linguistic theory is largely derived from syntactic structure (e.g. Chomsky 1965). Put differently, emphasis was placed on form over function of language. Linguistic forms were derived from a primarily

ideal speaker-listener, in a completely homogeneous speech-community, who knows his language perfectly and is unaffected by such grammatically irrelevant conditions as [...] errors (random or characteristic) in applying his knowledge to [his performance]. (Chomsky 1965: 3)

The underlying axiom of categoricity (Chambers 2009: 12; or maxim of categoricity Bod, Hay and Jannedy 2003a: 1) in which “all continuities, all possibilities of infinitesimal gradation, are shoved outside of linguistics in one direction or the other” (Joos 1950: 702) in order to make linguistics a “mathematics” or “quantum mechanics in the most extreme sense” (Joos 1950: 701)⁴ led Chomsky’s European counterpart and predecessor Ferdinand de Saussure (1857 – 1913) to view a science of the heterogeneous social uses of language (Saussure’s parole and Chomsky’s performance) as a contradiction in terms (Chambers 2009: 27; also cf. Joseph 2004: 51) and Chomsky to concur: “[it] surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline” (1965: 4). I understand this oversimplification of linguistic analyses⁵ – which was obviously not necessary for the European neogrammarians, as their sound laws were derived from real speech (Chomsky’s performance), despite the neogrammarians’ quite identical view of linguistics as a natural science – as an apparent display of immature⁶ inability and consequently unwillingness to take on the “lawlessness of social phenomena” (Sapir 1929: 213).⁷ Chomsky’s simpler methodological approach to linguistic analyses resulted in hordes of – in Schuchardt’s words “actually incompetent” (1972 [1885]: 65) – followers,⁸

³ Space does not permit a full discussion of this oblique relationship and its possible implications (cf. e.g. Chambers and Trudgill 2004).
⁴ Also see Schuchardt (1972 [1885]: 62) for a quite similar postulate by European neogrammarians.
⁵ As we know today, even the underlying linguistic system (Chomsky’s competence) is full of fuzziness, gradation and continua. It does not at all consist of discrete categories and categorical grammaticality criteria (Bod, Hay and Jannedy 2003a: 1), as, for instance, Joos (1950) and Chomsky (1965) postulated.
⁶ In Chambers’ words, Chomsky also had a “more mature view” to offer (2009: 27).
⁷ I include this quotation at this point to imply that the notion of language being socially embedded was also well established in the U.S. long before Labov’s opposition to Chomskyan linguistics as outlined below.
⁸ Fortunately, the 1990s with their increasing emphasis on microparametric variation (Adger and Trousdale 2007: 262) put an end to the ignorance of performance, which was characteristic of the early Chomskyan tradition.
removing the study of language "from its real-life performance" (Chambers 2009: 27; cf. Schuchardt 1972 [1885]: 64). Joseph even goes one step further in the sense that the isolation of language from its real-life performance is not only an oversimplification of analysis, but ultimately renders language itself useless, as it may "take the form of a pure abstraction for which the only use is to be worshiped as a kind of fetish" (2004: 24).

Among others, Hymes, Halliday and Labov questioned the limited and limiting framework of Chomsky, as Schuchardt (1972 [1885]) did with the limited and limiting framework of the European neogrammarians. Hymes (1997 [1974]: 12-13), for instance, argued for a review of Chomsky’s linguistic competence (parallel to Saussure’s langue) in order to not just merely include grammatical knowledge of an idealized speaker-hearer, but also social and cultural knowledge of real speakers - their communicative competence (cf. Coupland and Jaworski 1997: 5). He further postulated that the goal of sociolinguistics should include identification of “rules, patterns, purposes, and consequences of language use” and “the account for their interrelations” (Hymes 1974: 75). Halliday emphasizes that speakers’ display of different forms to convey the same meaning (function) is indicative of the speakers’ meaning potential. The meaning potential is similar to Hymes’ communicative competence, just that it avoids “the additional complication of a distinction between [performance and competence]” (Halliday 1997 [1973]: 33). Labov (1972b: 185-186) rightly observed the paradoxical nature of Saussure’s claim that data from one idealized speaker would suffice to study the communal langue, but in order to study the individualistic parole one must study language as it is used in the community.

With the advent of Labov’s (1963, 1966, 1969, 1972a, 1972b) work and that of his contemporaries (e.g. Fasold 1967, 1972; Wolfram 1969), variation and validity was (re-)emphasized, extending and challenging existing (Chomskyan) linguistics with data from the speech community (cf. Hymes 1997 [1974]: 14). The data required the application of a (re)new(ed) methodology, i.e. the collection of a (representative) sample from a population (via audio recordings), which naturally resulted in different findings – different from the results derived in a Chomskyan theoretical framework and from the results of sociolinguistic branches with qualitative approaches. In Labov’s early work on Martha’s Vineyard (1963), he put the social reasons for centralization of the nuclei of the diphthongs /au/ and /aʊ/ in focus:

[The fact that all of the linguistic cues cannot explain the observed variation] becomes all the more significant when it becomes apparent that the present trend on Martha’s Vineyard runs counter to the long-range movement of these diphthongs over the past two hundred years. (Labov 1972b [1963]: 9)

Thus, he overtly tried to oppose predominant Chomskyan linguistic theory. As he wrote himself in 1963, he wanted “to understand the internal structure of [English], including the systematic differences which now exist and the changes now taking place [...]”

Chomsky’s concept of performance was much more problematic for Hymes. A detailed description of Hymes’ criticism goes beyond the scope of this thesis (but cf. Hymes 1997 [1974]: 12-13).
The Beginnings

(Labov 1972b [1963]: 7), and, as he wrote in the introduction to his *Sociolinguistic Patterns*, that he wanted to “avoid the inevitable obscurity of [normative edited] texts, the self-consciousness of formal elicitations, and the self-deception of introspection” (Labov 1972b: xix). Bayley’s (2013b: 12) description of Labov’s (1963) Master’s thesis as “careful ethnographic work”, which led Labov to assume that young educated Vineyarders oriented themselves socially and consequently linguistically to traditional life, seems somewhat embellishing in light of studies such as Labov’s Ph.D. thesis on variation in the realization of */r*/ in New York City’s department stores (1966) and Eckert’s (1989a, 2000) careful ethnographic work on Belten High’s *Jocks* and *Burnouts*. What is remarkable about Labov’s earliest study instead is the strong focus on linguistic identity as the explanatory factor for the observed variation. According to Joseph (2004: 60), it was not before the 1990s “that the linguistics establishment was prepared to accept [such an explanation] as scientifically valid”.

After an interruption of some 13 decades since Schuchardt’s work, contemporary (socio-)linguistics is (once more) concerned with social and referential meaning; it is understood as part of the communicative endeavor as well as social action and thus from a functional perspective. In Schuchardt’s (1972 [1885]: 66) words, “we must learn to find the general rule in the specific detail”, or more generally, contemporary (variationist) sociolinguistics “starts from function and looks for the structure that serves it” (Hymes 1997 [1974]: 15; also cf. Tagliamonte 2013a: 383). An example of this approach to linguistic theory in Variationist Sociolinguistics (henceforth VS) may be the intensifier *very*, which can prescriptively be used with positively connotated adjectives, the intensifier *bloody*, which from a formal point of view should not have an intensifying function before an adjective with a positive connotation, and the intensifier *right*, which in example (1c) should prescriptively take its adverb form, *rightly*, in order to intensify an adjective:

(1)  
(a) She was *very* awesome last night. (My example)  
(b) He was *bloody* awesome on stage. (My example)  
(c) Tim was a *right* good fellow, only he drank rather too much. (COHA 1842  
FIC SportingScenes)

Those are three among many other variants of the same linguistic variable, intensifiers, in a contemporary VS framework (cf. Tagliamonte 2012: 3). Although *bloody* as an intensifier is usually attributed to British varieties of English, a quick look at the *Corpus of Historical American English* (COHA; Davies 2010-) reveals an increase in the use of *bloody* in the function outlined in (1b) in the course of the last two centuries. Similar linguistic change is evident in example (1c), as the use of *right* in this function had its peak in the late 19th century, but hardly occurs anymore today (COHA; Davies 2010-). Likewise, variation in the linguistic system occurs in short temporal intervals within or across social spheres, across styles, regions, generations, etc. Phrases such as “that were like sick” are common among teenagers, whereas elderly males may tell stories beginning
6

0. Prologue – Variationist Sociolinguistics

with “they was always workin’ in them days” (Tagliamonte 2012: 2). Correspondingly, references to car as automobile or wheels (cf. Chambers 2009: 13; Tagliamonte 2012: 4) within the same conversation by the same speaker are commonly encountered in everyday speech and can be described as a function of style.

The fundamental maxim of the VS framework is the idea that such variation is an inherent part of the linguistic system (cf. Schuchardt 1972 [1885]: 43) and that every study conducted within it strives traditionally to describe (the original sine qua non of the VS framework; Poplack 2000: 14), but more recently to explain or interpret (Tagliamonte 2012: 3) the rule-governed occurrences of linguistic variants in order to understand language (cf. Bayley 2013b: 11):

The key to a rational conception of language change – indeed, of language itself – is the possibility of describing orderly differentiation in a language serving a community [...] It is absence of structural heterogeneity that would be dysfunctional. [I]t is necessary to learn to see language – whether from a diachronic or a synchronic vantage – as an object possessing orderly heterogeneity. (Weinreich, Labov and Herzog 1968: 100-101)

As refreshingly new as this may have sounded to many contemporary readers, to a host of scholars of the time and to students some decades later, it was not (cf. Chambers 2009: 11). Unfortunately, not only the second generation of VS scholars, but even some of Labov’s contemporaries never seemed curious enough to go back further in time than to the quote above (e.g. Bayley 2013a,b; Bayley, Cameron and Lucas 2013a; D’Arcy 2013; Tagliamonte 2006, 2007, 2012, 2013b; Walker 2010). Notable exceptions that prove the rule are Jack Chambers and Kirk Hazen, among few others (e.g. Kiesling 2011: 4-6) who took the time to look into the origins of VS more carefully (Chambers 2009: 11-38 and Hazen 2007, 2010). As partially outlined above, all of the claims that Weinreich, Herzog and Labov made in (1968) had been well established and known long before Chomsky’s opposition to investigating performance: they originated in the neogrammarian school (and in the non-neogrammarian research as exemplified by Schuchardt’s) on the history of the European language families, in which laws and principles of sound change, i.e. the regularity of sound change, were established, such as Grimm’s law and subsequently Verner’s law (cf. Murray 1994; Schuchardt 1972 [1885]).

One of the most renowned American anthropological linguist of the first third of the 20th century was Edward Sapir (Joseph 2004: 55), who devoted his major academic output to testing the laws and principles established in comparative philology in Europe by applying neogrammarian methods to indigenous American languages (cf. Milroy 1997 [1987a]: 78). His understanding of society and language was surprisingly variationist, considering the year he published the following contemplations:

[The regularity and typicality of linguistic processes leads to a quasi-romantic feeling of contrast with the apparently free and undetermined behavior of human beings studied from the standpoint of culture. But the regularity of sound change is only
superficially analogous to a biological automatism. It is precisely because language is as strictly socialized a type of human behavior as anything else in culture and yet betrays in its outlines and tendencies such regularities as only the natural scientist is in the habit of formulating, that linguistics is of strategic importance for the methodology of social science. Behind the apparent lawlessness of social phenomena there is a regularity of configuration and tendency which is just as real as the regularity of physical processes in a mechanical world, though it is a regularity of infinitely less apparent rigidity and of another mode of apprehension on our part. (Sapir 1929: 213-214)

Put differently, sound change is a regular (i.e. heterogeneously ordered) linguistic process which needs to be methodologically accounted for (cf. Chambers 2009: 11) – again a job for the scientific field of linguistics. The view of the study of language as a primarily scientific endeavor unites Sapir with his contemporary and colleague Leonard Bloomfield (cf. Murray 1994: 113). The two are widely regarded as the founders of American Structuralism (cf. Joseph 2004: 53). Whereas Sapir was more of a creative thinker, Bloomfield tended to be methodological and relied strictly on evidence (Murray 1994: 114).

Labov clearly dissociates himself from Bloomfieldian Structuralism in his introduction to *Sociolinguistic Patterns* (1972b). However, the methods for observing sound change that he would later apply on Martha’s Vineyard in 1961 (1963) were largely anticipated by another key figure in American Structuralism, Charles Hockett (1958: 444-445). In addition to the strong emphasis on the development and practice of a rigorous and accountable set of field methods, the concern with accountability to the data (cf. Section 0.2) unites Labov with American Structuralists (Milroy 1997 [1987a]: 78). I thus consider – unlike widely claimed today – the primarily methodological contributions (the recording of data from real speakers in the speech community for subsequent quantitative analysis) that Labov has made to the field ever since his research on Martha’s Vineyard in 1961 (1963) as inspired by Sapir, Bloomfield, Hockett, and Fischer, after they had transported the European work on sound change and phonetic laws to the U.S. His theoretical contributions (language inherently possesses orderly heterogeneity, his principles, etc.) to the field are only secondary, i.e. they are a consequence of the quantitative methods, and thus rather a refinement of the – at the time – only recently established but almost immediately discontinued U.S. view on language as a social product. Somewhat more revolutionist were Sapir and Bloomfield, who

[w]ith their open contempt for how languages were studied at that time, their [questioning] of received categories and presuppositions [...] and their tendency to disregard any previous scholarly work [...] openly rejected the verities of academic language study and proposed a new paradigm of structural description. (Murray 1994: 123)

---

10 Auer (2004: 1717) and Milroy and Milroy (1985: 339) go much further: They consider Labov’s theoretical contributions as heavily borrowed from the neogrammarians.
Weinreich, Herzog and Labov’s (1968) heterogeneous condition of the speech community has its roots in the community’s linguistic economy (cf. Schuchardt 1972 [1885]: 59) rather than in the linguistic system itself. Such an economy is vital to meet the linguistic demands of every-day life (Tagliamonte 2012: 2). A speech community offers an array of linguistic choices which at least partially constitute its social organization, because these choices reflect the underlying linguistic (in Chomsky’s terms: grammatical) system, which is systematically constrained or motivated by linguistic forms and social variables (Bayley 2013b: 11). If the choices are primarily conditioned by age, but are not age-graded, they often reflect language change (Bailey 2002: 319; Bayley 2013b: 12; Cukor-Avila and Bailey 2013: 246; Labov 1994: 46). The conviction that linguistic choices partially constitute social organizations of the communities they are used in, led Labov and other variationist sociolinguists (e.g. Sankoff 1974: 24; Milroy 1992: 66) to believe that linguistic variation of natural language is most readily found in situations where respondents pay least attention to monitoring their speech, i.e. in their vernacular (or casual style in ‘the sociolinguistic interview’; Labov 1984; given Labov’s 1972b: 208 definition of style as attention to speech).

In the 1970s and ’80s, the VS framework began to be applied to data from other U.S. communities and communities world-wide, such as Anniston, Alabama (Feagin 1979); Cane Walk, Guyana (Rickford 1987); Glasgow, Scotland (Macaulay 1977); Norwich, England (Trudgill 1974); Ottawa, Ontario (Woods 1979); Panama City (Cedergren 1973); St. John’s, Newfoundland (Clarke 1985a,b, 1991); and Sydney, Australia (Horvath 1985). According to Eckert (2005), most of these VS studies can be grouped together to what she refers as the “first wave of variation studies” (2005: 1), because they all “established a regular and replicable pattern of socioeconomic stratification of variables” and “a regular stylistic stratification of variables at all levels in the socioeconomic hierarchy [...] across large communities” (2005: 2-3). The studies raised curiosity as to what underlies the primary social categories (socioeconomic class, sex, age) and gave “rise to a second wave”, which was primarily comprised of “ethnographic studies [focusing on] more locally-defined [and smaller] populations [for longer periods of time]” (Eckert 2005: 1, 5), such as Eckert’s (1989a, 2000), Labov’s (1963), Milroy’s (1980) or Rickford’s (1987). Studies of the third wave, such as Mendoza-Denton’s (2008) or Zhang’s (2005), have employed a community-
of-practice approach to the study of variation, in which variation is seen “as a resource for the construction of social meaning” (Eckert 2005: 1, 17).

0.2 The Framework

The fundamental maxims of the VS framework include notions such as the vernacular (see above), the linguistic variable, the principle of accountability and circumscription of the variable context (cf. Poplack and Tagliamonte 2001: 88-95; Tagliamonte 2012: 1-15, 2013a: 382). In order to describe the orderly differentiation or system underlying the co-occurring use of different forms with similar meanings, the VS framework requires the identification and conceptualization of the linguistic variable (Tagliamonte 2012: 2). A basic definition of the linguistic variable is that it consists of two or more variants with similar meanings (Mendoza-Denton, Hay and Jannedy 2003: 100; Poplack and Tagliamonte 2001: 88; Tagliamonte 2012: 2, 4), as outlined in examples (1 a) to (1 c), and that it should be relatively immune “from conscious suppression” by the speakers (Macaulay 1988: 156; Tagliamonte 2013a: 383). What is more, the variable should be frequent or quantifiable so that unstructured conversation or brief interviews with representatives of the speech community can capture enough tokens for a VS analysis; it should be structurally integrated into a larger system of functioning units (in the case of Martha’s Vineyard it was integrated into the phonological system of upgliding diphthongs: the variable, nucleus of PRICE\textsuperscript{13} /\textipa{ai}/ and MOUTH /\textipa{au}/, consisted of the basic variant [−centralization] or [a] and [aʊ], the intermediate variant [+some centralization] or [æ] and [ʊ], and the final variant [−centralization] or [æ] and [ʊ]; the variable should be highly stratified (in the case of Martha’s Vineyard, the primary social stratum was age) in order to chart the direction of linguistic change based on the sample, i.e. younger/older speakers use different variants more/less often (Labov 1972b [1963]: 7-9, 1966: 49). The number of allophones per phoneme or more generally variants per variable will necessarily vary according to region and social embedding. Linguistic variables can also be determined and analyzed on any other linguistic level: morphology-syntax, lexis, discourse-pragmatics, etc. In areas of linguistics and theoretical accounts outside of the VS framework, variation is incorporated into analyses as well, but often referred to as layering (Tagliamonte 2012: 2) or optionality (Tagliamonte 2013a: 383).

\textsuperscript{13} Labov’s and other work on phonology in North America (e.g. Boberg’s) use a binary phonemic notation, unlike the more commonly used unary notation based on the International Phonetic Alphabet (IPA; e.g. /æi/ versus /æ/; /i/ versus /i/; cf. Labov, Ash and Boberg 2006b: 11-12). Wells’ (1982a) lexical sets provide yet another way of notating the same phonemic systematicity. Instead of providing phonemic symbols, his lexical sets refer to those words in which a phoneme most typically occurs. Throughout the thesis, I will use Wells’ lexical sets in order to aid understanding of Labov’s and others’ binary notation when necessary, especially since most British work in sociophonetics uses Wells’ lexical sets (Boberg 2010: 153), instead of a more detailed notation.
The choices that speakers of a speech community make in the course of their linguistic performances vary systematically, not randomly, and can thus be quantitatively modeled (Tagliamonte 2012: 2, 2013b: 129). As outlined above, the linguistic variants co-vary not only with internal or linguistic patterns (e.g. phonetic context of diphthongs), they also have social meaning and thus correlate, for instance, with external or social patterns, such as prestige of a linguistic variant, speakers’ preferences, addressees’ expectations, group membership, identity, age, etc. The social interpretation of linguistic variants differs markedly along a sociogeographic continuum which starts at a microlevel (face-to-face communication between family members; social situation as in Communities of Practice) and expands to a macrolevel (intra-lingual intercultural communication between strangers; historical situation as in settlement patterns). Within the two poles of the continuum, there are additional regional differences in the social evaluation of linguistic variants. This evaluation reflects the organization and structure of the society in which it is embedded and the time period in which it occurs (Tagliamonte 2012: 6). At the same time, the linguistic choices speakers make and their social meanings accumulate and may ultimately lead to category shifts with the potential to change the underlying linguistic system (Mendoza-Denton, Hay and Jannedy 2003: 99).

The empirical task of an analyst working within the VS framework (or any other quantitative component of frameworks such as those in economics, sociology and social psychology; Sankoff 1988: 151) is to correlate the linguistic variation as the dependent variable (e.g. variants of the nuclei in PRICE and MOUTH) with social (e.g. age) and linguistic (e.g. preceding and following sound segment) categories as independent variables (Chambers 2009: 18). If two or more linguistic variants are available to the speakers of a speech community and if these variants are influenced by different independent variables, it is appropriate and often even necessary to invoke statistical techniques (Sankoff 1988: 151).

The primary target of any VS analysis is to model the simultaneous application of the many internal and external independent variables and their interaction. Such quantitative, probabilistic (as opposed to the neogrammarians’, Chomsky’s and others’ categorical) modeling reveals not only which independent variable favors or disfavors the variants of the dependent variable, but also how strongly or, put differently, how the independent variables are ranked hierarchically, in order to interpret and explain the historical and contemporary, as well as the social and linguistic patterns of language variation and change (cf. Mendoza-Denton, Hay and Jannedy 2003: 100-101; Paolillo 2002: 30-37; Tagliamonte 2013a: 382, 2013b: 129). In this regard, the VS framework combines quantitative (descriptive) and qualitative (interpretive) research (cf. Bortz and Döring 2006: 296). The primary statistical model for this type of analysis is logistic regression, which variationist sociolinguists traditionally apply in the form of the variable rule program VARBRUL.

According to Mendoza-Denton, Hay and Jannedy (2003: 103), the idea of ‘free variation’ has been entirely replaced by the concept of change in progress.
(Cedergren and Sankoff 1974; Rousseau and Sankoff 1978), but recently more flexible and 
accurate tools such as Rbrul (Johnson 2009) or general statistical programs, such as the 
Statistical Package for the Social Sciences (SPSS) and R (R Development Core Team 
2006-13), are increasingly used within the VS framework (cf. Bayley 2013b: 19-20).

The choices that speakers of a community make are regarded as reflections of the 
underlying linguistic system of each speaker and their speech community, which is why 
the linguistic variable must be integrated into this system. In such a contextual analysis 
of language use researchers are held accountable to their data, because they include those 
occurrences of variants of the dependent variable that potentially vary and not only those 
that actually vary (Wolfram 1993: 206). This requirement (Labov 1969: 738), later “the 
principle of accountability” (Labov 1972b: 72, 1994: 550), dictates that a VS analysis of, 
for instance, upgliding diphthongs must include all possible allophones of PRiCE and/or 
MOUTH and not only those which may serve to support the hypothesis of putative sound 
change.

Thus, researchers must remain accountable to all the data they purport to describe. 
They are not at liberty to pick and choose examples that support their arguments 
while ignoring or dismissing counter-evidence. To anyone familiar with academic 
argumentation in humanities fields such as literary criticism, the contrast here should 
be readily apparent. (Gordon 2012: 96)

If the principle of accountability is violated, it is impossible to gain access to functions 
of the underlying linguistic system (cf. Tagliamonte 2012: 10), and thus the linguistic 
analysis is reduced to a mere counting of individual occurrences. The VS framework is 
consequently concerned with one aspect of the linguistic system that is variable. The 
delineation of that linguistic subsystem varies in difficulty, depending on the linguistic 
level the analysis is concerned with. The variable intensifiers outlined in Section 0.1 
belongs to the morphology-syntax level of the linguistic system. Among the many possible 
variants, I outlined three in examples (1 a) to (1 c). Other variants will be in contexts 
that are ambiguous and where the same form can have an entirely different meaning. For 
instance, right is not per se an intensifier, it can be an adjective, an adverb, a verb and 
a noun. According to the accountability principle, “circumscribing the variable context” 
(Poplack and Tagliamonte 2001: 90) or defining the “envelope of variation” (Milroy and 
Gordon 2003: 180) includes only those contexts that are functionally equivalent, i.e. they 
have the same referential meaning.

Although Tagliamonte (2012: 10, 2013a: 385) claims that this form/function asym- 
metry can be managed in two different ways (descriptively by identifying the forms and 
their distribution, and interpretively by ensuring that each form is potentially variable), 
I consider these two approaches to the data rather as consecutive without any implica-
tion of exclusivity: first, forms are interpretively/qualitatively identified and their dis-
tribution is descriptively/quantitatively noted, and second, the analyst ensures inter-
pretively/qualitatively that the forms are not only potentially variable, but functionally
parallel (cf. Sankoff 1988: 149-150). Assessment of application sites and non-application sites in analyses of variation has made zero variants particularly infamous. In the example of intensifiers, the zero variant cannot be counted as part of an analysis of variation (non-application site), because if there is no intensifier, there is no intensifying function. Other morphosyntactic variables such as the relative pronoun system have a zero variant that is functionally equivalent and must thus be included in the analysis, but is notoriously difficult to count because the analyst cannot search for it. Other contexts that have to be excluded from the analysis are, for instance, categorical (i.e. one variant of many is used exclusively), exceptional or indeterminate ones (Tagliamonte 2013a: 385). On the phonetics-phonology level, the issue of which cases to count and which not to count is more easily assessed (Sankoff 1988: 141), as sounds do not carry inherent and referential meaning. They can thus be interpreted as true alternative forms with an equivalent function (allophones) when they occur in the same environment (Gordon 2012: 96).

In summary, the VS framework is comprised of the following analytical procedure (Tagliamonte 2012: 7):

1. Observation within the speech community
2. Identification of the linguistic variable
3. Reconnaissance of variation (if and where)
4. Systematic exploratory observation – define “the envelope of variation” (Milroy and Gordon 2003: 180), establish hypotheses and make claims about the occurrences of variation
   a) What is the inventory of forms constituting the variable context?
   b) What are the patterns?
   c) When does the variation occur and under which circumstances?
   d) Who uses which variant when, where and how?
5. Test hypotheses, observations and claims
6. Discover variable patterns, social and linguistic in terms of:
   a) Which independent variables are statistically significant?
   b) What is their relative contribution to the variation?
   c) What is the hierarchical order in which they influence the dependent variable?
   d) Does the hierarchy reflect the direction predicted in the hypotheses?
7. Interpretation and explanation of the patterns

0.3 Criticism and Reconciliation

Of course, VS does not come without problems of its own. I will not go into the details of the criticism that has been brought forward throughout the decades, but rather summarize some of the positions. Almost all of the extensive criticism goes back to the alleged
dichotomy of qualitative versus quantitative approaches to research and thus ultimately concerns different data sets, which naturally results in different means of data collection and analysis. This dichotomous view originates in the historical development of the social sciences in general, culminated in the German/Austrian positivism dispute of the 1960s between proponents of critical rationalism (e.g. Karl Popper and his student Hans Albert) and of the Frankfurt School (e.g. Theodor W. Adorno and Jürgen Habermas) about the methodology and value judgments in the social sciences (cf. Bortz and Döring 2006: 305) and has since been continuously proven to be without substance.

0.3.1 The Qualitative Approach in the Social Sciences

A brief summary of the historical development of the qualitative approach as employed in the social sciences can be found in Bortz and Döring (2006: 302-307). In essence: the qualitative approach developed from criticism of the well-established and frantically embraced quantitative approach in the natural sciences. In order to consolidate the social sciences as a scientific enterprise in the late 19th century, the methodology of the natural sciences was taken as a role model. Measuring variables, testing hypotheses against data, conducting experiments and axiomatizing knowledge into mathematical formulae were imported (cf. e.g. the neogrammarians' position and later Joos 1950: 701-702 and Chomsky 1965: 4; Section 0.1). Hermeneutics and Phenomenology initiated an alternative movement to the established methods of the natural sciences, and their adapted and modernized versions are still the primary means of data analysis in qualitative disciplines today (Bortz and Döring 2006: 303-304).

The development of the qualitative approach into a discipline in the 1970s and '80s has received major impulses from the positivism dispute (Bortz and Döring 2006: 302, 307): The Frankfurt School criticized that the empirical-analytic approach in the natural sciences is positivist or scientistic in the sense that it cannot accurately account for an understanding of human nature and behavior. Instead the approach would produce trivial results, sketch a mechanistic or deterministic image of humanity and disregard or even ignore the complexity of human and social reality through a particular preoccupation with single variables (Bortz and Döring 2006: 305). In addition, the empirical-analytic approach sought to produce pure factual information free of any value judgments, which was heavily criticized by proponents of the Frankfurt School for the potential danger of misuse of such information. They suggested a dialectic method which attempts to overcome the limits of a theory by changing back and forth from argument/thesis and counter-argument/antithesis and which includes the reason and responsibility of the researchers not to detach factual information from values, morals and ethics.

This dispute was marked primarily by misunderstandings and hostility between the disputants (Bortz and Döring 2006: 306). As Popper (1969: 112), for instance, notes, the critical rationalists never proposed that researchers are able to switch off their humanity
and work unethically or immorally by producing factual information regardless of its implications, e.g. research into nuclear power or genetics. The examples show that the awareness of ethic principles and their implications in the natural sciences in general and the social sciences in particular is still relevant today. For the VS study I conducted in Newfoundland, Canada, it was a prerequisite to apply for ethical clearance by outlining the details of my study before I was allowed to speak to any respondent – as opposed to the conventions in many other parts of the world, indicative of the perceived character of VS.

Reviewing recent findings in empirical or quantitative social research helps assess the value of formulaic and repetitive accusations that this research only yields trivial results or ignores human nature. In fact, such general criticism is far too often targeted only at the methods (e.g. being particular by only measuring variables, instead of being holistic by interpreting texts), but never at the suitability of the methods chosen in order to answer the research question (Bortz and Döring 2006: 302; also cf. Sanko ff 1988: 143). The fact that qualitative methods can equally not be used to research humans holistically in their biographic, social, cultural and historical dimensions is not considered – or at least not mentioned. In this regard, general statements such as ‘measuring variables is unsuitable to research human behavior’ are meaningless (Bortz and Döring 2006: 302).

In the specific case of VS, Sanko has demonstrated convincingly in 1988 that criticism of VS being positivist can hardly be taken seriously. As hinted at in Section 0.2, VS can neither be accurately characterized as an empirical-analytical (or experimental-evaluative; Sanko ff 1988: 142) approach to studying language nor as an introspective-generative one as, for instance, in Chomsky’s work. Instead, Sanko (1988: 142) describes VS as a descriptive-interpretive approach, because it models the independent variables and their statistical significance when influencing the dependent variable quantitatively (descriptively), but also requires identification of the function of the variable forms to include in the analysis qualitatively (interpretively; cf. Sanko ff 1988: 149, 151). Furthermore,

[t]he descriptive-interpretive approach typically sees the researcher deeply immersed in the speech community and intent on reducing the effects of his or her own role as an expert on and/or native speaker of (a more standard version of) the language under study, and as a (usually petit bourgeois intellectual) member of the wider society, with concomitant preconceived notions about communicative behavior. (Sanko 1988: 144)

0.3.2 Formalization and Simplicity

Other criticism Sanko refuted was that VS is fundamentally dependent on formalization in terms of the “community grammar” (Sanko 1988: 141) and the statistical method (Sanko 1988: 150). This goes back to general criticism in the sense of only measuring variables insofar that the social categories usually used to model interdependence of, for instance, age and diphthongs with centralized nuclei, are situated on a macrolevel and
0.3. Criticism and Reconciliation

...can only account for groups of speaker behavior (young versus old), but not individual behavior (human nature). This is only partially true in light of the linguistic variables that are modeled together with the social ones, as the former are independent of the latter, i.e. they hold true for each individual when social variables are not part of the model.

With regard to appropriateness of statistics in the study of human behavior (which excludes the free will component), the main goal of VS is to account for the distribution of linguistic patterns observed in large corpora of speech in numerous different comparable contexts. Whenever the individual’s choice of linguistic form is constrained or motivated by a large amount of linguistic and social factors, for most individuals, their free will is dedicated to the desire to be understood by adhering to the linguistic convention of the speech community, rather than attempting to be creatively busy with individualistic expression of linguistic uniqueness or to display mastery of poetic discourse (but cf. Mendoza-Denton 2002: 486). The issue is, of course, more complex that I can outline here (cf. Mendoza-Denton, Hay and Jannedy 2003: 109-110). The crucial point is that speakers can only actively alter their speech for those linguistic variables that are salient to them; such variables are more likely to be found on a lexical level than on a phonological one (Schilling-Estes 1998: 64-65).

Similar general criticism was provided, for example, by Cameron (1997 [1990]: 57), who considers the “dependence on a naive and simplistic social theory” (emphasis original) as problematic and demands explication of variables such as class, ethnicity and sex. The point she makes is almost entirely based on Romaine’s (1984: 36) claim that “sociology has no solutions to our [sociolinguistic] problems”. As much as Romaine’s view is overstated, Cameron exaggerates the issue in the manner described above (a description of correlations is not an explanation, and thus – following the Frankfurt School – the results are trivial and meaningless, etc.; cf. Cameron 1997 [1990]: 60-61), instead of recognizing that more qualitative approaches – which cannot model the interdependence of several social and linguistic variables – will produce findings that are more meaningful (to her).

0.3.3 Social Classes and Social Networks

In her pioneering work, Lesley Milroy dealt already in 1980 in detail with the problem of the abstract character of Labov’s (1966) social class conception (Milroy 1980: 14) as well as its ethnocentricity (Milroy 1980: 174) and modeled the social matrix of individuals of a speech community on a microlevel of social constraints and linguistic variation via social networks (Milroy 1987b; Milroy and Milroy 1985, 1992). Inspired by Milroy’s work, Rickford (1986) finds Labov’s conception of social class to be problematic for his research on Creoles in Cane Walk, Guyana, (Rickford 1987) for two reasons: 1) The multi-index scales that Labov (1966) used were devised by sociologists in the 1940s and ’50s as a
shortcut for ethnographic interviews in which respondents were asked to shed light on the
social groups that were locally relevant and recognized by the community. In this regard,
such multi-index scales document the results of the “time consuming ethnographic method
of ‘Evaluated Participation’”, and are hence not community-specific if used generically
prior to or instead of such interviews (Rickford 1986: 216). 2) If the method described
above is employed properly in the speech community, the emerging consensual picture
of social classes or groups relies heavily on status or prestige and disregards economic
interests and unequal distribution of power within the society. This functional model
thus fails to take the class struggles into account, which result from differences in interests
and values between classes, and which propel changes in society instead of individuals.
Among others, Karl Marx and Max Weber are prominent proponents of conflict models
which focus on those divergences between classes rather than the commonalities that
constitute society as a whole (Rickford 1986: 216). Rickford (1987) could demonstrate
that a community-specific conflict model of the socially powerless workers on the sugar
plantations (Estate Class) in Cane Walk, Guyana, fully explained the observed variation
by accounting for the complex ethnohistorical dynamics, unlike Labov’s (1966) functional
or consensus model (also cf. Ash 2013: 360; Bayley 2013b: 23; and Mendoza-Denton 2002:
486).

In communities where social class membership is less rigidly imbalanced in terms
of social power, Labov’s consensus model is still considered an appropriate model for
investigating language variation and change. The deep immersion of the researcher into
the speech community allows them to decide whether a conflict or consensus model – or
both – apply to a speech community. Recent studies such as Baranowski’s (2013a) show
that social class continues to be highly productive in VS research “despite the lack of
a single, unified theory of social class” (Ash 2013: 365). This lack is, however, not as
problematic as Cameron (1997 [1990]: 57) argues, since the object of inquiry of VS is not
social class itself, but the dynamics of language variation and change (Ash 2013: 365).

Cameron also rejects a social network approach as an explanation, because such an
abstract “theoretical construct [...] cannot therefore ‘make’ any individual speaker do
anything” (1997 [1990]: 61). I agree, but I find it hard to believe that this is really the
point here. It is the reality of the speakers that makes them choose a certain linguistic
form, whether we try to model this reality in form of social networks or not. I understand
the conflict model approach and the research on the close-knit networks among the upper
and lower social classes and the loose-knit networks in the socially and geographically
mobile middle classes, which conform to Labov’s principle that linguistic innovation em-
anes from the middle of the social hierarchy, as an explication of the variable social
class: “Different types of network structure seem to be broadly associated with different
social classes” (Milroy and Llamas 2013: 421). Similarly to VS, this approach to the
microlevel of social context continues to be quite successful and to be integrated into a
unified model of social network structures and social class (Milroy 1992, 2002; Milroy
0.3. Criticism and Reconciliation

Communities of Practice (Eckert 1989a; Eckert and McConnell-Ginet 1992; Eckert 1996, 2000; Lave and Wenger 1991; Wenger 1999) can be understood to constitute more strictly defined social networks within the social matrix (Eckert 2005: 16) and contributed further to the explication of the social variable (also cf. Joseph 2004: 65; Mendoza-Denton 2002: 486; Mendoza-Denton, Hay and Jannedy 2003: 99).

0.3.4 The Different Styles of Sociolinguistics

The variable style has been detailed likewise (e.g. Schilling 2013; Schilling-Estes 1998, 2002) and extended from Labov’s early definition (1972b: 208) of style as attention to speech to Bell’s Audience Design (1984, 1991, 1997, 2001, 2007) and beyond (Finegan and Biber 1994, 2001). Labov’s definition dissatisfied some of his contemporary linguists (e.g. Bell 1984; Cheshire 1982; Milroy 1987c) as well as, for instance, social psychologist Howard Giles and his colleagues (Coupland 2001: 185). They believed in social psychological, particularly motivational, processes as a basis for understanding style-shifting and developed a dynamic model in the traditional social psychological methodology – speech accommodation theory (cf. Giles and Powesland 1975).

Giles in particular understood Labov’s results in New York from the point of view that interviewees consciously adapted their speech to alleged needs of the interview situation. After the interviewer had entered the homes of their respondents, the established interpersonal relations between the two allowed the interviewee to continue to accommodate their speech in a way they assumed to be appropriate for the contents of the interview. Put differently, Labov’s speaker-hearer who just reacts to contextual formality of a situation is replaced by a proactive speaker-hearer who defines the situation they are in themselves (cf. Giles 2001: 211). I consider the postulate of treating speakers theoretically as active, rather than reactive, as substantial (as Schilling-Estes 1998: 75 has shown), although Coupland (2001), who additionally argues for style to be more than just a formal-informal continuum (but cf. Trudgill 1999 on the separation of formality and style in British English), and Giles (2001) base their criticism on the method per se:

\[
\text{computing averages of how people’s styles in taken-for-granted situations of formality-informality – as occurs in most quantitative, sociolinguistic surveys (as Coupland implies) – does “box away” intriguing variability that should, instead, be investigated. (Giles 2001: 211)}
\]

Cameron’s (1997 [1990]: 57) second criticism is leveled at the same problem Coupland and Giles identified: language does not simply reflect society, society also shapes language (i.e. speakers are active). Although I agree with them on the role of the speakers (based on empirical evidence), I disagree with Cameron’s argumentation. She states (1997 [1990]: 64) that only campaigns (specifically) rallying against sexism in language have actually managed to alter linguistic use, and that, despite a connection, (general) campaigns for equal opportunities (e.g. equal pay) for women and for nonsexist language have never
entailed or simply reflected one another in history. In other words, history shows that only campaigns specifically rallying against use of sexist language have changed language. To summarize this issue more: Only language change actively forced by feminists will change the language, because history shows that only language change actively forced by feminists has changed the language; or simply, in order for language to change, we have to actively force it to change, because historically it supposedly has always been like that.

Since the value of such argumentation should be readily apparent, I will refrain from outlining that such a hypothesis can never be tested in order to falsify it (which is not to imply that inductive conclusion is generally an invalid scientific method) and from referring to studies on the success (or rather lack of success) of language planning and language policy (e.g. l’Académie française). The crucial point here is that VS is firmly grounded in the belief that language change is propelled by social variation (Mendoza-Denton, Hay and Jannedy 2003: 102), which attributes – at least partially and only recently – an active role to humans in VS (see below for actual research on proactivity).

Giles’ contemplations inspired Bell, who in a manner of speaking domesticated Giles’ work by putting his ideas into a broader variationist framework. His audience design accounts for both the speakers’ behavior and the suspected motive for the behavior, because he believed that interviewees can consciously sound more non-standard (Bell 1997: 242), i.e. they consciously vary on an intra-speaker level which is derived from the differences they associate with a group of speakers (e.g. their family versus the researchers).15

Preston (1991) and Rickford and McNair-Knox (1994) put Bell’s design to the empirical test. Whereas Preston (1991: 36) generally finds support for Bell’s theory, Rickford and McNair-Knox’ (1994) results are not as straightforwardly supportive. They state that they did not control for the familiarity of McNair-Knox with Foxy Boston (1994: 251, 258), the interviewee (both of them are African American), so that her higher frequency counts of copula deletion and absence of plural and possessive -s when talking to McNair-Knox as opposed to the lower frequency counts when talking to an unfamiliar European American interviewer cannot be attributed to Foxy’s stylistic adjustment to the difference in audience (African American versus European American audience) alone (as they enthusiastically concluded; 1994: 266), but also to familiarity with the interviewer (also cf. Cukor-Avila and Bailey 2001: 255 for further issues). As Cukor-Avila and Bailey (2001: 263-266) could show in their replication of Rickford and McNair-Knox’ study, familiarity with the interviewer exerted the strongest social influence on the dependent variables, instead of style in the sense of designing the speech to the audience (Cukor-Avila and Bailey 2001: 267).

A similar assumption of speakers being proactive rather than merely reactive was proposed by Schilling-Estes (1998: 75), because her case study Rex O’Neal, resident of

---

15 In this sense, Bell’s audience design provides further explication of the social variable (cf. Preston and his social status axiom; 1991: 36-38).
Ocracoke island, showed that he was able to switch in and out of performance style (performing or role-playing the Ocracoke dialect), which was phonologically conditioned and quite systematic (also cf. Eckert 2001; Schilling-Estes 2002; Schilling 2013). Even more so, Schilling-Estes maintains that speakers always shape their speech stylistically to fit the need of any situation they become part of, so that each single style speakers exhibit represents a performance style (Milroy and Gordon 2003: 50; Schilling 2013: 332; Schilling-Estes 2008: 971). She could also show that his performed speech behavior could not be explained with divergence from and convergence to an imagined or present audience (as claimed by an audience design theory):

It is crucial to note that Rex is not playing the part of the quaint, heavily vernacular fisherman in RESPONSE to his focus on his audience of linguists. When he focuses on this audience, he could shift just as readily into exaggeratedly standard speech as into exaggeratedly non-standard speech. (Schilling-Estes 1998: 75; emphasis original)

Labov himself revised his conception of style as attention to speech insofar as he implicitly admitted that a differentiation of casual style (vernacular) and careful style (answering interview questions) in interviews based on channel cues (e.g. changes in the pitch of the voice or laughter) is indeterminate in the sense of too difficult to apply in an objective and reliable manner across different studies. Wolfram (1969: 58-59), for instance, rejected this differentiation, because laughter in an interview can mean a lot of things (dealing with stress or nervousness, happiness, etc.), the immediacy of the vicinity of channel cue and linguistic feature under analysis is unclear and the combination of channel cue and linguistic feature tends to be too infrequent to invoke statistical measures (also cf. Rickford and McNair-Knox 1994: 238; Schilling 2013: 332). In 1989, Labov (1989: 11) subsumed careful and casual speech under spontaneous speech (i.e. conversational, but real and natural speech), because it was sufficient to distinguish oneself as a sociolinguist theoretically from those who rely on introspective data (1989: 51).

### 0.3.5 Sex and Gender

Other noteworthy debates centered around notions such as the variable gender versus sex and the diffusion of change. The debate on the interpretation of biological sex and subsequent social construction of gender is most prominent in sociolinguistics, literature and other social sciences (cf. Cheshire 2002: 423-424). In the particular case of VS, studies by Eckert (1989b) and Labov (1990) have shown that the variable sex strongly interacts with other social variables. Eckert (1989b: 265), for instance, also pointed out that girls assert their category identity (female sex) more through language than boys do. She further asserts (1989b: 247) that sex should be replaced by gender, because the social role of the speaker may not conform to their sex (biological bias). Labov (1990:
209), however, emphasizes retention of sex in order to preserve the chance of replication of studies.

As Cheshire (2002) observes, many of the generalizations made, for example by Labov, lack empirical foundation and reveal little about the relation between language and social life. The emphasis of the VS framework is, however, placed on replication in order to gain the largest possible understanding of language change and the language faculty in general (Cheshire 2002: 428; Labov 1990: 208). Further research into the construction of gender and its interaction with language in the speech community was done, for instance, by Eckert and McConnell-Ginet (1992, 1999). With regard to the diffusion of change through a population, a summary of the debate whether, for instance, the implicational scale model was a good alternative to the quantitative model can be found in Rickford (2002).

0.3.6 The Influence of Age

In comparison to the variables discussed above, age has received the least amount of analytical attention in VS and has consequently been least criticized. According to Llamas (2007: 69), it is least understood, often modeled uncritically “and treated as a biological fact” similarly to speakers’ sex. The influence age has on any member of any speech community with regard to the social responsibilities, norms of behavior, constraints and the way they speak is apparent, and by no means uniform. One 18-year-old member of a speech community does not have the same social place as any other 18-year-old (cf. Eckert 1998: 155). In addition, the concept of age is itself culture-specific, as, for instance, many Africans do not count and memorize their absolute chronological age in the same fashion Europeans and Americans do. According to Cheshire (2004: 1552), ethnic groups in Africa conduct certain social or religious rituals which serve as the basis for social categories that may be relevant to the concept of age. Some of such rituals exclude women, so that they cannot be categorized in the same relevant age-related social categories (Cheshire 2004: 1552).

The potential issues age carries when conceptualized as an independent variable may not be so readily apparent, as chronological age (and biological sex) highly interacts with global categories such as social class and ethnicity and with local categories such as social networks, social/religious rituals and communities of practice (cf. Cheshire 2004: 1552; Llamas 2007: 71). The most arbitrarily categorized and least explored time span in the lives of speakers is adulthood (Llamas 2007: 71), covering the twenties after adolescence until the onset of the old years – whether they start at the age of 70 (Trudgill 1974: 28) or 60 (Labov 2001a: 170). The early years of adulthood are usually seen as a time of linguistic standardization according to the speakers’ professional and personal ambitions and thus stand in crucial opposition to adolescence (cf. Chambers 2009: 190). The most important influence on the speech of young adults is exerted by the perception of
0.3. Criticism and Reconciliation

the prestigious variants they use as socially legitimate – the linguistic marketplace (Ash 2013: 359; Llamas 2007: 72; also cf. Bourdieu 1991), especially for laborers whose income depends on servicing homes or public institutions and professionals whose occupations require linguistic intercourse, such as teachers, lawyers, actors, secretaries and announcers (Chambers 2009: 190-191). Speakers older than what is conventionally considered as adulthood are usually part of apparent-time VS studies in order to take their frequencies of variants as a point of comparison to those of younger speakers, but not to analyze “the state of being ‘old’” (Llamas 2007: 72; also cf. Cheshire 2004: 1553). The common-sense observation of older speakers’ linguistic behavior is that they are no longer exposed to the social constraints of the linguistic marketplace, so that their speech alters towards linguistic variants that are either local or carry covert prestige (but see below). Since speakers move through their life trajectories as individuals and at the same time as part of an age cohort (Eckert 1998: 151), it is “reasonable and convenient to group [them] by various stages in life” (Llamas 2007: 72), although the observed linguistic variation across age cohorts may not be explainable beyond doubt (Llamas 2007: 73).

The most crucial problem with regard to age and age cohorts in apparent-time VS studies is the question of whether linguistic variation reflects language change in a speech community or age-graded speaker behavior (Cheshire 2004: 1553). According to Cheshire (2004: 1555), Llamas (2007: 73) and McMahon (1994: 241), speakers usually show a U-shaped linguistic behavior throughout their lives with regard to the linguistic variants they use that are not undergoing change. Prestigious or standard variants are generally preferred in adulthood and covert prestigious or localized variants are usually used in adolescence and old age. This means that the individual behaves linguistically unstable throughout their life, but the community around them is linguistically stable (Labov 1994: 83). A comparison of a set of young speakers at a given point in time to a different set of old speakers at the same point in time does therefore not necessarily indicate language change (a linguistically unstable community), but the fact that the analyst does not have data to explore whether the younger speakers will still use the same variants when they grow old and whether the old speakers used the same variants when they were young, respectively (i.e. whether the individuals are linguistically stable in an unstable community), or not (i.e. whether the individuals are linguistically unstable in a stable community).

Chambers (2009), however, emphasizes that age-grading is rare and realized in such a predictable pattern (cf. e.g. Macaulay 1977: 47) that it does not refute the hypothesis underlying apparent-time studies: speakers will essentially speak the same as they grow older (i.e. the individuals are linguistically stable; Chambers 2009: 207; Labov 1994: 107). They do so because aging is often accompanied by “the linguistic reflex of [...] conservatism”, and because the “language-learning capability” of older speakers has slowed down (Chambers 2009: 197). In other words, the vernaculars of speakers are acquired during childhood, enhanced in adolescence and adjusted/extended in early adulthood, so
that their repertoire of socially significant variants has become sufficient in adulthood. As already mentioned in Section 0.1, Bailey (2002) and Cukor-Avila and Bailey (2013) generally find support for the language change explanations in various apparent-time VS studies. They further find that vernaculars on the morphology-syntax and discourse-pragmatics levels of language are not stable under the age of 20, so that VS investigations have to be careful in their claims about language change when such speakers are included. Both of these findings support Chambers’ reasoning above.

0.3.7 The Self-perception of Variationist Sociolinguists

General negative criticism towards an academic method per se is as meaningless as general positive criticism. Puzzlingly, even some contemporary sociolinguists are compelled to write “William Labov and his students have developed a quantitative research paradigm that [...] is often described as sociolinguistics” and that “this term is misleading in several ways”: first, because linguistics cannot exist without its social dimensions, second, because other paradigms have been described as sociolinguistics (e.g. ethnography of communication) “that are not quantitative and/or [rather thus] address rather different types of research questions”, and third, because sociolinguistics implies an “exclusive focus on social considerations” (Walker 2010: 2).

Another more implicit example of this narrow definition of sociolinguistics is provided by Mallinson and Kendall’s (2013) great review of the numerous studies that have employed concepts, theories and/or methods from other disciplines. They consider any combination of discipline, concept, theory or method with Labov’s VS (as the method of sociolinguistic inquiry) as an interdisciplinary approach. For instance, the Milroys’ “social network analysis has been incorporated into sociolinguistic research to some extent” (Mallinson and Kendall 2013: 156) and “provides an analytical framework for quantitatively analyzing social relationships [...]” (2013: 157). One possible explanation for such a – in my view – distortion of subdisciplines within sociolinguistics might be the fact that American first-year university students are (still) confronted with introductory books that answer the question of “What is Sociolinguistics?” (Van Herk 2012: 2) with “[m]any of us would trace the birth of modern sociolinguistics as a subdiscipline [of linguistics?] to the work of William Labov, starting in the early 1960s” (Van Herk 2012: 4).

Although the authors recognize the existence of related approaches (such as sociology of language), they claim that these are related fields rather than related subfields or centers of gravity within sociolinguistics, because the degree of emphasis on linguistics (rather than sociology) is higher in sociolinguistics (e.g. Van Herk 2012: 5). By the same token, such a non-specified determinant for inclusion and exclusion of subfields in and from a discipline such as sociolinguistics implies, for instance, that historical linguistics should be considered the history of language and thus part of history rather than linguistics. Similarly, it implies that proponents of a corpus-based approach should be
considered as corpus linguists, but proponents of a corpus-driven approach should not, and that functionalists should be considered as syntacticians, but formalists should not – and vice versa. The same is true for the definition of when to refer to a study as an interdisciplinary one: a narrow view implies that a corpus-driven study is interdisciplinary and that formalists’ work is interdisciplinary. Sociolinguistics is an interdisciplinary discipline which “originated at the intersection of sociology, anthropology, and linguistics” (Mallinson and Kendall 2013: 153).

As should be apparent, during the first wave studies of variation (Eckert 2005: 1), VS became institutionalized as a discipline, so that most American sociolinguists who were university students after the 1970s only learned Labov’s VS and some of them never questioned it for the sake of the convenience of only asking research questions answerable with that method. As a consequence the loosely defined determinant for inclusion/exclusion of subfields in/from sociolinguistics has been arbitrarily redefined (in North America) to an understanding of sociolinguistics to be VS only, so that the conception of what to consider as interdisciplinary takes VS as a starting point. This conception, however, does not change the reality of the research, and the names attributed to this reality have arbitrary accentuations within their senses which are indicative of an author’s knowledge of this reality and of who they want to be associated with rather than the reality itself (more on this below). In other words, the boundaries between fields or subfields are blurry and become recursively redefined with each individual author. In this sense, the Milroys’ social network analysis has contributed to an explication of the social class variable in a VS framework, regardless of whether we name their work part of sociolinguistics, part of VS or an interdisciplinary complement. The only real difference between the members of this ‘name race’ is their connotation. To me, the equation of sociolinguistics with only the work of Labov and his students has a negative connotation, because it disregards the work of other sociolinguists who work outside a VS framework.

Because these related approaches to the study of language use and social structure differ in their degree of emphasis on linguistics and because sociolinguists other than Labov deserve recognition of their work, I redefine the determinant for inclusion/exclusion further and consider the field of sociolinguistics as a continuum within which the study of language can range from the sociology of language (Fishman) to VS (Labov), but also to critical discourse analysis (Fairclough, Wodak). Such a view – arguably shared with other sociolinguists (cf. Bayley, Cameron and Lucas 2013b; Bucholtz and Hall 2008; Coulmas 1998, 2005; Fishman 1991, 2010; Hymes 1974; Kerswill, Johnstone and Wodak 2010; Mesthrie 2011; Tagliamonte 2006, 2012) – rests on the actual sense of the component parts socio- and linguistics. This negotiation of terms and their referential meaning is in a similar manner rooted in categorical thinking as the debate about quantitative versus qualitative approaches to science.\footnote{I may add here that researchers’ needs to differentiate their work terminologically (e.g. positivism, scientism) from that of others is in large part due to a putative and claimed demand of universal}
The separation of the two approaches based solely on data as outlined above is incomplete. According to the psychologists Bortz and Döring (2006: 298), the separation further reflects discrepancies in the researchers’ understanding of science and in their image of humanity. The discrepancy is usually expressed in contrastive pairs, such as idiographic versus nomothetic, fieldwork versus laboratory experiments, holistic versus particular, exploratory versus explanatory, understand versus explain or case versus sample, in order to characterize the qualitative versus quantitative approach. The categorical or mutually exclusive character that is implied in such a dichotomous classification reflects the values that researchers attribute to the approaches rather than the reality of research. If, for example, a quantitative approach is characterized as particular, its deficiency compared to a holistic qualitative approach is implied. In reality, however, methods usually attributed to the latter (e.g. qualitative interview) may only capture very few aspects, whereas quantitative methods (e.g. standardized questionnaires) may cover the whole range of relevant aspects (Bortz and Döring 2006: 299).

As the example of Howard Giles and his experimental or laboratory methods to research accommodation of speakers has shown, it was not as accepted by sociolinguists as Bell’s audience design. One possible explanation is that the experimental method was considered unnatural or artificial compared to a sociolinguistic interview setting, so that the findings are regarded as artifacts of the method rather than generalizable observations. However, as Labov (1972b: 209) states, sociolinguistic interviews aim at identifying and capturing the vernaculars of the respondents which they usually use when they are not systematically observed, but the sociolinguistic interview itself is a systematic observation (Observer’s Paradox; more generally Hawthorne effect). In other words, a sociolinguistic interview situation is not an everyday and thus not a natural situation. Even if it were, each of these situations would be completely different from all others, so that any generalization from one of them must be well explained (cf. Bortz and Döring 2006: 299).

validity of a certain approach (e.g. only VS is sociolinguistics or human nature cannot be measured in variables), so that, for example, Fishman never considered his work as sociolinguistic proper, but as sociological or macro-sociolinguistic in order to distance his work from that of Labov and other sociolinguists (Fishman 2010: xv; Spolsky 2010: 5). I specify the problems of universal validity of the quantitative approach to the study of language structure and society below; the problem with universal validity in science is, however, more general. It has a fatal logical consequence: in order to verify or prove unlimited validity of a theory, one has to test the theory in an infinite number of trials or studies. In other words, verification of a theory constitutes an inductive conclusion that falsely assumes universal validity of a theory based on a few single events. Even natural laws know anomalies, i.e. exceptions to the universal validity of a law. Every VS study is probabilistic, i.e. every VS study proposes hypotheses that are probably true, neither definitely nor universally, so that contrary cases are explicitly allowed (cf. Bortz and Döring 2006: 18). The probabilistic character of hypotheses in the social sciences requires certain conventional testing criteria in order to falsify them. The most important one is statistical significance (Bortz and Döring 2006: 10), which does not come without problems of its own, especially in VS (cf. Johnson 2009).

This contrastive pair was originally used in the 19th century to differentiate the natural sciences from humanities. Very generally, the natural sciences generally propose natural laws (nomothetic approach), whereas the humanities attempt to describe individualistically single historical events or products of culture (idiographic procedure). Such a definition of the two sciences is considered outdated (Bortz and Döring 2006: 298).
300). Such generalizations from a representative sample to a population are inductive, although the quantitative approach has generally turned away from inductive empiricism with the critical rationalists. They successfully developed a model in which hypotheses are deduced from theories and consequently falsified – not verified (cf. e.g. Cameron’s 1997 [1990] argumentation and footnote 16 above). Qualitative approaches do, however, work deductively as well whenever they apply a predetermined set of categories to a text in order to find out whether these categories occur in it (Bortz and Döring 2006: 301).

As a last example, I want to return to Cameron’s (1997 [1990]: 61) claim that a social network as a theoretical construct cannot explain any speaker’s behavior (cf. Subsection 0.3.3). As she herself outlines, this issue is grounded in the perception of the human being as active rather than reactive. In other words, she does not consider naming external and objectively observable variables that affect humans’ linguistic behavior as an explanation, because it reduces the image of humanity to that of a puppet or marionette that does not act self-determinately. Instead, we can only understand human behavior through communicative comprehension of the subjective or sentimental world view and inner reasons of the respondents (cf. Bortz and Döring 2006: 301). In this sense, qualitative approaches do fall back onto the same “explanation” whenever they use theoretical concepts in an analysis that are unknown to the respondents or outside their subjective world-view and inner reasons to behave in a certain way (cf. Bortz and Döring 2006: 302). Put differently, the qualitative and quantitative approaches to researching human (linguistic) behavior are not inherently one or the other. Instead of dichotomies, they constitute a continuum of possible approaches – in Bortz and Dörings’ words bipolar dimensions (2006: 298). Identifying the distance or closeness to one of its poles requires thorough reflection of differentiating criteria (as partially outlined above). Empirical studies should thus be evaluated by their results, function, scientific significance, and suitability of methods in order to answer the research questions, and not simply by the type of research method employed (Bortz and Döring 2006: 303).

0.3.8 Towards Unity in Sociolinguistics

During the course of VS studies throughout the decades, we can find numerous examples that incorporate concepts, theories and methods intra-disciplinary from other subfields of sociolinguistics. Gumperz (1971 [1968]: 120), for instance introduced the concept of “the speech community as a field of action” through which real speakers acquire and use language in order to construct their identities as part of the community (cf. Mallinson and Kendall 2013: 158). Labov (1963) used identity as the main explanatory factor for the observed variation on Martha’s Vineyard (cf. Section 0.1), and identity is still used when it comes to analyzing bedroom communities on islands (such as Petty Harbour in Newfoundland, Canada; Childs, Van Herk and Thorburn 2007; Childs et al. 2010; Van Herk, Childs and Thorburn 2007). Hymes’ (1964) concept of the ethnography of communica-
Prologue – Variationist Sociolinguistics

Eckert’s (1989a, 2000) study combines careful (qualitative) ethnographic analysis with VS methods (Bayley 2013b: 23). Cukor-Avila and Bailey (2001), Rickford and McNair-Knox (1994) and Schilling-Estes (1998) use qualitative case studies, which are part of larger VS studies, to review the conceptualization of the variable style. The central social variables in VS such as sex, ethnicity, social class and more recently ideology and identity are of equal importance in discourse studies, and are continuously explicated via the methods of critical discourse and conversation analysis: Gumperz and Cook-Gumperz (1982: 1) emphasize the mutual effect of these variables and identity, which can only be understood by the “communicative processes” that continuously constitute the variables; for Fairclough (2001), social power is exercised through discourse; for Wodak and Reisigle (2003) the same is true for racism (cf. Mallinson and Kendall 2013: 159).

By applying quantitative and qualitative methods, Eckert and McConnell-Ginet (1992, 1999, 2003) outline how speakers indicate and negotiate social power through language, but focus on the construction of sex within locally-based speech communities. Social power inequalities are usually accompanied by and expressed in aesthetic or value judgments about language. These language ideologies or linguistic stereotypes become relevant in endeavors of language planning and language policies, as for instance in the case of African American Vernacular English (AAVE) in European American schools. White teachers tend(ed) to (sub)consciously judge AAVE to be socially stigmatized as it sounds like a deteriorate and consequently deficiently learned variety of general American English. VS research into the linguistic status and structure of AAVE was conducted by Fasold (1967, 1972), Labov (1969) and Wolfram (1969), and it was more or less directly concerned with elaborating on the role of AAVE in the White classroom by explaining it to the White teachers. In the reprint of some of her most important and qualitative research on AAVE since the 1970s, Smitherman (2000) pointed out how the linguistic norms and the verbal rituals of African Americans are deeply rooted in the African verbal tradition. It helped the African slaves newly arrived in the U.S. to adapt English, which they were forced to speak, into a coded language, so that they could communicate secretly to one another in front of the White Master (slave owner). Most African Americans resist educational and linguistic inequality, because of the deep roots in African verbal tradition and the powerful capabilities AAVE offers to them to identify themselves as in-group members. Numerous other studies of the interactional character of language, race and ethnic identity can be found in Fishman and Garcia (2010).18

---

18 For an interdisciplinary approach to the study of language use and social structure see Mallinson and Kendall (2013).
A simple review of the literature might have convinced me that such empirical principles had no place in linguistics: there were many ideological barriers to the study of language in everyday life. (Labov 1972b: xix)

Despite the attempts of unity in sociolinguistics, which I have addressed in the previous subsection, it seems that many ideological barriers to the study of language in everyday life remain to this day: Over the past 40 years since Labov’s quote, proponents of VS often disregarded the qualitative approaches to the study of real language, which is not to imply that this disregard did not go both ways. Qualitative sociolinguists rejected VS, for instance, because of the lack of the explanatory power of its findings (cf. Section 0.3). A review of the literature convinced me that a VS study seems feasible and appropriate to fill a gap in research on the Canadian Shift in the speech community of St. John’s, Newfoundland, despite my academic socialization by predominantly qualitative (socio-)linguists.

The ideological barriers within VS become particularly apparent when adaptations of its methodology are suggested by subsequent research in later decades that is not necessarily quantitative in nature. However, I did not follow the rather strict methodological guidelines originally formulated by Labov (1984), but used a methodology adapted from the suggestions made by Tagliamonte (2006), among others. While some traditional variationists might feel obliged to criticize these adaptations, the process of my research and its results – particularly those yielded by, in my opinion, the much more accurate mixed-effects regression analyses – suggest that the adaptations are not only appropriate but also supportive of Labov’s earlier work.

The aim of my VS study is to provide empirical support for the presence of the Canadian Shift, a mainland Canadian English vowel shift, in the middle class of young urban residents of St. John’s on the island of Newfoundland. The length of this dissertation is owed to the fact that I employ many more than the one or two statistical methods typical
of VS studies. This thoroughness is in turn owed to the contradictory claims that have been made in the contemporary literature with regard to the linguistic behavior of this speech community.

Two large-scale projects, using acoustic phonetic methods, have been carried out recently that include a very small number of participants from Newfoundland: two in the *Atlas of North American English* (ANAE or the atlas; Labov, Ash and Boberg 2006b) and six in the project *Phonetics of Canadian English* (Boberg 2008b, 2010). In terms of the Canadian Shift in Newfoundland, the results of the former are diametrically opposed to those of Boberg (2008b), which are in turn contradictory to those of Boberg (2010). The value of these results for Newfoundland is negligible, because the focus of these studies has been on providing an overview of the dialect regions and the sound changes that are taking place on the entire North American continent (Labov, Ash and Boberg 2006b) and in the entire nation of Canada (Boberg 2008b, 2010), respectively.

Some sociolinguists and/or dialectologists have characterized the linguistic behavior of Canada’s urban middle class as relatively homogeneous – at least in comparison to the varieties of English spoken by members of the middle class in Britain and the U.S. Canada’s population is scattered across the west-east dimension of the nation, extending from British Columbia in the west to the Maritimes in the east. In the north-south dimension, Canadians almost exclusively reside on a small stretch of land near the U.S. border.

The linguistically most important influx of settlers to Ontario, Canada, and other provinces were the United Empire Loyalists, who left the northern American colonies after the American War of Independence. From Ontario as the linguistic center, these settlers moved to the west, which is largely taken as the reason for Canada’s relative linguistic homogeneity. One of mainland Canada’s vowel shifts is the Canadian Shift, first proposed in the 1990s, which, in line with this homogeneity, seems to be present across the whole nation – at least from Vancouver, British Columbia, to Montreal, Quebec. I will discuss and review all of these factors and their relationships in Chapter 2. Further details of its contents and of the other chapters described below will be given in each chapter’s respective introduction.

Newfoundland and Labrador, Canada’s easternmost Atlantic Province, has been unaffected by this settlement pattern, which is almost exclusively due to the island’s isolated geographical location. It was populated much earlier than the Canadian mainland by predominantly British and later Irish settlers. As a Dominion of the British Crown, the province joined Canada only in 1949.

Traditionally, the island of Newfoundland has been described as linguistically, geographically and economically distinct as well as isolated. And yet, recent research into the phonetics and morphosyntax of Newfoundland’s diverse English varieties attested the presence of many innovative mainland Canadian-English features, except for the Canadian Shift – at least in an undisputed fashion with more than six respondents. The settlement,
the sociolinguistic situation, the recent linguistic changes in the different varieties of Newfoundland English and the relationship, in terms of linguistic identity, between Canada and its youngest (and until recently poorest) province, Newfoundland, will be addressed in Chapter 3. With the review of the literature on the Canadian Shift in Canada (Chapter 2) and the Canadian Shift vowels in Newfoundland’s Englishes (Chapter 3) in place, I will derive and outline my research questions as well as hypotheses for this dissertation in the last section of Chapter 3, Section 3.4.

The subsequent and longest chapter is devoted to the methodology I employ and the sample I use (Chapter 4). As mentioned above, two large-scale sociophonetic studies have been conducted recently, which included a few Newfoundland participants and differed in their findings for the Canadian Shift on the island. Both of these studies inspired the methodology I use here in this dissertation. My methodological choices concerning the sociolinguistic interview differ somewhat from those suggested by Labov (1984). In the planning stage of my fieldwork in 2011, I relied on the suggestions made by Tagliamonte (2006) and much of the research that has already been reported in Section 0.3 of the Prologue. In addition to the make-up of my judgment sample, I will outline the methodological comparability and the structure of as well as motivation behind the sociolinguistic interviews I conducted in the first two sections of Chapter 4. Since I employ acoustic phonetic methods to analyze vowel quality, I summarize Acoustic Theory briefly in Section 4.3, which is followed by a section on vowel measurements and one on data preparation for the analytical statistical analysis (Sections 4.4 and 4.5, respectively). These two sections form the core of my methodology in terms of comparability with the studies conducted by Labov, Ash and Boberg (2006b) and Boberg (2008b, 2010). In the last two sections of this chapter, I outline the theoretical assumptions of the statistical tests I use, and I motivate the selection of tests for the respective vowels under investigation.

The results and their discussion are the focus of Chapter 5. I consider the results of my study as somewhat surprising, given the current state of research into the Canadian Shift. I will begin with a general overview of the results in form of a graphical representation (Section 5.1). This graph already hints at the results of the statistical analyses per vowel in the sections to follow. The structure of these four sections (5.2 - 5.5), containing the results for one vowel each, is in principle the same in that the sequence of the statistical tests I evaluate is identical per section. The last section in this chapter (Section 5.6) summarizes the individual findings per vowel and establishes the basis for combining these results with one another and with the relevant literature. In addition, I will investigate three case studies which are representative of the different age groups I included in my sample.

Via the combination of the linguistic details provided by investigating the case studies and the results of the quantitative analyses, I will draw my conclusions concerning the status of the Canadian Shift in St. John’s, Newfoundland. These are detailed in the first
section of Chapter 6, followed by the limitations of my study (Section 6.2) and finally an outlook regarding future research (Section 6.3).
Chapter 2

English-speaking Canada and its Vowel Shifts

The varieties of English in Canada are by and large characterized by a relatively homogeneous linguistic behavior of the urban middle classes, a notion which has not been undisputed, let alone the controversy surrounding it resolved. Canadian English is well-known for two peculiarities at least in the English- and to a lesser degree the non-English-speaking world, namely Canadian Raising and the Canadian Shift. Although both of these characteristics are by no means limited to Canadian English, the former in particular has been the target of linguistic stereotypes held by Americans towards Canadians. While the Canadian Shift is a vowel shift, Canadian Raising is not per se. However, some studies have found an incipient sound change within Canadian Raising: The nucleus of the MOUTH vowel has been found to be fronted or centralized, regardless of whether it is raised or not, among young females in particular. While this vowel shift has been noted as early as the 1950s, the Canadian Shift has been first proposed in the 1990s. Both Canadian Raising and the Canadian Shift have been used as criteria, among other linguistic features, to establish isoglosses along the U.S.-Canadian border that distinguish Canadian English from the varieties spoken in the U.S. I will outline the disputed notion of homogeneity and Canada as a distinct dialect region in Section 2.1.

The focus of this dissertation is on the Canadian Shift, a sequence of vowel movements that has been likened to vowel chain shifts in the U.S. such as the Californian Shift and the Northern Cities Shift. As I will outline in Section 2.2, the underlying linguistic principles of vowel chain shifts have been proposed by Labov (e.g. Labov 1972b; Labov, Yaeger and Steiner 1972; Labov 1991, 1994), after he had reviewed the work of the neogrammarians on, for instance, the Great Vowel Shift, and after he had conducted research on vowel shifts on Martha’s Vineyard (1963), in New York City (1966) and in Philadelphia (Labov, Yaeger and Steiner 1972). By incorporating earlier motivations for vowel movements in the vowel space, such as maximal dispersion, he proposed a set of principles for chain
shifts and mergers, which can be summarized as *Peripherality Hypothesis* (Labov, Ash and Boberg 2006b). In this theory, vowels move in a direction which is dictated by their non-/peripheral position in the vowel space. The peripherality of a vowel can change under certain circumstances which can in turn affect neighboring vowels. In pull-chain shifts, these may move subsequently towards the slot of the other vowel that has changed its peripherality. In such chains, the maximal perceptual contrast between two vowels is maintained and the vowel inventory is unaffected. Mergers and splits likewise vacate vowel slots but reduce and increase, respectively, the number of vowels in the inventory. This theory does, of course, not come without problems of its own.

The Canadian Shift has originally been proposed as a chain shift by Clarke, Elms and Youssef (1995). In this first proposal, the vowels involved in the shift moved in a pull-chain as the consequence of a merger. Later studies in the immediately following years found different results in terms of the vowels involved in the shift. Over a decade later, the phonetic nature of the shift had been criticized, which has sparked a host of research into the Canadian Shift. This has naturally contributed to an increased understanding of the shift, particularly in terms of the social variables involved and the shift’s regional distribution, which developed from partially unresolved controversies. I will outline this development chronologically in Section 2.3: The first subsection is devoted to the origins of the Canadian Shift (2.3.1), followed by the criticism of its chain-shift nature in Subsection 2.3.2. The first continent-wide study to investigate the Canadian Shift along with shifts in the U.S., the Atlas of North American English (Labov, Ash and Boberg 2006b), helped to resolve some of this controversy and will be discussed in Subsection 2.3.3.

After the first decade of research on the Canadian Shift, several reactions have been published, and most of them followed the lead of Labov, Ash and Boberg’s atlas in terms of the shift’s phonetic nature (Subsection 2.3.4). More recent studies focused primarily on some of the social variables that have not been under crucial investigation in relation to the shift, such as style and ethnicity, but also on rural areas on the Canadian mainland that have not yet been investigated on a large scale. The latter studies in particular will only briefly be mentioned in Subsection 2.3.5, due to my focus on an urban area of Canada’s youngest province, Newfoundland. The same is true for ethnicity, as St. John’s, Newfoundland, is predominantly populated by speakers of a mixed English-Irish heritage. Style is a more central variable in my thesis, but has hardly been researched. My chronological representation of the controversies about the Canadian Shift serves four main purposes: to review 1) the homogeneity of Canada’s middle-class English; 2) the vowels involved in the Canadian Shift; 3) the questioned chain-shift nature of the Canadian Shift; and 4) the role of the social and linguistic variables in the variation of the vowels involved in the Canadian Shift. Each of the studies I review is described in such a way that these purposes can be addressed, although this is not uniformly possible. The essence of the literature review on the Canadian Shift will be summarized in Subsection 2.3.6.
2.1 Canada as a Dialect Region

All regions in Canada are institutionally bilingual with English and French as the two official languages. Demolinguistic calculations show, however, that monolingual French speakers constitute roughly 20% of the 35.5 million Canadians (Statistics Canada 2012) and that 81% of these speakers live in the province of Quebec (Boberg 2008a: 352). According to Chambers (1991: 91), only 4% of speakers have French as their mother tongue when Quebec is left out of the calculations, and all of those who speak none of the official languages as their mother tongue learn English instead of French as their second language. Outside French Quebec and bilingual parts of New Brunswick and eastern Ontario, Canada is predominantly English-speaking (Boberg 2008a: 352).

Research on the variety of Canadian English was rare until the 1950s compared to that on the varieties of American and British English. Between the 1950s and the 1990s, research was predominantly carried out in the framework of traditional dialectology (Chambers 1991: 90). More or less detailed overviews of this research were published by Avis (1973), Bailey (1982) and Chambers (1979, 1991), who outline the four major thematic foci the research has contributed to: 1) the sociohistorical origins of English in Canada; 2) the use of alternating American and British word forms, spellings and pronunciations; 3) the documentation of relic areas; and 4) Canadian Raising (fronted and/or retracted onsets in the diphthongs PRICE and MOUTH before tautosyllabic voiceless consonants; cf. Boberg 2008a: 351; Chambers 1989: 79). Until recently, broad consensus about the homogeneous character of Canadian English was predominant (e.g. Avis 1973: 51; Chambers 1991: 91, 93; Chambers and Hardwick 1986: 24; Woods 1979: 33). As Chambers and Hardwick (1986: 26) point out, this view was already established in the last third of the 19th century, “when Ontario was itself the Canadian frontier” (cf. Figure 2.1). This homogeneity has to be qualified as follows: 1) it is prevalent in the Canadian provinces west of the Atlantic provinces and Quebec, namely Ontario, Manitoba, Saskatchewan, Alberta, and British Columbia; 2) it is prevalent in urban middle-class speech; and 3) it has emerged abruptly, instead of stepwise, as the indigenous speech of the first or second generation offspring of the early settlers (Chambers and Hardwick 1986: 24). In other words, the homogeneity resulted directly from the settlement history of the Canadian inland, starting in Ontario (Boberg 2010: 28). The southern Ontarians of the late 19th century, who were the offspring of English, Scottish and German settlers (Chambers and Hardwick 1986: 26), conformed remarkably to the linguistic standard set by the 7500 to 12,000 refugees of the American Revolution (United Empire Loyalists) from 1776 to 1793 (Chambers 1991: 91). Two other important waves of immigrants (Scots

---

19 Although Chambers’ observations are based on census figures from 1981, the picture has not profoundly changed within the past 30 years. He outlines, for instance, that 67% of Canadians are monolingual English speakers and 17% are monolingual in French (1991: 90). In comparison, the 2011 census data shows that 66% of Canadians are monolingual English and 20% are monolingual French speakers (Statistics Canada 2012).
and Irish from 1851 to 1861 and German, Dutch and Belgian from 1901 to 1911) settled in established towns in southern Ontario and assimilated to the standard Canadian accent within a generation or two, despite some linguistic traces that remain to this day (Chambers 1991: 91). The southern Ontarians that dominated the westward migration in large numbers beginning in the 1870s were largely comprised of white-collar workers such as doctors, teachers, bankers and merchants. Through the usual support of geography (valleys, prairies) and infrastructure (railroads, highways), their linguistic standard dominated newly founded rural communities in which other settlers (e.g. Britons, Irishmen, Germans, French Canadians) prevailed in agricultural work. This domination was enhanced by education so that the offspring of immigrant farmers soon conformed to the linguistic standard of the white-collar Ontarians (Chambers and Hardwick 1986: 26), as the Canadian provinces were founded and settled: Manitoba in 1870, Saskatchewan in 1885, Alberta in 1885 and British Columbia in 1886 (Chambers 1991: 91). As a result, urban, middle-class Canada, except Labrador and Newfoundland, Prince Edwards Island, Nova Scotia, New Brunswick and Quebec (cf. Figure 2.1), speaks white-collar southern Ontarian English – standard Canadian English of American English roots – generally homogeneously. The evidence that supports this characterization of Canadian English has been largely derived from research on Canadian Raising (e.g. Chambers 1973, 1980, 1989,
2.1. Canada as a Dialect Region


In Chambers’ (1973) earliest treatment of Canadian Raising in a generative framework, he analyzed the raising of the nucleus in price (/ay/ or /aI/) and mouth (/aw/ or /AU/) as a single process, because structural factors such as impressionistic vowel quality (higher), linguistic environment (voiceless) and vowel length (shorter) were identical for both nuclei. One of the explanations he sought was for the fact that Americans commented on the Canadian [2U] diphthong in lexical items such as house, but never on [2I] in items such as wife (cf. Chambers 1989: 76). His interpretation was that the nucleus in mouth is slightly higher and backer than that in price (Chambers 1973: 115), so that only the former is perceptually salient to Americans. Later research, however, found that Americans in the upper Midwest and the North in general increasingly raised their nuclei in /aI/, but not in /AU/ (e.g. Allen 1976: 25; Vance 1987: 195). In other words, only the latter can be salient to Northern U.S. Americans, because it is different from their realization (cf. Vance 1987: 207-208; also cf. Chambers 1989: 76). Among others, Labov pointed out that the raised nucleus in price was a feature of 16th and 17th century English and has been transported to U.S. regions such as New England, western New York, the Lower and Upper South and Martha’s Vineyard, Massachusetts, where it remained the favored form until the 19th century (1972b: 10). In comparison, the nucleus of /AU/ was less often raised and if so, only in rural areas of New England and eastern Virginia in the 19th century (Labov 1972b: 11). Only in the South, the raising of the price nucleus was as regular – before voiceless consonants – as the raising of the price and mouth nuclei in Canadian English (Labov 1972b: 10). Although raising also applied to both diphthongs on Martha’s Vineyard, “the voiceless environment is only the most favoured, and raising occurs elsewhere” (Chambers 1989: 77; also cf. Labov 1972b: 20). Thus, it is agreed that Canadian English is distinct from American English with regard to raising of price and mouth, because this raising is limited to following voiceless consonants. And yet, raising of the two diphthongs seems to function independently, as a change in progress was noticed for mouth but not for price in urban, inland Canada in the 1970s and investigated sociolinguistically in the late 1970s and in the 1980s.

Chambers (1980), for instance, analyzed the change in progress, mouth- or /aw/-fronting, in North Toronto (Ontario) in 1979 and found that speakers over the age of 40 exhibited the pattern predicted by the Canadian Raising rule: raised [2u] before voiceless consonants, but non-raised [2u] elsewhere. Younger speakers fronted both the raised and non-raised nucleus in mouth more frequently before voiced consonants than before voiceless segments, with females in the lead in each age group (1980: 24-25 and Figure 9). The change thus yielded the raised and fronted onset qualities of [2u], [wu] and [2u] before voiceless consonants and the same ones in addition to the fronted onset qualities of [2u], [Au] and [æu] before voiced consonants (Chambers 1989: 81). Mouth-fronting
was tentatively interpreted as eventually eliminating the Canadian Raising rule from Canadian English (CE) phonology (Chambers 1989: 82, 1991: 100).

A similar change was attested for Ottawa, Ontario (Woods 1979), Vancouver, British Columbia (Chambers and Hardwick 1986) and Victoria, British Columbia (Davison 1987), but not for Montreal, Quebec (Hung 1987). In Toronto, Ottawa, Vancouver and Victoria, young speakers in general and females in particular led the incipient sound change, while style showed a completely irregular pattern. The tendency was, however, for style to show a flat stylistic profile, meaning that in general there was no stylistic stratification (Chambers and Hardwick 1986: 32; Hung, Davison and Chambers 1993: 253, 261; Woods 1993: 161). Follow-up studies in Toronto in 1997 and Victoria in 2005 yielded real-time support for these findings (Chambers 2006: 113-115). In Montreal, the fronting index for /aw/ was between a minimum of 54 and a maximum of 76 and thus higher than among Toronto's 22-years-olds (Hung, Davison and Chambers 1993: 261). However, “[this] variation [...] appears to be between individuals, regardless of age and gender, whereas in Toronto the variation is between age and gender groups” (Hung, Davison and Chambers 1993: 262).

As mentioned above, this pattern was understood to be a function of Canada’s settlement history in that the founding population of the inland region, from Ontario to British Columbia, showed profound linguistic similarities, both regarding Canadian Raising and its concomitant, MOUTH-fronting, but also other linguistic variables on the phonological and lexical level of CE (Chambers 1991: 99). The motivation behind this tendency towards homogeneity is essentially a sociolinguistic and sociopolitical force: Hung, Davison and Chambers (1993: 265) emphasized that the sound change (/aw/-fronting) was “towards the standard American phonetic realization of the back-gliding diphthong [...]”, as a “result of a change in political heteronomy towards the United States”. This interpretation seems to receive further support from other studies on general leveling tendencies at the American-Canadian border (e.g. Boberg 2000; Sadlier-Brown 2012; Woods 1993; Zeller 1993). Moreover, the Canadian urban middle class appears to form one speech community sharing a set of norms. Hung, Davison and Chambers (1993: 265) concluded that Montreal differed from this speech community, as its anglophone population was not a community, but a non-francophone fringe, i.e. Montrealers were much more heterogeneous in terms of style, culture and fashion than, for instance, Torontonians. The instances of /aw/-raising in Montreal seemed to be incursions from the inland urban standard and thus would mildly reflect the larger speech community of anglophone Canada.

Later research by Boberg (2004a, 2009) showed, however, that MOUTH-fronting was present in Montreal English, but the key correlate was neither age nor gender, but ethnicity: While the Jewish community was converging to the British-origin community, many Italians resisted Canadian features, most likely due to the social and residential segregation in Montreal (Boberg 2009: n.p.; also cf. Boberg 2010: 218) – essentially the same interpretation as that of Hung (1987) and Hung, Davison and Chambers (1993), but Hung
2.1. Canada as a Dialect Region

(1987) did not include ethnicity. Other more recent research supports the homogeneity of inland, urban Canadian English. For instance, Hagiwara (2006: 136-138) outlined the presence of Canadian Raising in Winnipeg, Manitoba. In addition to Canada’s inland, Kinloch and Ismail (1993) reported that Canadian Raising and its concomitant sound change were also present in Fredericton, New Brunswick – an Atlantic Province. This finding was substantiated via acoustic phonetic methods by Boberg (2008b: 139-140): He attested both, Canadian Raising and MOUTH-fronting, for New Brunswick and Nova Scotia. Although Pratt (1982) found no such raising for PRICE (/ay/) on Prince Edwards Island, D’Arcy (2000, 2005) reported /aw/-fronting and raising for St. John’s, Newfoundland (also cf. Graham 2008). As I will detail in Section 3.1, the province’s settlement ceased in the late 19th century, approximately at the time when Manitoba, Saskatchewan, Alberta and British Columbia were founded and settled (1870-1886) and the United Empire Loyalists’ in-migration to Canada was not a factor in Newfoundland (Clarke 2008a: 93). In summary, the pattern seems to be that homogeneity is initially present in inland urban Canada and maintained when sound changes advance. This homogeneity subsequently spreads to the east to include the Atlantic Provinces as well, despite their differences in settlement histories. The view of CE homogeneity across thousands of miles as a sociolinguistic motive is maintained by some authors to this day (cf. e.g Chambers 2012) – at least on the phonetic level of CE (cf. e.g. Dollinger 2012; Dollinger and Clarke 2012).

Recent research as conducted by Boberg (2004a, 2005, 2008b, 2010) questioned this homogeneity of Canada’s urban middle-class speech on the lexical and phonetic level of CE. Boberg (2010: 209) maintained that for the phonetic variables Canadian Raising of MOUTH before voiceless consonants (awT), MOUTH-fronting, fronting of GOOSE (uw), fronting of GOAT (ow), unretracted START (ahr), and raising of TRAP before nasals (æN) and /g/ (æg), 20 six dialect regions of Canada could be identified (2010: 201): British Columbia; the Prairies, containing Alberta, Saskatchewan, Manitoba and northwest Ontario; the remainder of Ontario; Quebec; the Maritimes, based on respondents from New Brunswick and Nova Scotia only; and Newfoundland (see map in Figure 2.1). Since British Columbia and the Prairies united to form a western dialect region in some respect, the total number of regions was at least five (Boberg 2010: 209).

The picture of Canada as a dialect region outlined by Labov, Ash and Boberg (2006b: 148) differed greatly from Boberg’s depiction, as he stated himself (2010: 208). Since he apparently contributed a lot of information to the Atlas of North American English (ANAE or the atlas) with regard to CE, these differences may mainly be the consequence of the focus on structural dialectology and sound change (chain shifts) on a continental scale and less so one of the data sets (cf. Boberg 2010: 200). Labov, Ash and Boberg

20 As mentioned in the Prologue, I use IPA notation, Labov, Yaeger and Steiner’ (1972) binary notation as well as Wells’ (1982a) lexical sets interchangeably. More details on the individual notation systems can be found in Subsection 4.4.2.
(2006b: 220) acoustically analyzed 33 Canadian participants from coast to coast, aged 10 to 70 years (2006b: 28), and Boberg (2008b, 2009, 2010, 2011) acoustically analyzed 86 Canadian respondents, born and raised in the regions outlined above, all of them young students at McGill University in Montreal (2010: 201). These were part of a larger data set collected in a study called *Phonetics of Canadian English*, which was comprised of 108 McGill University undergraduate students from across North America (2010: 144).

ANAE identified only two dialect regions in Canada: “Inland Canada”, extending from Vancouver to Montreal, and the Atlantic Provinces, covering the remaining provinces east of Montreal, namely the Maritimes (New Brunswick, Nova Scotia and Prince Edward Island) and Newfoundland (2006b: 148). Labov, Ash and Boberg (2006b: 130, 146) defined Canada as a dialect region based on isoglosses for almost the same phonetic features that Boberg (2010) used: Canadian Raising of *mouth* before voiceless consonants, fronting of *goose*, (non-)fronting of *goat* and fronting of (or unretracted) *start*. The most important feature of CE that distinguished it from other dialect regions of North America in the atlas was the Canadian Shift (including the presence of the low-back merger as [o]), which I will outline in detail in Section 2.3. Boberg also investigated the Canadian Shift, but unlike Labov, Ash and Boberg (2006b: 220) he found no significant effect of region for this feature (2010: 203).

While both studies identified the Atlantic Provinces to be distinct from the rest of Canada (urban, inland Canada), Boberg (2010: 208) identified further isoglosses within inland Canada based on the above-mentioned features: While British Columbia was divided from the Prairies by the fronting of *goose* and *goat*, it was united with the Prairies by the raising of *trap* before nasals and /g/ and the fronting of raised *mouth* and *start* as opposed to Ontario. Thus, the separation of the Prairies from British Columbia was a less important division than that of the Prairies from Ontario. Atlantic Canada was distinct from Ontario regarding the fronting of *goose*, *goat* and *mouth*, but united with Ontario with regard to the raising of *trap* before nasals and fronting of *start* as opposed to the west (Boberg 2010: 208). Quebec was an intermediate region of uncertain status between Ontario and the Maritimes with variable occurrence of CE features. Newfoundland was distinguished from the Maritimes and the remainder of Canada. Overall, the English of young middle-class Canadians can be thus divided into at least five dialect regions (see above) based on the phonetic features outlined here. Boberg (2010: 209) added that this model of Canada’s varieties corresponded well to the one established with lexical data.

This depiction of CE into Inland Canada and the Atlantic Provinces Labov, Ash and Boberg (2006b: 148) offered is strikingly similar to the two-fold classification made in the decades before Boberg’s research outlined above: a dialect region considered as homogeneous west of Ontario and a non-homogeneous region east of Ontario. While Boberg’s analysis questioned the homogeneity of middle-class CE in his analysis of regional differences (2010: 199-209), he emphasized its correctness in a comparative sense of
middle-class Canadian English versus middle-class British or versus middle-class American English (Boberg 2010: 28).

2.2 Chain Shifts and Mergers

Before I continue to review the literature on the most important CE feature of this dissertation, the Canadian vowel chain shift, I will briefly embed it in the (socio-)linguistic theory of sound change. Sound change has been the target of investigation in historical linguistics for centuries, usually and not unexpectedly without accounting for the influence of social factors. This disregard among some historical linguists is rooted in their understanding of sound change: linguistic factors account for the origins of sound change and social factors for its propagation, which becomes apparent in the fact that most of their work is concerned with linguistic explanations of the origin of sound change. A review of these linguistic explanations can be found in Thomas (2011: 274-279), chiefly among them the controversy of teleology, maximal dispersion and the model of mis-perception.

An example of a teleological concept is linguistic economy or the principle of least effort (Schuchardt 1972 [1885]: 59), usually a conditioned sound change such as assimilation and deletion. In African American Vernacular English (AAVE), for instance, speakers make frequent use of consonant cluster reductions such as the deletion of word-final /k/ from -sk consonant clusters (desk as [dEsk]). If the velar is deleted, the effort of articulating it is deleted as well, similarly to the deletion of an off-glide in the monophthongization of a diphthong (cf. Thomas 2011: 274). Deletion does, however, not explain the diphthongization of monophthongs, which requires more tongue movement and thus more effort in production. The construct of maximal dispersion is an example of concepts that seek to explain sound production with increased effort. Vowels may be realized near the physiologically possible extremes of a speaker’s vowel space, acoustically measurable in formant-one by formant-two plots, when certain linguistic constraints are met. They are thus maximally dispersed in the vowel space and increase the ‘margin of security’ between one another, meaning that a vowel is no longer uttered coincidentally similar in perception to another vowel. In vowel chain shifts, one vowel can thus change its quality towards the original quality of the adjacent vowel that has already shifted, usually referred to as a pull chain, while maintaining the margin of security (e.g. lowered [i] is not confused with [e] when the latter is produced with an [æ]-like quality). Likewise, vowels can push adjacent vowels out of their original position in the vowel space, a chain shift usually referred to as push chain (cf. Thomas 2011: 275). This construct alone can, however, not explain a merger of two vowels such as the low-back merger of the LOT and THOUGHT vowels (/ɑ/ and /ɒ/, respectively).

The model of mis-perception is a phonetic model that seeks to account for sound changes that are motivated by perceptual confusion (Thomas 2011: 276). Taking AAVE as a case in point again, most speakers front their interdental fricatives: /θ/ and /ð/ are
often realized as /t/ and /v/, respectively. While these pairs are articulatorily distinct, they are acoustically quite similar and thus prone to perceptual errors. Arguably, however, they may also be indicative of the complexity of producing interdentals, but most other varieties of English stop the interdentals rather than front them. In any case, another example for which the change accounts is the fronting of the GOOSE vowel before alveolar stops – or more generally instances of a back vowel before coronals (cf. Thomas 2011: 276). In an /ut/ sequence, the coarticulatory effect of the following voiceless alveolar stop fronts the round high-back vowel to [u], which may be perceived – particularly in a language-acquisition setting – as the vowel target and consequently be learned.

As mentioned above, the problem of those models and concepts is their shortcoming in accounting for the role of social factors or the propagation of sound change. In sociolinguistics, Labov and colleagues in particular opposed such accounts on the basis of the fact that origin and propagation of a sound change cannot be separated. Especially the last example made the necessity to account for both factors apparent: just because one speaker improperly learns a variant, we cannot assume that we are looking at the origin of a sound change. Having stated that, I consider the strict postulation by Labov that one individual introducing a word or slip of tongue is not the origin of a linguistic change as too strong. He outlined that “[...] the origin of a change is its ‘propagation’ or acceptance by others” (1972b: 277; emphasis original). And yet, speech communities are not homogeneous and features that were originally introduced by one individual and have spread to others may coexist as variation before they change, if ever. As Thomas (2011: 279) notes, such inherent community variation means that innovations already have a constituency of speakers who acquired/learned it independently before the innovations spread.

The main concern of sociolinguistic accounts of sound change has been the compatibility with social factors such as prestige of innovations and linguistic identity of speakers. Prestige-related changes are considered as changes from above: “[The] stigmatization [of an innovative form] initiated change from above, a sporadic and irregular correction of the changed forms towards the model of the highest [social] status group – that is, the prestige model” (Labov 1972b: 179; emphasis original). The stigmatization of innovative forms by the highest social group is accomplished “through their control of various institutions of the communication network” (Labov 1972b: 179), which indicates some social awareness of the innovative forms. As Labov (1994: 78) later defined, change from above refers simultaneously to the level of social awareness and the status of social groups. Highest groups of one speech community view other speech communities as possessing higher prestige and borrow linguistic features from them, which occur more frequently in the formal styles and often, but not always, with women (Labov 2001b: 274). These changes from above are opposed to changes from below, which appear first in the vernacular of young adults as well as youth in late adolescence with females in the lead and represent the primary form of linguistic change that operates within the system (Labov
2.2. Chain Shifts and Mergers

1994: 156, 2001a: 279). Until such changes near completion, they are below the level of social awareness even for “phonetically trained observers” (Labov 1994: 78). Changes from below may be introduced by any social class, but one of the most consistent findings in variationist sociolinguistic studies is that changes usually originated at and spread from the interior of the social hierarchy (Labov 1972b: 294, 1994: 49, 53, 300; also cf. Baranowski 2013a: 275 on mergers). This interior is represented by “skilled workers, technicians, clerks, teachers, merchants, and leaders of local organizations and political parties” (Labov 1994: 156). Among these, the prototypical speakers are those with the highest local prestige: “upwardly mobile individuals from ethnic groups who have entered the community in the last three or four generations” (Labov 1994: 156). The innovations in changes from below are associated with covert prestige (cf. Trudgill 1972), which rests on expressions of solidarity with lower classes and socially inferior groups of a society (Thomas 2011: 288).

The systematic sound changes “that make up the major mechanism of linguistic change” are changes from below (Labov 2001a: 279). Despite the irregularity of style in the sound change shown by Chambers and Hardwick (1986) as well as Hung, Davison and Chambers (1993) I discussed above (i.e. innovative fronted MOUTH is not consistently most frequent in the vernacular), Labov considered this an example of a change from below (2001a: 281). The characteristics it shares with changes from below are: young, female leaders, mechanism of sound change (see below), phonetic conditioning (particularly manner of articulation; cf. Labov 1994: 543) and interior social class. This will become essential for the interpretation of my data in Section 5.6. Linguistic identity has usually been investigated in connection with speakers’ perceptions of membership in social classes, social networks, communities of practices and their ages as well as the constraints of standardization in adulthood (cf. Subsections 0.3.3, 0.3.5 and 0.3.6).

Sociolinguistic investigation of the ‘major mechanisms of linguistic change’ have been led by Labov, focusing on constraints on vowel chain shifting, vowel mergers and on testing the exceptionless operation of the linguistic constraints on sound changes – the Neogrammariam Hypothesis (cf. Section 0.1). Labov, Yaeger and Steiner (1972) reviewed historical data on sound change in different languages and acoustically analyzed varieties of English in order to address the motivations for chain shifting. They have proposed a set of principles governing chain shifts which were revised in subsequent work such as Labov (1991, 1994) and Labov, Ash and Boberg (2006b). The pivot in Labov’s major mechanisms of linguistic change has been summed up as the Peripherality Hypothesis by Labov, Ash and Boberg (2006b: 17). They outlined that in West Germanic Languages such as English, the historical vowel opposition, short and long, enters into a phonological opposition of lax versus tense vowels. The feature \[±tense\] refers to several phonetic features such as extended duration and maximal dispersion, or extremely peripheral articulation, with increased articulatory effort (Labov, Ash and Boberg 2006b: 16). The phonological vowel space available for North American English vowels is defined acoustically in a
In chain shifts, tense nuclei move upward along a peripheral track (Labov, Ash and Boberg 2006b: 16; also cf. Labov 1994: 116, 176).

II. In chain shifts, lax nuclei move downward along a non-peripheral track (Labov, Ash and Boberg 2006b: 16; also cf. Labov 1994: 116, 176).
III. In chain shifts, tense vowels move to the front along peripheral paths, and lax vowels move to the back along non-peripheral paths (Labov 1994: 200).

Labov (1994: 261-262) summarized his principles of chain shifts as “[i]n chain shifts, peripheral vowels become less open and non-peripheral vowels becomes [sic] more open”. All three of these principles are unidirectional and operate within subsystems such as short vowels, front and back upgliding vowels and ingliding vowels (cf. Labov 1994: 272; Labov, Ash and Boberg 2006b: 18), which does, of course, not ultimately lead to mergers at the highest and lowest extremes of the vowel envelope, as the vowels maintain their margins of security (maximal dispersion). Once lax vowels have reached the lowest point [a] and/or tense vowels have reached the highest points [i] or [u], they shift peripherality. From such a shift the lower and upper exit principles were derived:

IV. In chain shifting, low non-peripheral vowels become peripheral (Labov, Ash and Boberg 2006b: 18; also cf. Labov 1994: 280).

V. In chain shifting, one of two high peripheral morae becomes non-peripheral (Labov, Ash and Boberg 2006b: 18; also cf. Labov 1994: 281-284).

The other principles governing movement across vowel subsystems such as nasalized vowels in French (Principles VI to VIII) can be found in Labov (1994: 284-292). The Lower Exit Principle (IV) allows vowels that descended under Principle II to change subsystems, if pressured to move further, and consequently become subject to Principle I (Labov, Ash and Boberg 2006b: 18). In the Canadian Shift (Subsection 2.3), this happened with LOT (/ɑ/) as it migrated from the short subsystem to the long and ingliding subsystem of vowels (Labov, Ash and Boberg 2006b: 19). The Upper Exit Principle was derived from reviewing vowel movements across subsystems in the Great Vowel Shift and appears to be specific to West Germanic languages (Labov, Ash and Boberg 2006b: 18). Since both of the principles create vacant slots in the vowel space in the original subsystem, they may initiate chain shifts. When they are combined with the principles governing movement within subsystems, they cause varieties to diverge from one another on the phonetic level. If, for instance, /i/ vacates the highest slot of the system of long vowels under the Upper Exit principle, the remaining long vowels will rise under Principle I (cf. Labov, Ash and Boberg 2006b: 19).

The counterpart to such chain shifts is merger, where the number of oppositions and vowel classes that existed prior to any vowel shift are reduced. Like vowel movements, mergers are unidirectional. They are governed by two related principles:


Since “a word class is a historical accident”, any vowel has to be learned as part of a certain word when that lexical item is first learned (Labov 1994: 311). These assignments are arbitrary and inherited from the history of a language. Because the linguistic signs that make up historical word classes are arbitrary, restoring a merger would imply relearning each of the occurrences of a merged vowel in all words it occurs in (Labov 1994: 312). For instance, if the low back vowels /ɒ/ and /ɔ/, which are merged in many varieties of North American English, were to be restored or ‘unmerged’, native speakers of these varieties would have to relearn the differentiation between the low back vowels, i.e. that cot does not contain /ɔ/ but /ɒ/ and that caught does not contain /ɑ/ but /ɔ/, etc. Consequently, Labov (1994: 312) argued, it is much easier to learn word classes of a new language than to unmerge a merged vowel class. Garde’s Principle is based on the empirical observation that at no time in the history of languages such a reversal has been accomplished by enough speakers to restore word classes (Labov 1994: 312-313). Herzog’s Principle, or corollary of Garde’s, is a spatial reflection of historical events as they affect neighboring dialects (Labov, Ash and Boberg 2006b: 19). Linguistic changes normally advance along lines of communication such as river valleys and lag behind in mountainous areas. Newer forms overlap with older forms, but both new and old forms coexist in widely isolated relic areas (Labov 1994: 313). Over time, all the forms intermediate between various original vowel pronunciations and an ultimate merger are distinctions which eventually die out (Labov 1994: 314). Mergers often initiate chain shifts, as they may create a vacant position in a vowel subsystem or increase margins of security among the remaining elements of a subsystem (Labov, Ash and Boberg 2006b: 19). As mentioned above, the low-back merger includes the movement of short /ɑ/ or LOT from the subsystem of short vowels to the subsystem of long and ingliding vowels, whereby it becomes long open /ɔ/ or THOUGHT (Labov, Ash and Boberg 2006b: 19). This merger vacates the space in the subsystem of short vowels for /æ/ or TRAP to lower and retract, possibly initiating a pull chain among the front lax vowels.

Labov’s Peripherality Hypothesis, particularly the principles governing movement within and across vowel subsystems, has received some criticism. As Thomas (2011: 281) points out, Principles I and II apparently only apply to languages which have a lax/tense or short/long vowel distinction. When vowel systems lack such a distinction, Labov (1994: 121) responds, all vowels behave like tense vowels. Moreover, in languages with a ‘marked’ series of vowels such as nasal vowels in Romance or Tupi-Guarani languages, the marked system can possibly function as the equivalent of the tense vowels in Indo-European languages (Labov 1994: 290). According to Thomas (2011: 281), more empirical evidence is needed to substantiate these contentions. Another objection that has been pointed out for Australian English is that there are exceptions to some of the principles (Thomas 2011: 281). The LOT vowel in this variety is short but nevertheless peripheral, and as a consequence of its peripherality (Principle II), short LOT is raised. Furthermore, Thomas (2003: 156-162) has shown that peripherality of a vowel may be
2.2. Chain Shifts and Mergers

only an incidental effect of shifting, but not a controlling factor, since his research in the South of the U.S. indicated simultaneous deperipheralization and lowering of the tense nucleus in face (Thomas 2011: 281). According to the theory, vowels first shift peripherality and then move on the track they are positioned on. Labov (1994) outlines some exceptions to various principles and emphasizes that they are tendencies rather than strict and inviolable rules. Another problem is the motivation behind these principles. Thomas (2011: 281-282) summarizes the possible motivations for Principles I to V and VII, but VI and VIII go unmentioned in this regard. Nevertheless, I use Labov’s *Peripherality Hypothesis* as the theoretical framework for the vowel movements of the Canadian Shift, as Clarke, Elms and Youssef (1995) have done when they first proposed its existence (cf. Section 2.3).

In terms of the neogrammarian dogma, mentioned in the beginning of this subsection, Labov invested much effort to sort out the controversy between the postulates of neogrammarians and more recently lexical diffusionists. Neogrammarians proposed (lexically) regular sound change via sound laws – linguistic conditioning factors – that operate without exception, meaning that whole classes of sounds – phonemes – are affected as the change proceeds. Sound change is thus only conditioned by phonetic environment, and not by the need to convey information (Labov 1994: 421). In lexical diffusion, sound change advances word by word through a language, but the change does not affect a language’s whole lexicon at once (cf. Labov 1994: 472). Neogrammarians ascribed lexical properties that affect sound change to analogy and dialect borrowing, whereas lexical diffusionists consider lexical effects to be the vehicle of sound change (cf. Labov 1994: 422, 438; Thomas 2011: 285-286). Analogical change, “which involves conceptual relations” that are not phonetic in nature, and dialect borrowing, which involves “social relations of relative prestige” that are not phonetic, are the two types of exceptions to the neogrammarian exceptionlessness (Labov 1994: 422-423). In an attempt to resolve the controversy, Labov reviewed numerous studies and examples, providing evidence “that both Neogrammarian regularity and lexical diffusion exist” (Labov 1994: 438), although the examples of neogrammarian sound change outnumbered those of lexical diffusion (Labov 1994: 471). These examples were also derived from research conducted by historical linguists, proponents of lexical diffusion, in the field of dialect geography (Labov 1994: 472). The main case he investigated was the historical split of /æ/ into raised and tensed /æh/ and lax /æ/ in New York City and Philadelphia (cf. e.g. Labov 1994: 503), which I will outline in some more detail in Subsection 5.5.2.1. He concluded that lexical diffusion is not the basic mechanism of sound change, but it appears alongside neogrammarian change (Labov 1994: 501). His proposition for a solution of how both the neogrammarians and the lexical diffusionists can be right is as follows:

I do not propose to resolve the original confrontation into a simple dichotomy – that here words change, there sounds change. I have exhibited two polar types, and have analyzed the clusters of properties that created these types. The whole array of sound changes will undoubtedly show many intermediate combinations of these properties of discreteness, abstractness, grammatical conditioning, and social
conditioning. [...] We will find some discontinuous shifts that are regular [...] We will also no doubt find some lexical irregularities within subsystems [...] (Labov 1994: 542).

Bearing these reservations in mind, he roughly equated neogrammian sound change with (regular) change from below and lexical diffusion with (abrupt) change from above and tentatively continued: “[...] we would predict that the realms of regular sound change and lexical diffusion would display complementary distribution [...]” (Labov 1994: 543). While I understand ‘complementary distribution’ to be seemingly contradictory to ‘polar types with many intermediate combinations’, he postulated to employ research strategies that allow investigation of “[...] the full range of properties that determine the transition from one phonetic state to another [...]”, rather than asking whether the neogrammarians or lexical diffusionists “were right” (Labov 1994: 543). Particularly since Bybee (2002) provided empirical evidence that both types of sound change can be embedded in one another (also cf. Oliveira 1991), this is precisely what I intend to do regarding the interpretation of my findings, especially in terms of KIT lowering (cf. Chapter 5).

2.3 The Canadian Shift

This section serves several purposes: 1) to continue the discussion of the homogeneity of Canada’s middle-class English I began in Section 2.1; 2) to establish a comprehensive overview of the vowels involved in the chain shift referred to as the Canadian Shift; 3) to ascertain whether the chain shift relationship between the front lax vowels can be maintained throughout the studies that have investigated this subsystem in Canadian English; and 4) to establish which variants of the social and linguistic variables that have been investigated influence the variation in the vowels. These purposes will be served by focusing on three studies in particular: the one that first proposed the existence of the Canadian Shift (Clarke, Elms and Youssef 1995), detailed in Subsection 2.3.1; the one that questioned the chain-shift nature of the vowel movements (Boberg 2005), detailed in Subsection 2.3.2; and the Atlas of North American English (ANAE or the atlas; Labov, Ash and Boberg 2006b), which used the most elaborate methodology and indirectly helped to resolve some of the controversy between the former two studies (cf. Subsection 2.3.3). As I will outline in Chapter 4, my study will follow the methodology employed in the atlas and by Boberg (2005, 2008b, 2010), which is in large part identical to that of the atlas. The overarching theme of this section, including the subsections, is the diachronic development of research on the Canadian Shift. I will thus discuss the reactions to the studies mentioned above in Subsection 2.3.4 and relatively recent extensions to the knowledge concerning the Canadian Shift in Subsection 2.3.5, particularly regarding ethnicity and style. The multitude of studies on the Canadian Shift does not provide substantially different evidence from that provided in the three core studies in terms of the purposes outlined above. For such studies, I will restrict my discussion to methodological
2.3. The Canadian Shift

concerns as well as the results only in order to outline the geographical distribution of the sound change and to address the notion of homogeneity.

2.3.1 The Original Proposal

In his *The Three Dialects of English*, Labov (1991: 33) defined the third dialect of English by the absence of any vowel shift affecting the entire system such as the Northern Cities Shift and the Southern Shift in the United States. The geographical distribution of the third dialect included all areas in North America that exhibited, among other features, the merger of the low back vowels LOT and THOUGHT as well as raising and fronting of TRAP before nasals, while maintaining stability elsewhere. These areas were the western U.S., including a transition zone towards the east, Eastern New England and Canada (1991: 31). The other features involved in this third dialect were the fronting of GOOSE and of GOAT (Labov 1991: 30).

At the same time, Esling (1991) published his study conducted in Vancouver, British Columbia, Canada, with data collected from 1979 to 1981. He presented his results for ten phonemes, yielding 1200 vowel tokens uttered in a reading passage by 32 females and 32 males between the ages of 16 and 35 (born between 1963 and 1946, respectively). These 64 respondents represented four social classes, middle working (MWC), upper working (UWC), lower middle (LMC) and middle middle class (MMC), each represented by eight respondents (1991: 124). With regard to gender, he found that all ten vowels (/i, ɪ, e, ɛ, æ, a, o, ə, u, ʊ/) were significantly differentiated by social class for females, revealing an acoustic shift in vowel quality. For the two social classes at the interior of the social hierarchy, UWC and LMC, he summarized that they were most difficult to differentiate in terms of vowel realization, but all vowels separated these two social groups, except for (merged) LOT-THOUGHT, KIT, TRAP and STRUT (1991: 125). I understand this finding as a strong indicator that those two groups typically advancing sound changes are converging their realizations of these four vowels in the course of an acoustic change. Unfortunately, Esling did not provide more details, except that the “majority of shifting to signal social status appears in half-open to open and in front to central vowels: /i ɛ æ æ/” (1991: 126), i.e. the lax vowel subsystem of young urban middle-class females born and raised in Vancouver.

Based on Esling’s data, Esling and Warkentyne published their rather methodological work exclusively concerning the retraction of TRAP two years later (1993). They found that ash was most often retracted among and appears to have originated in individuals of the highest social status, and that women led men by one generation. In addition, middle middle-class females were considerably more advanced than middle working-class females in the oldest age group. As mentioned above, their data was based on the rather formal reading-passage style (1993: 242).
ongoing vowel changes that Boberg and Labov used to establish the dialect region of mid to low central to low back (1995: 212; Principles II and III; cf. Clarke, Elms and Strut established by [lower and more central position), is triggered in the direction of central open /a/, except in pre-nasal position. With variably rounded to [ə], which is merged vowel remains a tense peripheral vowel (1994: 280; Labov, Ash and Boberg 2006b: 18-19). It creates the conditions for a pull-merger (Clarke, Elms and Youssef 1995: 211-212). The conceptualization of the dialects of English (cf. Clarke, Elms and Youssef 1995: 210). In addition to results of earlier studies, their evidence was based on a sample of eight male and eight female middle-class speakers of ages ranging from 22 to 26 (except for one 33-year-old male) who read a list of words from which the authors extracted a sufficient amount of 1900 tokens representing the lax vowels /ɪ, ɛ, æ, a, ʌ/ in monosyllabic words. 14 speakers were from southern Ontario, one from Vancouver and one from Edmonton who had lived in Toronto for six years. The tokens were impressionistically coded for shifting or non-shifting, double-checked by the three authors and supplemented with acoustic analysis (Clarke, Elms and Youssef 1995: 211-212). The conceptualization of the vowel chain shift they observed is schematically represented in Figure 2.4.

In response to Labov’s (1991) paper, Clarke, Elms and Youssef (1995) offered Canadian evidence that went counter to Labov’s notion of a relatively homogenous third dialect area, defined predominantly by the absence of a major vowel chain shift when contrasted with the ‘first’ and ‘second’ dialects of English (cf. Clarke, Elms and Youssef 1995: 210). In addition to results of earlier studies, their evidence was based on a sample of eight male and eight female middle-class speakers of ages ranging from 22 to 26 (except for one 33-year-old male) who read a list of words from which the authors extracted a sufficient amount of 1900 tokens representing the lax vowels /ɪ, ɛ, æ, a, ʌ/ in monosyllabic words. 14 speakers were from southern Ontario, one from Vancouver and one from Edmonton who had lived in Toronto for six years. The tokens were impressionistically coded for shifting or non-shifting, double-checked by the three authors and supplemented with acoustic analysis (Clarke, Elms and Youssef 1995: 211-212). The conceptualization of the vowel chain shift they observed is schematically represented in Figure 2.4.

As the figure illustrates, Clarke, Elms and Youssef suggested that the LOT-THOUGHT merger (cot/caught) serves as the pivot of the Canadian Shift (Principle IV; cf. Labov 1994: 280; Labov, Ash and Boberg 2006b: 18-19). It creates the conditions for a pull-chain shift, because the merged vowel remains a tense peripheral vowel [a], which is variably rounded to [b], and as a consequence the lowering and retraction of TRAP (/æ/) is triggered in the direction of central open /a/, except in pre-nasal position. With TRAP in a lower and more central position, DRESS (/ɛ/) simultaneously lowers and retracts to the position formerly occupied by TRAP, and subsequently KIT (/i/) follows the trajectory established by DRESS until it has lowered to the slot formerly occupied by DRESS. The STRUT vowel tends to lower and/or centralize in their data to a position anywhere from mid to low central to low back (1995: 212; Principles II and III; cf. Clarke, Elms and Youssef 1995: 222; Labov 1994: 116, 176, 200, 262; Labov, Ash and Boberg 2006b: 16). In addition to the lax vowel movements, Clarke, Elms and Youssef also noted those other ongoing vowel changes that Boberg and Labov used to establish the dialect region of
2.3. The Canadian Shift

Canada (cf. Section 2.1), such as fronting of GOOSE, GOAT, FOOT and MOUTH (1995: 213).

Their original hypothesis of similar linguistic constraints on all the vowel shifts involved proved as too simplistic. They included three linguistic predictor variables in their multivariate analysis: following manner and place of articulation and following glottal state. The strongest predictor was manner, followed by place and glottal state. Manner of articulation yielded significant findings for DRESS, TRAP and STRUT, with nasals inhibiting the retraction of TRAP and centralization of STRUT (but promoting lowering of DRESS), fricatives favoring all of the vowel shifts (DRESS in particular) and laterals showing no consistent effect of lowering (Clarke, Elms and Youssef 1995: 214). Following place of articulation was only significant for STRUT, but no generalizations could be made. Glottal state was only significant for DRESS, indicating that following voiceless consonants promote its lowering (1995: 216).

For extra-linguistic factors, Clarke, Elms and Youssef could only provide results for gender, due to the nature of their data. Females clearly led the advancement of the sound change, as the difference between genders was significant for all vowels but KIT; gender was additionally more significant than any linguistic predictor examined (1995: 216). On the formality level, they observed that the shift was more advanced in spontaneous speech, but did not have enough data to substantiate this claim (cf. Clarke, Elms and Youssef 1995: 211). In terms of the other non-linguistic factors, age, geographical distribution and social class, Clarke, Elms and Youssef had to rely on findings by others. They quote, for instance, Esling and Warkentyne (1993: 240-242) and Woods (1993: 170-171) to place their gender results in a larger context and to establish an apparent time trajectory concerning the low-back merger with its variable state of lip-rounding and the retraction of TRAP. As mentioned above, the results of Esling and Warkentyne (1993) are based on the data in Esling (1991); those of Woods (1993) are based on the data collected for his doctoral dissertation in 1979. Woods (1979) reported both vowel movements, i.e. the LOT-THOUGHT merger and the retraction of TRAP, for 100 Canadian-English speakers from Ottawa, Ontario. He found that young females exhibited the greatest amount of TRAP retraction of all speakers in his data set. At the same time, they were also most advanced concerning the rounding of the merged low back vowels LOT and THOUGHT (Woods 1979: 152-153; also cf. Woods 1991: 142, 144). For the low-back merger, he found the highest frequencies of lip rounding in the two most formal styles, minimal pair and word-list style (Woods 1991: 142), and an almost categorical use of the rounded variant in minimal pair style for middle-class speakers (near 100%), as opposed to 20% to 40% in free speech (cf. Woods 1991: 144). Although Esling (1991) mentioned the front lax vowels in particular when talking about an acoustic shift in vowel quality, both studies Esling and Warkentyne (1993) and Woods (1991, 1993) concentrated on the retraction of TRAP only, which is not necessarily connected to a front lax vowel chain shift (cf. e.g. Hagiwara 2006: 133). As mentioned above, Clarke, Elms and Youssef’s data did not allow
for establishing an apparent-time trajectory for the Canadian Shift, since they did not have respondents in their 40s, 50s, and 60s and only one respondent in his 30s. They supplemented their original analysis with six middle- to upper middle-class Torontonians in their 50s and 60s and found that their lax vowel subsystem was much more conservative. Having stated that, they cautioned that their sample was too small to allow for such a generalization (1995: 217, 219).

As far as geographical distribution and social class are concerned, Clarke, Elms and Youssef (1995: 220) suggested that the shift covered the regions from Canada’s west coast to eastern Ontario and that it was a middle-class phenomenon, emphasizing Esling and Warkentyne’s (1993) findings of TRAP being more retracted among middle-class speakers than working-class speakers. As outlined above, Esling and Warkentyne (1993: 242) indeed found middle-class females to be more advanced than working-class females, but only among the oldest speakers. In the young cohort, upper-class females were most advanced. Clarke, Elms and Youssef (1995: 220) continue: “The retraction of /æ/ is also an incipient change in the English of St. John’s, Newfoundland; there, it is most prevalent in the speech of women of the uppermost social class investigated (Clarke 1991: 116-118).” Recall that lowering and retraction of TRAP may not necessarily be connected to the systematic movement of the front lax vowels TRAP, DRESS and KIT. In essence, the shift proposed by Clarke, Elms and Youssef seems not unlike other North American vowel shifts (1995: 220).

In the immediately following years, the lax vowel lowering as a part of the Canadian Shift has been further investigated, predominantly in undergraduate and graduate students papers such as De Decker and Mackenzie (2000), Hirayama (2000), De Decker (2002) and Lawrance (2002). De Decker and Mackenzie, for instance, impressionistically investigated five male and eight female speakers aged 10 to 59 years from predominantly Metro Toronto and also from Eastern Ontario, North York and Quebec. They included two styles, formal (reading/word list) and informal (interview), but only collected ten interviews and twelve reading passages/word lists, yielding a total of 3439 tokens of /æ/ and /ɛ/ (2000: 2). The three social predictors they investigated, age, gender and style, were complemented by three linguistic predictors, place of articulation, manner of articulation and syllable type (checked/closed versus free/open). Their results mainly confirmed those of Clarke, Elms and Youssef (1995), with following manner having the strongest influence on KIT and DRESS lowering. For the former, young males led in interview style and for the latter young females led in interview style. They also found 471 lax KIT vowels (6% lowered) and 516 lax DRESS vowels (21% lowered) in open syllables (De Decker and Mackenzie 2000: 5, Tables 1.2 and 1.3). De Decker (2002) added a micro-level and rural perspective to the Canadian Shift (also cf. Hirayama 2000 for western Canada): Among nine teenage females and one teenage male from a small town near London, Ontario,

Note that some conference papers were presented prior to these publications such as Meechan (1996) and Hoffman (1998, 1999); later D’Arcy (2002) and Hollett (2005) for St. John’s, Newfoundland.
forming a community of practice, De Decker found that those without regular contact with young urban speakers accommodated to a greater extent to the urban speech pattern than those with regular contact to young urban speakers. Lawrance (2002), a graduate student at McGill University at the time, conducted research for her Master’s thesis in small towns in Ontario, as well as Toronto, but her results opposed the front lax vowel movement of kit and dress suggested by Clarke, Elms and Youssef (1995) insofar as they showed retraction rather than lowering among the 27 female respondents of ages 18 to 25. This observation was taken up by Boberg (2005), professor at McGill University, in his peer-reviewed paper on the Canadian Shift in Montreal, Quebec, in which he sought to re-examine the nature of the Canadian Shift (2005: 137).

2.3.2 The First Criticism of the Phonetic Nature of the Shift

While acknowledging its value, Boberg (2005) criticized Clarke, Elms and Youssef’s study on three levels. First, the small range of speakers in terms of age stratification caused a lack of “statistically valid generational differences” (2005: 135). Second, 14 Ontarian speakers did not represent eight million native English-speaking Ontarians; and the two remaining speakers, one from Alberta and one from British Columbia, did not rule out the possibility that this shift was in fact an Ontarian phenomenon (2005: 135). Third, the impressionistic coding of vowel lowering may have produced errors that could be accounted for by acoustic phonetic analysis (2005: 136). This, of course, only holds true if the various sources of errors potentially introduced in acoustic phonetic analysis are avoided (cf. Section 4.4).

Boberg further outlined that preliminary analyses of /ɛ/ of six females and four males aged 17 to 55 from various towns across Ontario, collected for the atlas (Labov, Ash and Boberg 2006b), indicated “no correlation [of] the F1 of /ɛ/ (r = 0.16)” (Boberg 2005: 136). The r value here is a correlation coefficient, Pearson’s r, which was, in this case, derived from a correlation of formant one and age. This positive correlation means: the lower the ages, the lower the formant one values. The first formant indicates tongue height and the second formant tongue advancement: Low first formant values indicate high vowels and high first formant values indicate low vowels; low second formant values indicate back vowels and high second formant values indicate front vowels (cf. Peterson and Barney 1952; Hillenbrand et al. 1995). Since lowering first formant values indicate raising of a vowel, the ten speakers slightly tended to raise dress vowels when they were younger – opposite to what Clarke, Elms and Youssef (1995) proposed, based on an impressionistic analysis.

As I will outline in Subsection 4.6.3.2, Pearson’s r is, however, problematic for two reasons: First, the evaluation of 0.16 as ‘no correlation’ is subjective, because others

---

22 Because the contrast of eight million to 14 seems profound, I would like to add that certainly many fewer Ontarians are middle class and young, although this number is certainly still of seven digits (census data does not provide such details).
consider a correlation between 0.05 and 0.2 to be low (e.g. Gries 2009: 139; also cf. Sullivan and Feinn 2012: 280). A statistical analysis of this correlation would have been more revealing. Second, and more importantly, Pearson’s $r$ requires the data to be normally distributed and outliers to be absent. If the data distribution violates the assumption of normality, Pearson’s $r$ may indicate a higher correlation than a scatterplot of all ages and all formant one means would justify, or it may even yield a falsely negative or falsely positive value. In cases where normality is violated or outliers are present, Kendall’s $\tau$ (tau) is the better choice (cf. Gries 2009: 140, 145, 147).

In essence, assuming that the criticism of lacking representativeness that Boberg (2005) brought forward against Clarke, Elms and Youssef (1995) is justified, both Lawrence’ (2002) thesis and the preliminary analysis of atlas data are subject to the same criticism and indicate that DRESS is retracted rather than lowered. Based on two indicators, a Pearson’s $r$ value of 0.16 and the results of a Master’s thesis, Boberg (2005: 137) observed that “the essential nature of the Canadian Shift [is called] into question”. Despite this seemingly exaggerated criticism, his Montreal study is a valuable contribution to expand the knowledge about the Canadian Shift – at least in his later publications (Boberg 2008b, 2010). Boberg (2005: 138) investigated more than 1000 tokens of the six lax vowels, /i, e, æ, œ, u, o/, produced by 35 native speakers of Montreal English from three ethnicities (9 Irish, 15 Italian and 11 Jewish). The speakers were stratified according to three age groups (13 were born before 1946, 11 between 1946 and 1965 and 11 after 1965) and were skewed towards females (21 females and 14 males). The respondents were recorded in their homes by undergraduate students from McGill University, while they provided demographic data, read a word list and talked about life in Montreal (in this order; cf. Boberg 2005: 138). The 1000 tokens in the analyses were exclusively derived from the word list, which contained 115 common monosyllabic words. Of these, 41 lexical items contained the lax vowels in pre-stop, pre-nasal and pre-lateral position. Although Boberg never mentions the social status that he attributes to his respondents, a table with their education (university yes/no) and occupations, ranging from “repairman for railroad” via “elementary school teacher” to “supervisor at electrical supply business” (and university students; 2005: 139), suggests their middle-class background. The independent variables (predictors) he investigated were social ones only: age, sex, ethnicity and education (Boberg 2005: 141).

His results from a multivariate analysis of covariance (MANCOVA) suggest that age had a significant effect on formants one and two of the six lax vowels, which was independent from sex, ethnicity and education. Within his age groups, only formant one of TRAP and formant two of TRAP, DRESS and KIT were significant, indicating that the two front lax vowels DRESS and KIT did not lower and that STRUT was stable in apparent time. The low back vowels LOT and THOUGHT did also not move significantly in apparent time but they were merged, indicating that the low-back merger was present in Montreal English but was no longer a change in progress (Boberg 2005: 141-143).
2.3. The Canadian Shift

between Ontario and Montreal productions of the vowels involved in the Canadian Shift. Of course, the small size of the Ontario sample means that this conclusion can only be tentative at this point, but it seems likely that a robust regional difference would have appeared even with a small sample.

Clearly, the nature of regional differences in the operation of the Canadian Shift, like the effect of social factors, represents a promising opportunity for future research. Acoustic analyses of the speech of large samples of comparable subjects from all of Canada’s regions should prove particularly valuable in this respect. Labov et al. (forthcoming) offer a first view of the national picture. Based on a limited sample of only a few subjects in each of Canada’s urban regions, they find that the Shift does not operate consistently in Atlantic Canada, but serves as a reliable indicator of Canadian speech in the rest of the country, from Quebec to British Columbia, distinguishing it from the American varieties spoken across the international border. The much larger sample of Montreal’s English-speaking population examined here confirms this view, at least with regard to Quebec.

If Montreal’s participation in the Shift now seems clear, however, the phonetic nature of the Shift remains a puzzle. The multivariate analysis of acoustic data on Canadian English carried out by Labov et al. (forthcoming) reveals a pattern that conforms to both versions of the Shift discussed here: that of Clarke et al., in which the major development of /ɛ/ is a descent towards /æ/ (Figure 1); and that of the present analysis, in which the major development of /ɛ/ is a centralization parallel with the retraction of /æ/ (Figure 4). In other words, Labov et al. found that /ɛ/ is moving diagonally, both down and inward.

Figure 2.5: A schematic of the Canadian Shift in Montreal according to Boberg (2005). Figure taken from Boberg (2005: 149).

though FOOT centralized when age was investigated in isolation, it was stable when sex, ethnicity and education were included. As Boberg interpreted it, this movement of FOOT may be coincidental and not structurally connected to the Canadian Shift (2005: 145).

In terms of the remaining predictor variables he included in the analysis, only sex exerted significant influence on vowel realization, but ethnicity and education did not (Boberg 2005: 146). More specifically, sex influenced formant two of /i/, /æ/ and /u/, but not in the expected pattern: Like De Decker and Mackenzie (2000: 5), Boberg (2005: 147) attested a significant retraction of KIT for males; unlike De Decker and Mackenzie (2000: 5), he found no significant retraction of DRESS for either sex. In the retraction of TRAP and LOT-THOUGHT, females were in the lead. For the latter two vowels he offered a chain-shift explanation: “The further back the articulation of /u/, the more room there is in the low central region for the retraction of /æ/” (Boberg 2005: 147-148). For KIT, he did not provide an explanation and offered differences in “sociosymbolic values” as a possibility but dismissed this possibility worth noting in passing by stating that it should be subject of future research (2005: 147).

Based on these findings, Boberg concluded that a movement in the front lax vowels was active in Montreal English, but its phonetic character was different from the one reported by Clarke, Elms and Youssef (1995). He emphasized that the front lax vowels showed a set of parallel retractions or centralizations, but not a vowel chain shift. In this set, the retraction of DRESS was the most active component, which indicated that it was “more marked” than the other changes (Boberg 2005: 145). This relationship was shown in Figure 2.5. The vowels involved in a change that is similar to the Canadian Shift as proposed by Clarke, Elms and Youssef (1995) is indicated by arrows. The vowels /u, ɔ, ʌ,
Looking more closely at the arrow Boberg drew for *dress* retraction, we can see that the originally proposed “essential nature of the Canadian Shift” is not called into question despite earlier claims to the contrary (Boberg 2005: 137). This interpretation of the results of his Montreal study was instigated by his discussion of the results in relation to those of Clarke, Elms and Youssef (1995) and those of ANAE (Labov, Ash and Boberg 2006b), which at the time of Boberg’s publication was yet to appear, and this interpretation was substantiated by his later publications (Boberg 2008b, 2009, 2010, 2011). Since Boberg was a member of the research team of ANAE, he knew what the atlas would essentially propose about the Canadian Shift, namely exactly what the *Peripherality Hypothesis* suggests about lax vowels: they move downward along the non-peripheral track (Principle II; cf. Section 2.2). In this theory, the vowel space is conceptualized as a triangle and not a rectangle, implying that a movement along the track is essentially a diagonal one. In terms of the formant values, this means that formant one increases and formant two decreases in value simultaneously. In case of Boberg’s Montreal data, he should have found a significant effect of age on formant one of (at least) *dress* in addition to that on formant two, according to the theory. I claim that this is precisely the reason for Boberg’s dubious claim regarding *dress* retraction: “It must be admitted that the effect of generational group on the F1 or height of /ɛ/ is fairly close to being significant, at *p* = .126, [...]” (2005: 143). With such a *p*-value, the effect of age is not close to significance, it is simply what it is: insignificant. In contrast, his F2 of /ʊ/ was much closer to significance, at *p* = 0.099, but goes unmentioned in this regard (cf. Boberg 2005: Table 5 on page 143). In his discussion, he claimed that

> [t]he multivariate analysis of acoustic data on Canadian English carried out by Labov et al. (forthcoming) reveals a pattern that conforms to both versions of the Shift discussed here: that of Clarke et al., in which the major development of /ɛ/ is a descent towards /æ/ (Figure 1); and that of the present analysis, in which the the major development of /ɛ/ is a centralization parallel with the retraction of /æ/ (Figure 4). (Boberg 2005: 149)

As I will outline below, the atlas did not conform to both views of the shift, it only conformed to that of Clarke, Elms and Youssef (1995) – at least in part. For these two vowels, Labov, Ash and Boberg (2006b: 130) defined the Canadian Shift by TRAP retraction and *dress* lowering only; they attested no parallel F2 movement whatsoever.

Boberg (2005: 150) continued that the diagonal movement of *dress* in the vowel space was “particularly difficult to reconcile” with the correlation coefficient, *r* = 0.16, between age and F1 of *dress* in the Ontario data from the atlas (i.e. no lowering of *dress* in apparent time). He assumed that this result was due to regional differences, a small sample in each city and the incomparability of the vowel tokens by each speaker due to the fact that the data were derived from spontaneous speech (2005: 150). He did not
consider the possibility that the distribution of the data may not meet the assumptions the correlation coefficient is based on.

Although Boberg (2005) criticized Clarke, Elms and Youssef’s (1995) study for having employed impressionistic methods only, he admitted that Clarke, Elms and Youssef’s (1995) and ANAE’s (2006b) view of the Canadian Shift in terms of DRESS lowering (and simultaneous retraction)

gains further support from anecdotal, impressionistic observations of the speech of young Canadians. Among young Canadian women in particular, the pronunciation of /ɛ/ is sometimes low enough to produce potential confusion with /æ/, at least when taken out of context, as when left and bet sound somewhat like laughed and bat. It is not clear why this development is not reflected in the data presented here. If it is not simply a regional difference, the discrepancy may result from characteristics of the sample on which the present study is based, such as the high proportion of certain ethnic groups [...]. (Boberg 2005: 150)

That he did not find via acoustic means what he found via impressionistic methods is substantiated in the empirical data that he would present in his later publications. In a paper in 2009, for instance, he pointed out that the British-origin Montrealers participated most in recent Canadian English (CE) sound changes, Jewish Montrealers participated to a lesser degree, and Italian Montrealers did not participate at all (Boberg 2009: n.p.). As I discussed above, his Montreal study on the Canadian Shift consists of 9 Irish, 15 Italian and 11 Jewish respondents (cf. Boberg 2005: 138).

By outlining Boberg’s statements in his discussion of his findings, I intend to make two things apparent: First, “the phonetic nature of the Canadian Shift” does not remain “a puzzle” (Boberg 2005: 149); and second, Boberg falsely shifted his focus from the underlying phonetic principles of sound change to the terminology Clarke, Elms and Youssef (1995) used to describe the front lax vowel movements that form the Canadian Shift. They referred to the movement of DRESS (and KIT) as lowering, i.e. an increase in formant one values only, but meant that the vowel(s) move(s) to the position formerly occupied by TRAP (or DRESS) under Principle II (cf. Clarke, Elms and Youssef 1995: 212, 222).

In essence, Boberg (2005) showed that the Canadian Shift is present in formal (word-list) style of Montreal’s middle-class English, that it operates similarly to what Clarke, Elms and Youssef (1995) suggested and that it is predominantly constrained by age and to a lesser degree by gender, when the vowels occur in pre-stop, pre-nasal and pre-lateral position. Boberg did neither investigate linguistic environment, nor stylistic profile. In terms of the vowels that move as part of the Canadian Shift, Boberg (2005) listed TRAP, DRESS and KIT but not STRUT; the centralization of FOOT is an independent movement.

In the same year that Boberg’s Montreal study appeared, D’Arcy published her peer-reviewed article which sought to impressionistically investigate /æ-/fronting and “/æ/ Retraction and Lowering (RL)” among young middle-class females in St. John’s, Newfoundland (2005: 330). Although she emphasized that /æ/ RL, as reported by Esling
2. English-speaking Canada and its Vowel Shifts

(1991) and Esling and Warkentyne (1993) for Vancouver English, was part of the Canadian Shift (D’Arcy 2005: 328), Esling and Warkentyne (1993) reported it independently as a feature of Vancouver English (also cf. Hagiwara 2006: 133). D’Arcy collected her data in 2000, using a friend-of-a-friend approach to sampling which yielded four pre-adolescent (8 to 11 years of age) and four adolescent (16 and 17 years of age) respondents of local parentage as well as four pre-adolescent and four adolescent respondents of non-local parentage (2005: 332). She collected 801 tokens of (æ) RL from a word list and spontaneous speech in sociolinguistic interviews, and identified three variants of (æ): raised [æ], unmarked [æ] and retracted [a] (2005: 333). In addition to the social factors (parentage and style), she investigated the effects of following linguistic environment, namely manner and place of articulation as well as glottal state. Her results showed that the frequency of the innovative mainland CE variant is increasing, progressing from local pre-adolescent (3%), via local adolescent (9%) and non-local pre-adolescent (14%) to non-local adolescent (21%). In terms of the linguistic variables, place of articulation generally exerted the strongest influence, followed by glottal state. Unlike in Clarke, Elms and Youssef’s (1995) study, manner of articulation was not selected as significant by the stepwise regression analysis. Voiced segments favored the innovative variant [a] in the speech community, and place of articulation exhibited a behavior that phonetically makes perfect sense: velars and alveolars favored the innovation, labiodentals were intermediate and interdentals and bilabials disfavored it (2005: 339). Following voiced velars in particular showed the highest frequency for the use of the innovative CE variant in St. John’s, whereas following nasals inhibited it categorically (2005: 338). A further breakdown of the results according to the two groups of parentage revealed that the linguistic constraints of place of articulation and glottal state only held for the young females with mainland Canadian parents, but not for those with parents born and raised in St. John’s. For the latter, style and age were significant, but not the linguistic factors. In particular, casual style and being 16 to 17 years of age favored the use of (æ) RL, whereas word-list style and being pre-adolescent disfavored it (D’Arcy 2005: 339-340).

2.3.3 The Stance of the Atlas of North American English

Unlike the studies conducted by Boberg (2005) and Clarke, Elms and Youssef (1995), ANAE’s (Labov, Ash and Boberg 2006b) goal was to describe and outline the vowel chain shifts currently underway in North American varieties of English that serve as defining criteria for cross-continental dialect regions. Between 1991 and 2000, ANAE’s research team recorded 805 respondents from North American cities both via a telephone survey (Telsur) and face-to-face sociolinguistic interviews in four styles: elicited minimal pairs, semantic differential tasks and word lists as well as spontaneous speech (2006b: 21, 28, 36-37). The sampling strategy was to represent the largest possible population, with special focus on those participants that are typically most advanced with regard
to sound change. North America was thus categorized according to the concentration of the population, ranging from central cities via urban centers to zones of influence – a set of counties with more than 100 households (2006b: 21). In order to locate speakers representative of each sampling site, local telephone directories were searched for names associated with the most prominent national ancestry groups (2006b: 22). This method resulted in a skewed sample in which the “Middle Middle Class [sic] is over-represented” and the “upper working class [sic] is under-represented”, because working-class members are less often listed in such directories (2006b: 23). A similar bias was detected for national ancestral groups: German-ancestry participants constituted the largest group in most parts of the United States, Scots-Irish in Canada (2006b: 24).

The social status of the respondents was assessed using Census data on income and education as well as data on ratings of occupational titles (based on the results of the National Opinion Research Council, NORC, in 1947 and 1992). The three levels were used to calculate Socioeconomic Index (SEI) scores. This method posed several problems: for one, matching a speaker’s occupation to one of the 503 occupations in the NORC/Census list; moreover, students cannot properly be assigned an SEI on the basis of their breadwinner’s data; and finally, data elicited from speakers was often insufficient (e.g. I work in an office). Thus, some speakers, such as the 35-year-old male respondent from St. John’s, had not been assigned an SEI (cf. Labov et al. 2006: n.p.; also cf. Subsection 5.2.1).

In terms of age and gender of the respondents, ANAE’s sample is skewed towards “early adult years” (2006b: 27), and females outnumber males by a ratio of 1.7:1. Most of the respondents were aged from 30 to 39 (182) and 40 to 49 (157) years. The early adult years (20 to 29) were represented with 119 speakers; the remaining decades – the speakers ranged from 12 to 89 years of age – were represented by fewer than 90 speakers each. In the 20-to-29 decade, females outnumber males by 1.9:1. For Canada, 39 speakers participated, with 10 speakers in their 20s and 14 speakers in their 30s (2006b: 28).

762 of the 805 speakers were ultimately included in the impressionistic vowel coding procedure. 439 of them were additionally analyzed acoustically, including 33 speakers from Canada (2006b: 36, 220). On average, 305 vowel tokens were extracted per speaker, yielding a total of 134,000 vowels for multivariate regression analysis (2006b: 36-37) in formal and casual speech style of predominantly young urban females from the upper middle classes. Particularly relating to the Canadian Shift vowels, tokens in liquid or glide and pre-nasal context as well as TRAP before /g/ were excluded (2006b: 77).

As outlined in Section 2.1, Labov, Ash and Boberg (2006b: 130) defined Canada as a dialect region predominantly by virtue of the Canadian Shift. Their definition of the shift was based on quantitative measures: the shift is triggered by the low-back merger, with LOT-THOUGHT (/o/) in low back position as [ɔ]. Since the backing of /o/, defined by a mean formant two value of less than 1275 Hz, extends the margin of security between TRAP (/æ/) and LOT-THOUGHT, the low-back merger triggers the retraction of TRAP,
defined by a mean formant two value of less than 1825 Hz, which in turn is followed by the lowering and retraction of DRESS (/e/), defined by a mean formant one value of more than 660 Hz (i.e. lowering of DRESS; 2006b: 130) – a pull-chain shift. This view of the shift, shown in Figure 2.6, excluded the Atlantic provinces of Canada (2006b: 130) and also differed from the two other versions of the shift, outlined by Boberg (2005) and Clarke, Elms and Youssef (1995), respectively: According to Labov, Ash and Boberg (2006b: 219), the shift was found in the following Canadian cities, proceeding from west to east: Vancouver, British Columbia (three speakers); Edmonton and Calgary in Alberta (two speakers each); Saskatoon and Regina in Saskatchewan (one speaker each); Winnipeg, Manitoba (two speakers); Thunder Bay, Sault Ste. Marie, Arnprior, Ottawa (one speaker each), London, Windsor (two speakers each) and Toronto (four speakers) in Ontario; and Montreal, Quebec (two speakers). Boberg’s (2005) observations for the latter were summarized with a brief adverbial clause: “[He] has observed the Canadian Shift in progress in Montreal, though with more retraction than lowering of /e/” (2006b: 220). Furthermore, “[i]t is evident that /e/ is moving backward and downward in apparent time, and /æ/ is moving backward” (2006b: 220). In other words, there was no parallel retraction of DRESS and TRAP in the ANAE data set, as claimed by Boberg (2005).

In response to Clarke, Elms and Youssef (1995), Labov, Ash and Boberg (2006b: 220) acknowledged and confirmed the underlying principles or phonetic effects of the Canadian Shift that they proposed, but emphasized that “[n]o shift of /i/ is indicated in the ANAE data for Canada”. In a footnote (Fn. 3 on page 220), they confirmed Boberg’s findings for Montreal in that the ANAE data did not replicate the centralization of STRUT as part of the Canadian Shift, as originally proposed by Clarke, Elms and Youssef (1995). In terms of the regional distribution of the vowel movements, Labov, Ash and Boberg’s findings supported those reported by Clarke, Elms and Youssef (1995: 211) for the 14 speakers from Ontario (mostly Toronto, some from London), the one from Edmonton, and the one from Vancouver, and those reported by Boberg (2005) for the 35 speakers from Montreal. They did not support those reported by D’Arcy (2005: 338-340) for the

23 This list is to be understood in the context of the studies discussed up to this point and the studies that I will outline below.
16 speakers from St. John’s, Newfoundland. Thus, Labov, Ash and Boberg (2006b) were the first to provide empirical evidence that Canada’s upper middle class behaves linguistically homogeneous in terms of the Canadian Shift – at least between Vancouver and Montreal.

Among the linguistic and non-linguistic variables included in the multivariate regression analysis of the Canadian Shift vowels were: preceding and following manner as well as place of articulation, following glottal state, number of following syllables, age, city size, education and gender, but style was excluded from the analyses. Given the nature of the sampling procedure I outlined above, most of the 3021 vowel tokens analyzed were most likely extracted from casual speech. These factors accounted for 30% to 50% “of the variance for the sound change in progress” (2006b: 220), referring to the $R^2$ value of the statistical model. The details of the results of the statistical assessment were only revealed regarding age – most likely in order to emphasize the apparent time trajectory of the sound change. Despite their claim that KIT was stable in apparent time, the table including the regression coefficients for age and formant one of KIT showed that there was indeed a significant change in the vowel height in apparent time. In unexpected resemblance to the Pearson’s $r$ value Boberg reported for age and DRESS height in the ANAE data for Ontarian speakers, the age coefficient for KIT height among all Canadian speakers was positive. As I outlined above, this means that the younger speakers are (low age), the more raised are their KIT vowels (low F1) – opposite to what is expected to happen to the quality of KIT in the course of the Canadian Shift. In addition, this shift of /i/ was significant, at $p = 0.04$ (2006b: 221, Table 15.1). In terms of the other non-linguistic variables, these findings are most likely representative of urban females from an upper middle class (Middle Middle Class) background.

The linguistic factors Labov, Ash and Boberg (2006b) discussed were limited to TRAP before nasals (/æN/) and /g/ (/æg/) and the following environments of the low back vowels. For the former, they stressed that /æg/ almost showed a merger with the FACE vowel (/ey/ or /eI/) and that some raising of TRAP occurred before nasals and /d/. Canadian respondents, however, showed regional differentiation with regard to the raising of TRAP before nasals. In general, raising in this environment was not as pronounced as in U.S. regions with a split short-a system (2006b: 221, 223). For the latter, the results of the acoustic measurements for vowel production were triangulated with data on perception, based on minimal pairs or rhymes (e.g. hot and caught or Don and dawn). Merged LOT-THOUGHT before nasals (Don-dawn) behaved uniformly across Canada, including the Atlantic Provinces, and was considerably more advanced than in other environments. Before /t/ (hot and caught), the merger was not uniformly perceived to be present. Eight Canadian respondents were ‘close’ in production or perception, and two stated “they did not rhyme even though they said them as rhymes” (Labov, Ash and Boberg 2006b: 217).
2.3.4 The Reactions to the First Decade of Research

In the same year as Labov, Ash and Boberg, Hagiwara (2006) published his acoustic phonetic study on 15 contrastive English vowel phonemes in Winnipeg, Manitoba, which was part of a much larger project. These vowels were placed in real monosyllabic lexical items of the form $hVd$ and $hVt$ in a word list and read by five females and five males aged 18 to 25 who were born and raised in Winnipeg (2006: 128-129). He emphasized that he did not control for ethnic, geographic or cultural (social) factors possibly affecting the speech of his ten respondents (2006: 128). The exact amount of vowel tokens for his analysis seemed to be limited to a total of 300, although he did not mention this explicitly (cf. Hagiwara 2006: 129). He consequently cautioned that “the data from the ten speakers under discussion are not enough to provide conclusive results” (2006: 129). Among his tokens, some represented the lax vowel phonemes that form the Canadian Shift, which Hagiwara plotted in an F1xF2 plane together with the formant values reported by Peterson and Barney (1952) for General American English. Based on a direct comparison of his Winnipeg data with Peterson and Barney’s General American data, he concluded that trap, dress and kit were retracted rather than lowered, because lowering seemed “to be something occurring to front vowels generally, if at all, including tense /i/” (2006: 133). Thus, if the lax vowels lower relative to the height of /i/, the Canadian Shift is absent from his Winnipeg sample (2006: 133; my emphasis). In other words, Hagiwara found retraction and lowering of trap, dress and kit among young Winnipeg respondents.

In her Master’s thesis, Hollett (2007) investigated the Canadian Shift in four old middle-class females from St. John’s recorded in the 1980s, four old females recorded in 2003 and four young females recorded in 2003. Her acoustic analysis was based on word-list data and yielded a total of 1156 vowel tokens (2007: 36). Her results indicated that females aged 55 and older showed the greatest use of trap retraction, but she did not indicate whether she analyzed possibly raised trap in pre-nasal context and contexts before /g/ separately from possibly retracted trap elsewhere (2007: 52-53). As it represents one of the very few acoustic (socio-)phonetic studies on the Canadian Shift in St. John’s, I will discuss her study in further detail in Subsection 3.2.6 and in Section 4.1.

A year later, Boberg published his results of the *Phonetics of Canadian English* (PCE) study, conducted at McGill University, for the first time (2008b: 132). Like ANAE has done predominantly for the regional differences in U.S. English, Boberg’s intent was to outline a more detailed view of regional differences of Canadian English on a national scale. He emphasized the differences between PCE and ANAE as follows: ANAE investigated mostly spontaneous speech data from 33 Canadian speakers covering a broad social range, whereas PCE investigated word-list data from 86 speakers, comprising young, university-educated speakers (2008b: 129). In particular, Boberg criticized that the sample of Canadian speakers in the atlas was too small considering its social and regional
2.3. The Canadian Shift

diversity, so that differences between two locations that appeared to be regional could have been social as well. In addition, the focus on spontaneous speech limited the control over kind and frequency of allophonic environments in the data of the individual respondents (2008b: 131). PCE was concerned with vowel production “in Standard Canadian English” affected by region, sex and attitudes (2008b: 132), but only the regional effects were reported (2008b: 134). As in ANAE, the sample was skewed, with females outnumbering males by a ratio of 1.5:1. McGill University students conducted sociolinguistic interviews with fellow students from different regions in Canada. The regional breakdown was motivated by the findings of ANAE and consisted of those regions that I outlined in Section 2.1, including 13 Montrealers, 16 respondents from New Brunswick and Nova Scotia and six Newfoundlanders. The participants and their parents grew up entirely in the respective regions they represented, and the first contact to non-local peers was at the age of 18 when they moved to Montreal in order to attend McGill University. Boberg considered them to be representatives of their original local speech community in terms of age and middle-class social status (2008b: 133). His analysis was limited to word-list items only, on which each vowel of interest appeared at least once before /t/, /d/, /n/, /l/ and /r/, and tense vowels appeared additionally in syllable-final position. The acoustic analysis followed the same method as that of ANAE (2008b: 134).

In terms of the lax vowels, /i, e, æ, a, ʌ, o/, Boberg found no effect of region, so that he considered the Canadian Shift to be “[...] proceeding on a nationwide basis”, including the Atlantic Provinces New Brunswick, Nova Scotia and Newfoundland (Boberg 2008b: 138). This result contributes once more to the view of Canada’s middle classes behaving homogeneously concerning the Canadian Shift. Although he did not specify which vowels he included in the shift or how he interpreted the movements of the front lax vowels, he emphasized that /o/ was lowering from eastern to western Canada but “[...] completely independent[ly] of the Canadian Shift” (2008b: 138). As it stands, the study cannot establish an apparent-time trajectory of the shift in the Atlantic Provinces where Labov, Ash and Boberg (2006b) found the shift to be absent. Furthermore, Boberg did not reveal which cities or urban areas his respondents came from. For Newfoundland, he mentioned briefly that at least one respondent was from Gander (2008b: 140) and others were from St. John’s (2008b: 142). In 2011, Gander had a population of 11,000, whereas St. John’s had one of 197,000 (cf. Statistics Canada 2012). In addition to the fact that the rural-urban divide is a substantial factor in Newfoundland’s culture, attitudes and language, Boberg did not detail how his sample was representative in terms of city population versus small-town population, unlike Labov, Ash and Boberg (2006b) did for ANAE. Moreover, Gander as a town owes its existence entirely to the airport built in the late 1930s, with no local population nearby prior to and at that time. During World War II, Gander served predominantly as a base for British, American and mainland Canadian military personnel, with 13 homes for locals and 1000 civilian workers living in barracks. In the 1980s, Gander was an economic and transportation hub (cf. Riggs 1984: 468-470).
Boberg’s respondents from Gander can thus hardly be “taken to be good representatives of the local speech” (2008b: 133), neither for rural Newfoundland nor for urban middle-class Newfoundland.

In the same year, Sadlier-Brown and Tamminga (2008) succeeded in establishing an apparent-time trajectory of the Canadian Shift in Halifax, Nova Scotia, by supplementing Boberg’s (2008b) data with 14 old speakers from their personal social networks or on-campus advertisements in that city (Sadlier-Brown and Tamminga 2008: 1, 5). They compared the Halifax sample with a sample from Vancouver using acoustic phonetic methods in order to find differences between the two cities in terms of the Canadian Shift vowels. The younger speakers were undergraduate students at McGill University from each region and also part of Boberg’s PCE study. In total, Sadlier-Brown and Tamminga (2008: 5) investigated a sample of 26 middle-class participants, 14 from Halifax and 12 from Vancouver. They subdivided each group into old (born 1922-1972) and young (born 1981-1986), with both genders represented approximately evenly in each. The sociolinguistic interviews consisted of three sections: demographic data, a word list representing all English vowels in six environments each, and a conversational component (2008: 6). Vowel tokens in pre-/r/ and pre-/l/ environments were excluded, but not those before nasals. For the definition of the presence of the Canadian Shift in both cities, they used the same quantitative defining criteria that Labov, Ash and Boberg (2006b: 130) had established: F1 of /e/ > 650 Hz, F2 of /æ/ < 1825 Hz and F2 of /oh/ < 1275 Hz (Sadlier-Brown and Tamminga 2008: 7). In addition to outlining the mean formant values per vowel and region in order to show that they met the criteria, they used Pearson’s r values to establish the apparent-time trajectory. They found that the correlations of both formant values with age per region suggested a stage-like pattern of the shift, with TRAP showing the strongest, DRESS intermediate and KIT low correlations. Unlike Boberg (2005) suggested for his Montreal sample, Sadlier-Brown and Tamminga attested lowering and retraction for all three front lax vowels, i.e. they moved diagonally in the vowel space, in apparent time (2008: 9). Unlike Labov, Ash and Boberg (2006b), they thus found all three front lax vowels to be involved in the shift in both regions, confirming the suggestions made by Clarke, Elms and Youssef (1995), except for the centralization of STRUT as part of the shift. Via t-tests, they showed that retraction of TRAP and lowering of DRESS was significantly more advanced in Vancouver, a result that was not unexpected (Sadlier-Brown and Tamminga 2008: 11). Unfortunately, they did not discuss the role of style.

More recent studies concentrated predominantly on Toronto, making Canada’s largest city the best-studied research site regarding the Canadian Shift. According to Roeder and Jarmasz (2009: 3), this is in order because the variety spoken there is most representative of mainstream Canadian English and most innovative. Roeder and Jarmasz (2009: 4) investigated 35 speakers of Toronto English, recorded by Tagliamonte from 2003 to 2006, who had aimed at eliciting the most natural speech possible. Unfortunately, it remains
unknown which social class these speakers belong to. The sample was comprised of 14 male and 19 female speakers, stratified according to three age groups: young (20-39 years of age), middle (40-54) and old (70-85). From these the authors extracted 1800 vowel tokens for acoustic analysis; 75% of these comprise /i, ɛ, æ/ and 25% comprise /i, u, ɔ/, and normalized them using the procedure of Nearey (1977). As I will outline in Subsection 4.5.2, this procedure only normalizes accurately when all the stressed vowels of the vowel inventory are included. For the statistical assessment, vowel tokens in “velar, liquid or glide, and pre-nasal context” were excluded (Roeder and Jarmasz 2009: 4). The only independent (predictor) variable they assessed was age for both formants as the dependent variables (MANOVA); males and females were investigated separately (2009: 4).

They interpreted their results in such a way that they confirmed those of Boberg (2005) and of Hagiwara (2006): Roeder and Jarmasz (2009: 8) found that retraction of TRAP and DRESS was stronger than lowering, and concluded that the Canadian Shift was not a chain shift but a redistribution of the front lax vowels (2009: 9). These findings are based on a comparison of speakers aged 70 and older with speakers aged 55 and younger (cf. Roeder and Jarmasz 2009: 5). The motivation for changing the three original age groups into a binary classification remains unmentioned. Additionally, like Labov, Ash and Boberg (2006b) did with their spontaneous speech data, a continuous modeling of age could have yielded different results or provided more support to their interpretation. The statistical tests did not indicate a shift of KIT between the two age groups (2009: 5), but their discussion suggested that KIT was redistributed as a consequence of the low-back merger, along with TRAP and DRESS (2009: 9).

Based on Boberg’s and Hagiwara’s data, Roeder and Jarmasz (2009: 10) suggested a timeline of the spread of the Canadian Shift from Toronto to Montreal and Winnipeg. Recall that Boberg’s (2005) Montreal sample was skewed towards an ethnic group (Italians) that does not participate in recent changes and that Hagiwara’s (2006) comparison for Winnipeg was based on ‘General American English’ from the 1950s. ANAE has made clear that any notion of ‘General American English’ is relatively meaningless when not specified or defined, due to its regional and social diversity (also cf. Boberg 2010: 124). In addition, Hagiwara used a normalization technique for both data sets that has not been tested in relation to established procedures such as Nearey (1977). Each tested normalization technique has serious drawbacks, so that choosing one over another is motivated by the methodological decision of which drawbacks are tolerable (cf. Subsection 4.5.2). Furthermore, Hagiwara (2006: 129) himself cautioned that his results were not conclusive. In addition to the problems of the studies Roeder and Jarmasz base their interpretation on, their study has methodological issues itself. Since they did not indicate the social status of their respondents, the possibility that they were biased towards a certain social group which behaves in a certain way cannot be ruled out – similarly to Boberg (2005). Likewise, they did not outline how they controlled for the linguistic environments remain-
64  2. English-speaking Canada and its Vowel Shifts

...ing in the analysis, not to mention that they disregarded glottal state as a linguistic factor that might equally have a confounding effect. Although they emphasized that they only included vowels in primary-stressed content words (2009: 4), they did not include the effect(s) of vowel duration – or number of following syllables, as Labov, Ash and Boberg (2006b) did, which can function as a substitute of vowel duration. This could confound the relative strengths of F2 versus F1 movements, i.e. shorter vowels tend to be backer than lower (i.e. central; smaller F2s than greater F1s). Finally, the discussion of whether the front lax vowels retract and/or lower was initiated by Boberg (2005) on the basis of having found no movement in F1 in apparent time and not by a stronger movement in F2.

A year later, Boberg published his book on the English language in Canada, which is based in part on the PCE and the *Phonetics of Montreal English* (PME) data sets (Boberg 2010: 143, 199), but naturally more detailed than his peer-reviewed paper (Boberg 2008b). In terms of the Canadian Shift results based on the PCE study, Boberg emphasized the most important difference to his Montreal article as follows:

The Canadian Shift appears to be initiated by the retraction of /æ/, first described in Vancouver English by Esling and Warkentyne (1993), which is in turn activated by the low-back merger of /o/ and /oh/, a well-established feature of Canadian English listed among the variables of phonemic inventory in Table 3.5. As /æ/ shifts down and back into the low-central space made available by the merger, /i/ and /e/ shift down and inward in parallel fashion. The original formulation of Clarke, Elms and Youssef emphasized lowering of /i/ and /e/ to fill the space created by the retraction of /æ/, a classic chain shift, while a subsequent study of the shift in Montreal English by Boberg (2005a) emphasized parallel retraction, finding the F2 of /e/ to be the variable that showed the strongest movement in apparent time. Labov, Ash and Boberg (2006: 221) and the apparent-time analysis presented here in Chapter 5 (Figure 5.3 and Table 5.12) find significant effects of speaker age on both the F1 and F2 of /e/, suggesting that the shift involves both lowering and retraction of that vowel. (Boberg 2010: 146-147)

Boberg further emphasized that TRAP before nasals and TRAP before /g/ have to be treated as separate analytical categories, because the vowel was raised and fronted in these environments in his PCE data and would thus skew the overall distribution if they were treated as a single category (cf. Boberg 2010: 146). In keeping with the view that the shift involved retracting and lowering of DRESS, he outlined which vowels participated in the shift in his PCE data:

[...] Figure 3.1 shows /i/ in upper-mid position opposite to /ey/; /e/ in lower-mid position near the raised allophones of /æ/; and /æ/ approaching low-central position, [a], well separated from the /ah-o-oh/ double merger [Palm-Lot-Thought] in the low-back quadrant. The mean formant values for these vowels in Table 3.12 all conform to the thresholds established by Labov, Ash and Boberg (2006: 219)
2.3. The Canadian Shift

for the shift: the F1 of /e/ [/e/] is greater than 650 Hz; the F2 of /æ/ is less than 1825 Hz; and the F2 of /o/ [/o/] is less than 1275 Hz.24 (Boberg 2010: 147)

In his fifth chapter, Boberg concentrated on differences between regions, age and sex in Canadian English as well as ethnicity in Montreal English and extended his data set by the PME study and others that he did not further specify. His Montreal article was most likely based on his PME study focusing on ethnic variation in Montreal. Both studies, PCE and PME, were representative of middle-class speakers (2010: 199). Unlike in his (2008b) paper, he outlined the division of the communities the respondents came from based on their relative size in population (for < 100,000 n = 40; for 100,000 - 1,000,000 n = 20; for > 1,000,000 n = 26) and included city size as a variable in his multivariate analysis of co-variance (MANCOVA) along with region and sex (Boberg 2010: 202). The results were the same as in 2008: region did not have an effect on the Canadian Shift vowels, indicating its presence among young middle-class Canadians, when speaking formally, from British Columbia to Newfoundland (2010: 204).

In terms of sex, women were ahead of men in the shift of /e/ and /æ/ in every respect (2010: 210). The fact that kit was neither included in the statistical assessment of the effect of sex on formants one and two for all speakers (2010: 210) nor in the graph (2010: 212) remains uncommented. After having cited other studies, Boberg emphasized that “[t]he convergence from these data from several separate studies leaves little doubt that the Canadian Shift fits the typical model of a sound change led by women” (2010: 212). Ethnicity was only investigated for Montreal (2010: 216). Unlike in his (2005) article, the sample was no longer skewed towards Italian respondents and had increased in size from 35 to 93 participants, of which 29 were British, 30 Italian and 34 Jewish (2010: 217). For the Canadian Shift, Boberg did not find any effect of ethnicity (2010: 218).

In order to investigate the role of age, Boberg (2010: 225) complemented his PCE data with PME and additional data. This so-derived sample yielded young (born after 1965) and old (born before 1965) speakers from Vancouver, Halifax and Montreal. The older participants from the former two cities were those collected by Sadlier-Brown and Tamminga (2008), as Boberg outlined later (2010: 227). His results suggested that age had a significant effect on both the F1 and F2 of the Canadian Shift vowels and co-varied with region (2010: 226; also cf. Boberg 2011: 23).25 With the effect of age in place, he examined the results of tests of between-subjects effects in order to determine which of

---

24 Note that all Canadian speakers had a mean F2 of /æ/ of less than 1725 Hz, so that Labov, Ash and Boberg (2006b: 220) refined this threshold from 1825 Hz (2006b: 219) to 1725 Hz for Canada. Boberg’s “interregional mean of 86 participants from eight Canadian regions” is 1727 Hz with a standard deviation of 97 Hz (2010: 145). Labov, Ash and Boberg’s thresholds are to be understood with caution as they vary from page to page. For instance, the threshold of /e/ is an F1 of greater than 650 Hz on page 219 and greater than 660 Hz on page 130.

25 Note that Boberg (2011) is an article in the University of Pennsylvania Working Papers in Linguistics. It is primarily a methodological paper, introducing an index of phonetic innovation in CE which is based on the (reversed) distance between fronted GOOSE and retracted TRAP (F2 [æ] – F2 [uw]; 2011: 21-22), but also repeats some of the findings in Boberg’s (2010) book.
the phonetic measures were significantly affected by age (2010: 225-226) and interpreted them as follows:

As shown in the first seven lines of Table 5.12, age is correlated with the F1 of /i/ and with the F1 and F2 of /e/ and /æ/ in ways that are consistent with the established view of the shift: all three vowels are moving downward in apparent time, while /e/ and /æ/ are also retracting, creating diagonal trajectories. That /e/ is moving both down and inward suggests that the systematic relation among the individual developments of the front vowels involves parallel shifts rather than a classic chain shift, in which each vowel moves into the space vacated by the one in front of it. There was no age effect for /o/, suggesting that the phonetic quality of the low-back merger acts as a diachronically stable initiating condition for the lowering and retraction of the front vowels. Table 5.12 indicates that the shift also involves co-variation between age and region. This is true only for the F1 of /e/, which shows a more dramatic increase in Montreal than in Vancouver. However, this difference reflects lower F1 values for older Montreal speakers rather than higher values for younger Montreal speakers: there is virtually no regional difference in the mean F1 of /e/ for younger participants. (Boberg 2010: 227)

In summary, Boberg claims that the Canadian Shift is not a chain shift, because KIT lowers but DRESS lowers and retracts (similarly to TRAP). Thus, the Canadian Shift consists of parallel shifts. There are two problems with this interpretation: First, Boberg did not include formant two of KIT in his statistical analysis, or he merely did not include it in Table 5.12 (Boberg 2010: 226). The implicated result for formant two of KIT, namely that “age is [not] correlated” with it, seems highly doubtful (Boberg 2010: 227), because second, Sadlier-Brown and Tamminga (2008: 9) did find a correlation of birth year (age) and both formant one and two of KIT in Vancouver and Halifax: for F1, \( r = 0.38 \) in Vancouver and \( r = 0.23 \) in Halifax; for F2, \( r = -0.32 \) in Vancouver and \( r = -0.21 \) in Halifax. The positive F1 correlations indicate lowering of KIT (the higher the birth year, the higher the formant value), and the negative F2 correlations indicate retraction of KIT (the higher the birth year, the lower the formant value). Recall that Kendall’s \( \tau \) might have been a better methodological choice than Pearson’s \( r \) (cf. Subsection 2.3.2). Further recall that in his Montreal paper (2005), which was based on one third of the respondents in his PME study, Boberg himself observed that KIT was retracting but not lowering (2005: 141-143, 147).

In essence, Boberg’s (2010) results, both the ones based on PCE only and the ones based on PCE, PME and other data, conformed to the view of the Canadian Shift as a chain shift as proposed by Clarke, Elms and Youssef (1995) and Labov, Ash and Boberg (2006b). The changed situation between Boberg (2005) and Boberg (2010) casts further doubt on the propositions made by Roeder and Jarmasz (2009) outlined above, which were partially based on Boberg’s (2005) Montreal study, for the Canadian Shift in Toronto (retraction of TRAP and DRESS) and in general (no chain shift but redistribution of the front lax vowels). The vowels involved in the shift seem to be TRAP, DRESS and KIT, but not STRUT; centralization of FOOT seems to be an independent movement (cf. Boberg
2.3. The Canadian Shift

The Canadian Shift seems to be present in at least one Atlantic Province, Nova Scotia, suggesting that the homogeneity of CE can be maintained as a marker of Canada’s urban middle-class speech – at least in formal styles.

Boberg’s final concern were some of the individual vowel spaces of speakers from the PCE and PME data sets, which represent one city each (Boberg 2010: 232). For Montreal, Boberg discussed a Jewish respondent who he found to be less advanced in terms of innovative CE features, particularly the Canadian Shift, but nevertheless participating. Since ethnicity was not significant for the 93 speakers of Montreal English regarding the shift, he explained this lack of advancement “with the speaker’s age [rather] than with his regional or ethnic origin” (Boberg 2010: 238). Although he initially suggested that the Canadian Shift is present in Newfoundland as well, he attested an incomplete merger of the low back vowels for his young middle-class female from St. John’s, “which has prevented a full development of the Canadian Shift” (Boberg 2010: 239). Her vowel plot does, however, reveal that five out of six TRAP tokens have a formant two value of less than 1825 Hz and two of less than 1725 Hz. Furthermore, the three DRESS tokens are all well below 660 Hz in formant one space (cf. Labov, Ash and Boberg 2006b: 130). The lax vowel KIT is not included in this vowel plot (cf. Boberg 2010: 240).

2.3.5 The Social Extensions to the Shift

Hoffman (2010) and Hoffman and Walker (2010) further contributed to the knowledge about the Canadian Shift in Toronto, particularly its social and ethnic dimensions. Hoffman (2010: 126) conducted sociolinguistic interviews with 30 Torontonians, including 24 younger speakers (17-27) and six older speakers (53-80). The younger speakers were additionally stratified according to their ethnic background: Italian, Chinese (Cantonese) and British. The older speakers were exclusively of British origin. At least five, but for many speakers seven or more, vowel tokens of TRAP, DRESS and KIT were extracted from spontaneous speech for acoustic phonetic analysis in obstruent environments only (2010: 127). The shifted status of the vowels was determined via the thresholds outlined by Labov, Ash and Boberg (2006b), similar to Boberg (2010) and Sadlier-Brown and Tamminga (2008). For KIT, Hoffman (2010: 128) used Boberg’s (2005: 142) Montreal mean and his inter-generational mean derived from his PCE data (Boberg 2008b: 137).

Hoffman’s results showed that all three lax vowels were retracted and lowered in apparent time. For KIT, young males had higher F1 values than females (in line with De Decker and Mackenzie 2000: 5 and with Boberg 2005: 147 for KIT retraction); both sexes showed lowering and retraction of DRESS in the young age group; and TRAP was more lowered among females and young females in particular, while young speakers led its retraction. Ethnicity did not play a role in the shift of these three vowels (Hoffman 2010: 122; in line with Boberg 2005: 146 and Boberg 2010: 218). Like De Decker and Mackenzie (2000), Hoffman (2010) reported a shift of KIT in spontaneous speech, in contrast to
Labov, Ash and Boberg (2006b) and Roeder and Jarmasz (2009). Unfortunately, Hoffman (2010) did not mention which social classes she attributed to her speakers.

Hoffman and Walker (2010) focused exclusively on the influence of ethnicity in multicultural Toronto on the (stable) linguistic variable t/d-deletion and the (ongoing) Canadian Shift. Their sample was comprised of 60 speakers of Chinese and Italian origin. For the Canadian Shift they included a further subsample of 22 speakers of British origin. They impressionistically coded *kit*, *dress* and *trap* vowels as shifted or non-shifted and for following manner of articulation, distinguishing among obstruents, nasals and laterals. During the coding process, they found very low rates of *kit*-shifting, so that they only included the remaining two vowels in their regression analyses (2010: 53). Their results showed that *dress* shifting was favored by laterals but disfavored by obstruents, with nasals showing a mixed behavior across all three ethnic groups. For *trap*, the results were different in that laterals showed mixed effects, obstruents favored and nasals disfavored across the ethnic groups (Hoffman and Walker 2010: 54). Women were in the lead for both vowels (2010: 56).

In the same year, Roeder and Jarmasz (2010) presented a revised and edited reprint of Roeder and Jarmasz (2009) discussed above (cf. Subsection 2.3.4). The revisions concerned, for instance, the sample: It was reduced by two speakers to a total of 33 Torontonians, and differently stratified according to age and gender. The imbalance of 14 male and 19 female speakers was evened out to yield 16 men and 17 women. The limits for the age groups were re-defined: the young age group was changed from 20-39 to 22-32 years, the middle-aged group from 40-54 to 37-53 and the old group from 70-85 to 72-85 (cf. Roeder and Jarmasz 2010: 390). Unfortunately, the social status of these speakers was still not mentioned. In terms of the vowel tokens extracted from the speakers for acoustic analysis, the authors mentioned more than 1700 (instead of 1800) for the same vowels as in the (2009) version: /i, e, æ, i, u, ɔ/. The former three constituted a total of 1493 tokens from the “more than 1700”, instead of 75% of 1800 tokens (1350; cf. Roeder and Jarmasz 2010: 390). Unlike in the earlier study, the authors statistically assessed the independent variable age group and sex (2010: 391), instead of only including age group and investigating males and females separately (2009: 4).

These two independent variables were entered into a “MANOVA [sic]” along with F1 and F2 as the two dependent variables (Roeder and Jarmasz 2010: 391). The discussion of their results is difficult to follow. They indicated that TRAP and DRESS were both lowered and retracted in apparent time, and also significantly influenced by sex; KIT was not affected by either variable (2010: 391) – similar to the earlier study (2009: 9). In addition, Roeder and Jarmasz (2010: 392) found that /ɔ/ was significantly retracting, conditioned by age and gender. For all three vowels, /ɛ, æ, ɔ/, the young and middle-aged speakers did not differ significantly from each other, but both differed significantly from the 72-85-year-olds (2010: 393), which adds post-hoc motivation to the binary division of age in younger and older than 70 years in the early study. In addition, it suggests that the
Canadian Shift might near completion in Toronto English (cf. Roeder and Jarmasz 2010: 401; their youngest respondent was born in 1984). In (2010), the MANCOVA results were triangulated with a continuous modeling of age in linear regressions, separately for males and females, which confirmed the results for males: TRAP and DRESS lowered and retracted simultaneously in apparent time, /ɑ/ retracted in apparent time. For women, the findings in apparent time were different: only the retraction of TRAP and DRESS was significant, and neither retraction nor lowering was significant for /ɑ/ (2010: 394).

The motivation for statistically assessing the variables age and gender separately in the linear regression remained unmentioned. Since they did not do so in the MANCOVA, a comment concerning why seems necessary, particularly because the different statistical models are thus not comparable.

Further inconsistencies between the categorical and continuous modeling of age arose: the shift of DRESS correlated more strongly with age for males than females, but no such difference between the sexes was found when age was modeled in three groups (2010: 396). The differences in the results of the MANCOVA and the linear regression suggest that the age ranges in the groups might be incorrectly defined and/or that the effect of age is confounded by that of gender; possibilities that Roeder and Jarmasz (2010) did not mention. Crossing the two groups, age and gender, in the MANCOVA could have shed more light on this and on the linguistic behavior of young males versus young females in particular. As it stands, the MANCOVA results only show no differences between the sexes concerning DRESS in all age groups.

Without addressing any of these different possibilities, Roeder and Jarmasz suggested that the Canadian Shift had two stages: “A combination of lowering and retraction characterizes the earlier stage of the shift, while retraction alone characterizes a later stage” (2010: 396). The earlier stage is thus characterized by simultaneous DRESS and TRAP lowering and retraction, and the later one by DRESS, TRAP and /ɑ/ retraction without lowering. They based this suggestion on Labov (2001a: 501), who stated that in linguistic changes, female speech can be ahead of male speech by one generation (Roeder and Jarmasz 2010: 396). They attempted to “provide evidence in support of this conclusion” by “a linear regression analysis run on the correlation between F2 of /æ, ɛ/ and F2 of /ɑ/ [...]” (2010: 396), including gender. The results of this regression model showed that only the F2 of females’ DRESS and TRAP vowels significantly correlated with that of /ɑ/.

The R² values they presented were 0.305 for DRESS and 0.571 for TRAP (2010: 396). As they correctly noted, “R² can be thought of as the percentage of the variation [...] that is explained by the independent variable” (2010: 394). In other words, formant two of /ɑ/ explains 57% of the variation in formant two of TRAP for females (cf. Roeder and Jarmasz 2010: 396). Recall that for women in general, /ɑ/ was not retracted in apparent time, whereas TRAP and DRESS were. If there was indeed a stage two in the Canadian Shift where leaders (here: females) only retract /ɑ, æ, ɛ/, then why do these leaders not retract /ɔ/? And, if there was a stage one in which non-leaders (here: males) retract
and lower simultaneously, then why do they retract /a/ but not lower it? Further, the inconsistency in the results of three different statistical models might also be due to the make-up of their data.

Based on these results they repeated their interpretation of the Canadian Shift to be no chain shift, even though the results they outlined for their male speakers fit exactly that pattern of a chain shift Labov, Ash and Boberg (2006b) proposed (cf. Subsection 2.3.3). Instead of the Peripherality Hypothesis, Roeder and Jarmasz (2010) proposed other phonological theories in order to account for their results, which go beyond the scope of this dissertation, particularly in light of their methodology: Their interpretation is no longer based on the studies conducted by Boberg (2005) and Hagiwara (2006), but some of the methodological concerns I raised for their (2009) study remain. In addition to the possibility of a confounded effect of age and gender, the unclear social status of the respondents might introduce a bias towards a certain social group; controlling for the linguistic environments remaining in the analysis was not explained; and the possible effects of glottal state and vowel duration (or number of following syllables) was not mentioned. The latter is particularly important, since Roeder and Jarmasz (2010) centered their propositions around the notion of ‘competition between retraction and lowering of the Canadian Shift vowels’. The notion of ‘filling the position formerly occupied by another Canadian Shift vowel’ was not considered.

Two years later, Roeder (2012) published a study in which she investigated the Canadian Shift in Thunder Bay, a small town in Toronto. She compared her young speakers, who were aged 12 to 20, with young speakers from Toronto, who were aged 18 to 24, in order to investigate regional differences between the two distant Ontario cities (2012: 481). In addition to the young speakers from Thunder Bay, she investigated middle-aged (no males) and old speakers in order to establish an apparent-time trajectory for that community (2012: 482). She did not find a significant effect of age, except for THOUGHT (2012: 487), but one of city for TRAP and DRESS retraction as well as KIT lowering (2012: 484). The main aim of this paper was to investigate the geolinguistic diffusion of the shift in Ontario, the province in which most of the research concerning the Canadian Shift had been conducted. The diffusion of the shift on the Canadian mainland is, however, beyond the scope of this dissertation.

A year later, Roeder and Gardner (2013) were concerned with the presence of the Canadian Shift in the Atlantic Provinces. They pointed out that young speakers from Industrial Cape Breton, on the eastern part of Cape Breton Island in Nova Scotia, showed a similar linguistic behavior concerning the Canadian Shift vowels as young speakers in Thunder Bay, Ontario. The sample from Cape Breton comprised 29 speakers, stratified in four age groups: 18-21, 24-50, 53-70 and 82-92 with a balanced gender ratio (2013: 166). Their social status was not mentioned. The acoustically analyzed vowel tokens were taken from a word list, which excluded lexical items with vowels before nasals or liquids, and vowels in a glide context (2013: 166). The results showed that age was not
2.3. The Canadian Shift


The most recent study on the Canadian Shift in Toronto by Hall (fc.) is noteworthy in three respects: First, it tests Roeder and Jarmasz’s (2010) propositions of two stages of the Canadian Shift in Toronto (cf. Hall fc.: 7-8); second, it considers the role of the two formal styles, word list and reading passage, in particular; and third, it reports on the constraints of the preceding and following environments (except for glottal state; cf. Hall fc.: 23). Although Labov, Ash and Boberg (2006b) included both environments in their statistical analysis along with the linguistic and non-linguistic variables mentioned in Subsection 2.3.3, they did not report their findings for any other predictor variable than age. Since Hall attempts to “examine the current state of these [Canadian Shift] vowels”, her judgment sample is restricted “to birth-years between 1983 and 1993, in order to represent the age group immediately after the subjects surveyed by Roeder and Jarmasz (2010)” (Hall fc.: 7). The sample she uses is twice as large as that of Roeder and Jarmasz (2010), with 60 respondents stratified according to age and gender. She divides her participants into two age groups, younger (18-23) and older (24-28), with the first one comprised of 13 females and 15 males and the second comprised of 17 females and 15 males (fc.: 9). The speakers are of a middle-class background: 46 have some university education, approximately half of these are students and half are university graduates in employment; eleven have high school diplomas, and three are college graduates (Hall fc.: 8). From these speaker, she extracted more than 3000 vowel tokens of /i, e, æ, o/ from a word list and a reading passage, among others, for acoustic analysis (fc.: 9). Pre-nasal environments were excluded for TRAP. Unlike Roeder and Jarmasz (2010), Hall only uses linear regression analysis rather than a MANOVA and also employs correlation tests, based on Pearson’s r (cf. Hall fc.: 18). Unlike Labov, Ash and Boberg (2006b), she does not model the social factors, age, sex and style, together with the linguistic factors, preceding and following environment.

Her results show that the two-staged pattern proposed by Roeder and Jarmasz (2010) for the Canadian Shift cannot be confirmed: According to their proposition, speakers leading the change (here: females and younger speakers) should only retract their TRAP, DRESS and THOUGHT vowels, as they are in stage two. Non-leaders (here: males and older speakers) should lower and retract their vowels, as they are in stage one of the Canadian Shift. Hall’s results, however, make clear that women lead men in both lowering and retraction of TRAP and DRESS. Further, the correlation tests show that none of the correlations between F2 of /a/ and the F2s of DRESS and TRAP are significant for women (Hall fc.: 26), which is the opposite of the “evidence” that Roeder and Jarmasz (2010: 396) based the two-stage proposition on. Hall adds that a visual comparison of mean formant values between her data set and that of Roeder and Jarmasz (2010) reveals that her younger speakers seem in fact more lowered than retracted in terms of the Canadian Shift vowels (fc.: 27).
More generally, Hall’s findings indicate no effect of age, which would support the notion that the shift was completed in Toronto, although she cautions that the ten-year span of speakers she included does not allow for such a claim. However, she emphasizes the possibility that her findings in terms of sex may mark gender identity, if it is the case that “the shift had largely run its course by 2012”, as Roeder and Gardner (2013: 165) had put it. What is most prominent in her Canadian Shift vowels are the gender differences: females lead males in terms of lowering and retraction of TRAP and DRESS, lowering of KIT and retraction of THOUGHT (Hall fc.: 26). Concerning the stylistic profile of the shift in her data, she finds that TRAP, DRESS and KIT are more lowered in word-list style, while TRAP, DRESS and THOUGHT are more retracted in the reading passage (fc.: 28). She explains this finding partially on phonetic grounds and partially on social ones: The retraction of vowels in reading style is a natural concomitant of less careful speech, as vowels tend to be undershot in connected speech (centralized towards schwa), whereas in more careful speech, vowels are pronounced more peripherally in order to maximize contrast between them. The lowering of vowels in word-list style may have social meaning. Recall that De Decker and Mackenzie (2000) found more of KIT and DRESS lowering in informal style (spontaneous speech) than in word-list style for their 13 speakers (cf. Subsection 2.3.1). They explained this finding very traditionally in that speakers considered lowered vowels less acceptable than non-lowered ones (2000: 8; also cf. Hall fc.: 29). Since Hall’s results point to the opposite, she suggests that today, speakers may consider non-lowered vowels less acceptable than lowered ones. She further adds that this lowering of vowels in formal styles is consistent with contrastiveness, as lower or more open vowels are clearer than undershot ones (fc.: 29). In addition, she suggests that this phonetic motivation of the vowels to be more retracted in connected speech may also explain Roeder and Jarmasz’s findings due to the “informal nature of their data” (Hall fc.: 30). The preceding and following environments she investigates in isolation from social variables differ in their effects on the Canadian Shift vowels. Hall suggests that preceding and following nasals and liquids should either be controlled for or excluded in future investigations (fc.: 31). Although the influences of the linguistic environments are not uniform for all vowels, particularly preceding nasals and liquids seem to inhibit the participation of the vowels of the shift (Hall fc.: 30).

2.3.6 Lessons Learned

In order to summarize the literature review about the Canadian Shift in mainland Canadian English, I will return to the four purposes outlined in the beginning of this section: 1) the homogeneity of Canada’s middle-class English; 2) the vowels involved in the Canadian Shift; 3) the questioned chain-shift nature of the Canadian Shift; and 4) the role of the social and linguistic variables in the variation of the vowels involved in the Canadian Shift.
2.3. The Canadian Shift

The research on the Canadian Shift so far seems to support the notion that the urban middle class of Canada behaves linguistically relatively homogeneously (cf. Section 2.1) as far as the quality of TRAP, DRESS, KIT and LOT-THOUGHT are concerned. This is indicated by the nation-wide studies in particular, Boberg (2008b, 2010) and Labov, Ash and Boberg (2006b), and by the comparison of the community-wide studies in different research sites in general. The geographical extension of this homogeneity is, however, far from clear. While Boberg (2008b) postulates the Canadian Shift to be a pan-Canadian phenomenon, Boberg (2010) excludes Newfoundland and Labov, Ash and Boberg (2006b) the Atlantic Provinces altogether. However, at least one community-wide study has confirmed the presence of the shift in urban Nova Scotia (Sadlier-Brown and Tamminga 2008).

Which vowels are involved in the Canadian Shift has been a matter of dispute throughout the decades of research. While some authors propose that this is a matter of regional differences, others find disparate results within a single community. If the studies are viewed in the framework of their respective methodologies, the inconsistent findings across different studies/methodologies seem to make sense. Clarke, Elms and Youssef (1995) originally proposed the participation of KIT, DRESS, TRAP, STRUT and LOT-THOUGHT and further found the back vowels, GOAT, FOOT and GOOSE, to be fronting as a movement independent of the Canadian Shift (1995: 223). Almost 20 years after their proposal, agreement exists that TRAP and DRESS are involved in the Canadian Shift, with the merger of LOT and THOUGHT in place. There is less agreement about the participation of STRUT in the shift, because some more recent studies disregarded this vowel altogether. For KIT, there seems to be only disagreement: the host of research suggests that ten studies found a movement of the vowel and five did not. Among those who attested KIT participation in the shift are Esling (1991); Clarke, Elms and Youssef (1995); De Decker and Mackenzie (2000); De Decker (2002); Lawrance (2002); Boberg (2005); Hagiwara (2006); Sadlier-Brown and Tamminga (2008); Boberg (2010); Hoffman (2010). Among those who did not attest a shift of KIT are Labov, Ash and Boberg (2006b); Roeder and Jarmasz (2009); Hoffman and Walker (2010); Roeder and Jarmasz (2010); Hall (fc.). The remaining studies did either not investigate KIT, did not report on it or did not find an apparent-time shift in the community under scrutiny (Woods 1979, 1991; Esling and Warkentyne 1993; Woods 1993; D’Arcy 2005; Hollett 2007; Boberg 2008b; Roeder 2012; Roeder and Gardner 2013). After ANAE was published, most acoustic (socio-)phonetic studies used the thresholds established by Labov, Ash and Boberg (2006b) in order to determine the Canadian Shift in different regions. They defined the shift as follows: it is triggered by the low-back merger with LOT-THOUGHT in the position of [b] (F2 < 1275 Hz); followed by retraction of TRAP (F2 < 1725 Hz); and subsequently by the lowering and retraction of DRESS (F1 > 660 Hz or F1 > 650 Hz; cf. Labov, Ash and Boberg 2006b: 130, 151 fn. 5, 217). Such a threshold is, of course, not available for KIT, since Labov, Ash and Boberg (2006b) did not find a movement for this vowel.
The chain-shift nature proposed by Clarke, Elms and Youssef (1995) has been challenged by Boberg (2005) in his Montreal study. His questioning the phonetic nature of the Canadian Shift was based on the result that DRESS and KIT did not significantly lower but only retract. As a consequence, he proposed a set of parallel shifts of the three front lax vowels, rather than a chain shift where DRESS and KIT lower after TRAP has retracted. This notion was taken up by Roeder and Jarmasz (2009), who found that DRESS and TRAP retraction was stronger than DRESS and TRAP lowering. In a revised version of their (2009) article, Roeder and Jarmasz (2010) did not find lowering of TRAP and DRESS anymore in the same data set. A year after Boberg (2005), ANAE did not support his findings but emphasized a diagonal movement of DRESS, i.e. lowering and retraction. Boberg (2010) adopted ANAE’s view of the Canadian Shift. ANAE reported no movement of KIT and Boberg (2010) based his postulation of parallel retractions on the fact that KIT lowered but did not retract in Montreal, Halifax and Vancouver English – unlike DRESS and TRAP. The results by Boberg (2005) and Sadlier-Brown and Tamminga (2008) contradict these findings for KIT. The results by Hall (fc.) contradict those of Roeder and Jarmasz (2010).

This whole discussion principally originated from the interpretation of the terms ‘lowering’ and ‘retraction’ which Clarke, Elms and Youssef (1995) used to describe the vowel movements that are part of the Canadian Shift. In their impressionistic study, they conceptualized the vowel space of speakers on the basis of the vowels’ articulatory dimensions in relation to the cardinal vowels. In acoustic phonetic analyses, lowering is represented by an increase in formant one values and retraction by a decrease in formant two values. The resulting fine-grained differentiation between lowering and retraction is thus unmatched by impressionistic analyses. However, Clarke, Elms and Youssef (1995) did not propose that the front lax vowels were merely lowering and/or retracting to some place in the vowel space, they proposed that the vowels moved to the positions formerly occupied by other vowels: “Thus, /i/ lowers to /ɛ/; and /ɛ/ lowers to the slot occupied by /æ/ in more conservative varieties” (1995: 212). Whether they move to these slots via more lowering or more retraction is, if at all, of secondary importance. The primary importance is that the lexical item bit sounds like bet, bet sounds like bat and bat sounds like [bat], similar to the vowel in bath in varieties of, for instance, British English with a “TRAP-BATH split” (Wells 1982a: 232). Given the fact that the results of those studies that questioned the chain-shift nature were refuted in follow-up studies, the Canadian Shift seems to be a pull-chain shift as originally proposed by Clarke, Elms and Youssef (1995).

The review of the studies on the Canadian Shift has clearly shown that the role of the social and linguistic variables is quite complex, so that it is difficult to establish agreement on the effects of these variables. Very generally, the shift has been attested for members of Canada’s middle to upper middle classes in predominantly urban speech communities, focusing on word-list speech data. The role of social class membership has
2.3. The Canadian Shift

not been reported by Labov, Ash and Boberg (2006b), although they included speakers from different social backgrounds, unlike all other studies I reviewed. The role of style has either not been investigated or the results not been reported, except by Hall (fc.) and De Decker and Mackenzie (2000) for Toronto. However, the stylistic profile Hall (fc.) outlines is limited to the two rather formal styles, word list and reading passage, and contradicts the findings by De Decker and Mackenzie (2000). Although she explains the contradiction with the fact that speakers today may evaluate lowered front lax vowels positively, unlike at the time De Decker and Mackenzie (2000) conducted their research, it may also be possible that the difference between word list and reading passage she finds is indeed a purely phonetic phenomenon related to articulatory clarity. I propose this possibility because De Decker and Mackenzie (2000) grouped the two formal styles together and contrasted them with spontaneous or more casual speech. Although De Decker and Mackenzie (2000) did not investigate this, it may well be that the difference in vowel movement is no longer significant between word list and reading passage in the Canadian Shift when interview style is included in the analysis. Another caveat is that this reported role of style is limited to the speech community of Toronto, which is believed to be the linguistic center of Canadian English innovations, i.e. the shift might near completion there. The role of age seems to be undisputed in the majority of the studies: younger speakers are most advanced. The role of sex is, however, disputed and depends on the vowel under discussion. For TRAP, females lead; for DRESS, females tend to lead variably, depending on the community, or sex is not significant; and for KIT, if it participates, sometimes females lead and sometimes males lead. In terms of the linguistic variables, very few studies have investigated and reported their influence on the Canadian Shift vowels. The few that did so seem to agree that following manner of articulation is the most important variable, with laterals and nasals promoting vowel lowering and retraction (e.g. Clarke, Elms and Youssef 1995; De Decker 2002; Hall fc.). Hall (fc.) adds that laterals and nasals seem to inhibit vowel movement when preceding the Canadian Shift vowels. The role of glottal state (except Clarke, Elms and Youssef 1995) and number of following syllables has not been reported (Labov, Ash and Boberg 2006b). The lessons learned from approximately 20 years of research on the Canadian Shift are: a) young urban middle-class females from Vancouver to Halifax lead in the shift of KIT, DRESS and TRAP, based on the investigation of formal styles; b) the Canadian Shift is best described as a pull-chain shift triggered by the merger of LOT and THOUGHT and c) the roles of social class, style and linguistic factors are underresearched.
This chapter ties in with the previous one insofar as it outlines the reasons for the distinctiveness of Canada’s youngest province on the one hand and its recent convergence to mainland Canadian (linguistic) norms on the other hand. Both have shaped and still shape the linguistic varieties spoken on the island of Newfoundland. The distinctiveness is immediately apparent from the island’s settlement history, which I outline in Section 3.1. Two main geographical locations are virtually exclusively the origin of Newfoundland’s contemporary population, southwest England and southeast Ireland. Unlike on the Canadian mainland, the settlement patterns of United Empire Loyalists who migrated northwards after the American War of Independence was not a factor in Newfoundland and Labrador (e.g. Clarke 2008a: 93). Once the English and later Irish arrived on the island, the former settled most of the shores of the island, whereas the latter settled almost solely on the Avalon Peninsula in Newfoundland’s southeastern corner.

Apart from larger towns such as St. John’s on the Avalon Peninsula, both settler groups virtually never came into regular and long-term contact throughout the island’s settlement history, which started in the 16th century and ceased towards the late 19th century. Consequently, Newfoundland has been characterized by a relatively small population which is scattered along the shores in small and isolated fishing communities. A second factor contributing to the distinctiveness of Newfoundland and its Englishes is its remote geographical location east of the Canadian mainland. Except for the settler’s ships, island arrivals were sparse until the World Wars when military bases and international airports were built. With joining the Canadian Confederation in 1949, contacts between urban Newfoundlanders and North Americans increased further. Since that time, the influence of the Canadian mainland has only grown – socially, economically and linguistically. Contact between traditionally isolated rural Newfoundlanders, urban Newfoundlanders and mainland Canadians has received further boost by a government-run resettlement program in the mid-20th century, the total cod moratorium in 1992 and the
discovery of offshore oil in the late 1980s, because these three events reignited ongoing small-scale migration patterns within, from and onto the island.

Improved infrastructure and employment prospects have caused general leveling tendencies in the conservative and standard varieties of the island. Having stated that, many of the traditional dialect features are not lost during the adoption of innovative forms from the mainland. Since this thesis is only concerned with the quality of the vowels, merely their system in Newfoundland’s Englishes will be sketched in Section 3.2. In the subsections, I outline the lax vowels of Newfoundland’s Englishes as they have traditionally been described in the literature, which consists predominantly of Clarke’s extensive work (e.g. Clarke 2004b, 2008a, 2010a). I put particular focus on those lax vowels that are typically considered to be part of the Canadian Shift in mainland Canadian varieties: LOT-THOUGHT (Subsection 3.2.1), TRAP (Subsection 3.2.2), DRESS and KIT (Subsections 3.2.3 and 3.2.4). The remaining lax vowels are addressed in Subsection 3.2.5. In the final subsection (3.2.6), I discuss the only elaborate “first wave” (Eckert 2005: 1) variationist sociolinguistic study on the speech community of St. John’s, which has been conducted in the early 1980s: Clarke (1985a, 1991). Although she focused generally on linguistic features of Hiberno-English origin, she also found that young upper-class females do in fact show an innovative retracted TRAP variant – the pivot of the Canadian Shift. The research that followed up on Clarke’s (1991) study provided mixed evidence in support of this finding, which led her to conclude in Clarke (2012) that mainland Canadian English speech patterns do not increasingly incur into the local speech of St. John’s. As I will point out in the discussion of the studies she cited in her recent article, I do not come to the same conclusion. Instead, I argue that the studies are incomparable and (maybe consequently) contradictory in terms of St. John’s’ innovative vowel variants.

In order to shed more light on the situation of the vowels in the standard variety of St. John’s, I review the recent developments that have been reported for attitudes, linguistic identity and other phonological as well as morphosyntactic features in Section 3.3. The first subsection (3.3.1) is devoted to the linguistic identity of the population of St. John’s, which seems to have changed very recently towards refocusing on the pride of having a Newfoundland accent. At the same time, Newfoundlanders are generally very aware of the negative views that are held by outsiders and communicated to Newfoundlanders about their conservative varieties, and they have internalized most of them. As I will outline, the situation is quite complex: The refocusing process currently taking place does not necessarily result in an increase of conservative dialect features (cf. Martha’s Vineyard; Labov 1963), as these are attributed to working-class residents of St. John’s and rural Newfoundlanders. St. John’s residents who are members of the middle to upper classes have been distinguishing themselves from rural Newfoundlanders socially and linguistically for centuries. Due to the recent in-migration from rural areas and the mainland, young middle-class residents of St. John’s may thus feel metaphorically “trapped” between rural Newfoundland’s language and culture as well as mainland Canadian language and
3.1. The Settlement History and Recent Migration Patterns

Newfoundland is the most easterly province of present-day Canada, lying on the northern flank of the St. Lawrence river to the south of Quebec and Labrador in the southeast of the large peninsula which forms eastern mainland Canada (Hickey 2002: 285). The earliest reported contact with natives of the island of Newfoundland with European settlers was at around 1000 A.D., when Norse fishermen established short-term settlements on the island that was to become Newfoundland. They left shortly after, most likely due to animosities with the First Nations (O’Flaherty 1999: 8-9). However, the island of Newfoundland was officially discovered roughly 500 years later by the Genovese John Cabot (Giovanni Caboto) in 1497 (a citizen of Venice, also known as Zuan Cabotto; cf. Clarke 2010a: 4), sailing on an English ship (Hickey 2002: 285; Hiller 2014: n.p.). Shortly after
the discovery, English fishermen used the fishing grounds off the coasts of Newfoundland during the summers (Handcock 1989: 24), after Sir Humphrey Gilbert officially claimed the island for the English Crown (Clarke 2010a: 5). Approximately seven decades later, Basque whalers came down from Labrador to settle along the west coast of Newfoundland and started its famous cod fishery (O’Flaherty 1999: 19) – Newfoundland’s prime economic resource until the 1990s (Clarke 2010a: 3), peaking in the 1880s (Clarke 2010a: 8). In 1610, the English made an attempt to establish cod fishery year-round in the form of permanent settlements (O’Flaherty 1999: 22-23), but these did not become successful until some considerable time later (Kirwin 2001: 441).

Conflicts with the French, who shared the English’ interest in cod fishery off the coast of Newfoundland, dominated the early 18th century. Soon both nations expressed their desire to “settle and possess the island” (O’Flaherty 1999: 39). Between 1702 and 1713, numerous battles were fought on Newfoundland soil (Queen Anne’s War), culminating in the Treaty of Utrecht in 1713 which adjudicated Newfoundland and the adjacent islands, along with Newfoundland’s continental portion, Labrador, to Great Britain. The French were tolerated legally on the northeastern and after 1783 on the western and northern coasts of the island, which would come to be known as the French shore (Clarke 2010a: 7; O’Flaherty 1999: 61). For more than 200 years, the island, as one of Britain’s earliest transatlantic colonies, was part of the Commonwealth as a Dominion of the British Crown (in 1907) until it joined the Confederation as the tenth Canadian province in 1949 (Clarke 2010b: 72; Hiller 2014: n.p.; Kirwin 2001: 445).

Newfoundland English is thus peculiar as it can be described as the oldest and newest New World English. That is, Newfoundland was “discovered” in 1497 and settled by the Europeans shortly after, but became part of Canada as late as 1949. In addition, it is also peculiar in that the precise make-up of the settler groups per ship can be accurately recounted (cf. Handcock 1989: 15-16). The ships’ logs provide detailed documentation of settlement in comparison with similar “outposts”. It is thus not surprising that Handcock (1989: 66) can provide detailed figures such as that 5047 men from South Devon settled in St. John’s between 1675 and 1681. Kirwin (2001: 442) estimates that in 1700, roughly 3500 settlers stayed on the island during the winter, over 15,000 by 1800 and over 100,000 by the 1850s.

Two very distinct main settler groups can be identified throughout the centuries: Particularly from the area around Poole in Dorset, England, but also from Devon, Somerset and Hampshire (Clarke 2010a: 6), the first group arrived in the mid-1700s to 1830 and settled practically everywhere along the shores of the island. The second main settler group came from Waterford and the surrounding areas in southeast Ireland, such as Wexford, Carlow, Kilkenny, Tipperary and Cork (Clarke 2010a: 6; Hickey 2002: 286), particularly from the 19th century onwards (Clarke 2010b: 74). This picture suffers from simplification in that it ignores both a temporal and a social component: The English settled the island as a whole first and had already established themselves as merchants. The Irish settlers
arrived after the success of the English; particularly between 1800 and 1835, they made up about 75% of recorded passengers arriving in Newfoundland at that time. Soon they became infamous for their low social status: According to most historians, they were the poorest of the poor in Newfoundland, most often unable to support themselves and as a consequence responsible for the majority of committed crimes (cf. e.g. O’Flaherty 1999: 124-126, 156, 201). For most of the Irish immigrants, St. John’s was the first and last port of disembarkation, since they did not have the financial means to move elsewhere. Since the English were comparatively wealthy and consequently of a higher social status, Irish arrivals included a large number of future servants in English households near the immigrants’ place of arrival. Thus, the Irish did not settle much of the island of Newfoundland, but stayed predominantly on the Southern Avalon Peninsula, i.e. near the port of St. John’s. After the mid-19th century, immigration to Newfoundland ceased so that the island missed the large influxes of 20th century immigration typical of the U.S. By 1884, 97% of Newfoundland’s population was native-born (Clarke 2010a: 8).

Throughout the centuries, the Avalon Peninsula in the southeast corner of the island of Newfoundland (cf. Figure 3.1) became the homeland of Irish settlers, after it had originally been dominated by the English (Handcock 1989: 81-82, 89). Intermarriages were generally rare due to the social (i.e. religious) differences between the two main settler groups (Kirwin 1993: 65), nevertheless some English merchants married their newly arrived single Irish female servants (O’Flaherty 1999: 47). In addition to the fact that Roman Catholics were allowed to practice their religion from the 18th century onwards, intermarriage may have been the reason that the Irish slowly became the culturally dominant group (Handcock 1989: 89-90).

As late as 1950, much of Newfoundland’s population was distributed in approximately 1300 small fishing or ‘outport’ communities scattered over roughly 29,000 km (ca. 18,000 miles) of coastline (Clarke 2010a: 9). The only link to the world outside of those communities were originally small fishing boats for virtually all of these communities. In later years, a government-run coastal ferry service connected the outport communities (cf. Figure 3.1). Although the railway provided a means of inland transportation since its completion in 1898, the rail journey, which only connected St. John’s with Port aux Basques in the southwestern-most corner of the island (cf. Figure 3.1), was slow, taking 28 hours in 1898 (Collier 2010: n.p.). Newfoundland’s poor road system was slightly improved with the completion of the Trans-Canada Highway across the province in 1965, following the same route as the railway through the province. By the end of the 1960s, passenger trains were eliminated altogether, since the Canadian National railways introduced a Bus service, taking 14 hours for the same route in 1968 (Collier 2010: n.p.). Although the term Highway seems to imply high quality roads, it has to be emphasized that of a total of 948 kilometers, 929 kilometers of the Trans-Canada Highway across Newfoundland were unpaved roads in 1955 (TransCanadaHighway.com 2013: n.p.). To
Figure 3.1: An old map of Newfoundland’s coastal ferries and railway in the 1950s. Map taken from Gagnon (2013: n.p.).
3.1. The Settlement History and Recent Migration Patterns

this day, coastal ferries remain the only link to the outside world for those communities located on the south coast of the island (cf. Canadian Government 2013: n.p.).

In addition to Newfoundland’s relatively isolated geographical location at the northeastern tip of the North American continent, this state of affairs is by and large responsible for the generally conservative nature of local speech and the high distinctiveness on the phonological and morphosyntactic level of modern Newfoundland English from other North American varieties of English (Clarke 2010b: 72). Newfoundland’s linguistic uniqueness has been vividly and maybe exaggeratedly described, for instance, by Herbert (1950: 257):

If you fell ‘blindfolded’ by parachute onto any part of Newfoundland and you listened to the talk, you might say you were in Devonshire, in Dorset, Cornwall, or Somerset, or Yorkshire: you might say you were in Ireland or Scotland (not often in Wales).

But you would never say you were in Canada or the United States.

Newfoundland English preserves numerous features no longer used in other Englishes (Clarke 2010b: 72; also cf. Clarke 2004a,b). Additionally, Newfoundland English shows great internal variation for the same reasons, which can be subsumed under Newfoundland English of British origin, Newfoundland English of Irish origin and standard Newfoundland English (Clarke 2010a: 2-3), but in reality means numerous individual communitylects within a few miles of one another (Clarke 2010b: 75).

After World War II, the English began to settle in St. John’s again from other areas of the island, so that today almost all speakers from St. John’s are of mixed ethnicities (Clarke 1985a: 68). In terms of religion, which traditionally played a crucial role in the social and consequently linguistic reality of Newfoundland settlements outside of St. John’s (cf. Handcock 1989: 241; also cf. “Tocque Formula”, Handcock 1989: 145), St. John’s also shows a mixed picture. In fact, most, if not all, of my respondents did not practice any religion, including some of the oldest respondents in their late 50s to early 60s. The latter group also includes respondents who have converted to a religion other than Protestantism and Catholicism. In addition, practically all of my informants have a typical St. John’s background in being of mixed cultural heritage – English and Irish. Due to the mixture of traditional social statuses and religions, social class membership in present-day St. John’s is a very difficult concept for the inhabitants (referred to as townies by the locals). Although census data would suggest a certain membership, based on occupation, income, neighborhood, housing value, education, etc., most locals do not necessarily feel they belong to such a social class.

It has to be emphasized, however, that the structure of Newfoundland’s society in general led to a maintenance of an almost segregationist pattern far into the 20th century (Clarke 2010b: 74), including the capital St. John’s, as witnessed by the almost 300-year-old denominational school system. It was not abolished until 1998 through an amendment to the province’s constitution (Bergman, Stoakes Sullivan and Fisher 1997; also cf. Clarke 2012: 505). While this segregationist pattern had a profound impact on the formation
of linguistic varieties spoken outside St. John’s, namely Newfoundland English of British origin and Newfoundland English of Irish origin, it did not affect the rather standard variety spoken in the city itself as much (cf. Clarke 2010b: 74). Having stated that, the Irish heritage is still visible, both socially and linguistically, especially in the working class of St. John’s (Clarke 2010a: 11).

In the past 60 years, the island’s economy and society has changed profoundly. Many Newfoundlanders who had been living in the province’s outport communities with fewer than 300 inhabitants were relocated, enforced by a government-sponsored resettlement program. By 1975, 300 such communities had ceased to exist. The cod stocks were seriously declining, which abruptly put an end to Newfoundland’s prime economic resource via the total cod moratorium in 1992. Deprived of a taken-for-granted future in the fishery (and a traditional Newfoundland identity), younger generations from traditional outport communities (referred to as baymen by the locals) were forced to move in large numbers not only to St. John’s and other urban areas in the province (e.g. Corner Brook, Grand Falls-Windsor or Happy Valley-Goose Bay; Clarke 2010a: 17), but also to mainland Canada’s west coast (predominantly to Alberta; cf. Davis 2003). This further led to a serious decline of small outport communities (Clarke 2010a: 9). The majority of the present-day population of Newfoundland resides on the Avalon Peninsula, which includes the capital and largest city, St. John’s. The metropolitan area houses more than one third of the total of Newfoundland’s population of 509,950: approximately 197,000 people (Statistics Canada 2012: n.p.).

The rural-urban divide in Newfoundland is a very salient social concept to the locals, visible in the self-references bayman versus townie. These stand for the rather conservative, traditional and isolated rural live (sometimes with derogatory connotations), which even today is unusually harsh both economically and socially, on the one and the rather innovative, open and wealthy urban life (with an implication of superiority) on the other hand (cf. DNE 1999a,b). Due to its geographic location, St. John’s has always been the gateway to the world outside of Newfoundland: historically, the harbor marked the entrance for new arrivals – permanent settlers from Ireland and England as well as summer visitors –, and the presence of an international airport, built by the military in the Second World War (Collier 2010: n.p.), increased this contact between townies and non-locals, including soldiers from the U.S. and overseas on the military bases.

The numerous linguistic consequences resulting from the relatively mainland-oriented social life in St. John’s have been documented extensively by Clarke (1985a,b, 1991), who found a decline in the frequency of traditional Newfoundland English forms (of Hiberno-English origin, e.g. interdental stopping) among younger generations of speakers in St. John’s long before the cod moratorium in 1992, which was to cause profound and sustained socioeconomic change. Clarke attributed this change in the usage of traditional

26 She sampled her respondents from 1981 to 1982 (Clarke 1991: 111).

As he rightly observes, the discovery of offshore oil in Newfoundland in the late 1980s once more resulted in economic changes, in large part due to the development of the oil industry (De Decker 2012: 42; also cf. Fusco 2007). 35% of Newfoundland and Labrador’s Gross Domestic Product was generated by the oil industry in 2007, an increase of more than 10% since 2004 (De Decker 2012: 42), which was identified as the strongest economic growth in Canada in 2007 (13.4%; Higgins 2010: n.p.). Naturally, the development of such an industry was accompanied by in-migration, mostly from the Canadian mainland. According to The Daily (2011: n.p.), Newfoundland witnessed the highest rate of in-migration within the past 30 years in 2007, while out-migration slowed so that the province experienced a population increase for the first time in 15 years (also cf. De Decker 2012: 42). However, the net in-migration to Newfoundland was approximately 1000 in 2007, the same amount as in 1983 (The Daily 2011: n.p.). It thus seems doubtful to attribute the decline in the use of traditional Newfoundland English features De Decker (2012: 51) observes (he also discusses interdental stopping) merely to the “forces of non-local influences” (De Decker 2012: 42), particularly since these forces have already been attested in the same urban speech community for at least two decades. Likewise, the classification of the decline in interdental stopping found in Petty Harbor, a bedroom community 15 kilometers outside of St. John’s, by Van Herk, Childs and Thorburn (2007; also cf. Childs, Van Herk and Thorburn 2007) as an indicator of rapid social change solely due to recent economic changes does not add further support to such a claim, as the same patterns were found, for instance, by Colbourne as early as 1982 in a rural Newfoundland context (also cf. Clarke 1991: 119). It could thus equally well be the case that townies feel the need to maintain their perceived social differences to baymen, who began to move to St. John’s after the cod moratorium in 1992 and the development of the oil industry in the 1990s (and still visit their parents and younger siblings who stayed behind frequently). One means to do so is using more standard linguistic forms in order to keep perceived advantages over baymen in an increasingly competitive economic climate. That is, the forces that operate on the linguistic choices are not exclusively non-local (also cf. Clarke 1991: 110, who emphasizes that the incoming North American standard forms are not due to in-migration).

3.2 The Sociolinguistic Situation

Newfoundland’s sociolinguistic situation in general and the phonological system in particular has been recently researched and described in detail by Clarke (2004a,b,c, 2008a,b, 2010a,b)\(^{27}\) and in part by her students (D’Arcy 2000, 2004, 2005; Hollett 2006, 2007) and colleagues (Childs and Clarke 2006; Childs, Van Herk and Thorburn 2007; Childs et al.

\(^{27}\) Note that Clarke (2008b) is a reprint of Clarke (2004b).
The latter two groups are, however, concerned with recent sociolinguistic changes in certain phonetic (e.g. interdental stopping) or morphosyntactic (e.g. quotatives) subsystems of Newfoundland English (NLE), but not with a description of the variety as a whole. Thus, this section will almost exclusively rely on Clarke’s work.

According to Canadian census data, 98% of the 509,950 inhabitants of the province of Newfoundland and Labrador have English as their mother tongue. 1% indicated that French was the language spoken in their home, and 1% named one of many aboriginal languages such as Atikamekw or Innu/Montagnais (Statistics Canada 2012: n.p.). As mentioned in the section above, Newfoundland is characterized by a rich inventory of several different outport community varieties of English within a few miles of one another, which Clarke categorized into NLE of Irish origin (NIE), NLE of British origin (NBE) and standard NLE (SNLE). NIE and NBE are predominantly spoken by very conservative or traditional rural Newfoundlanders, whereas SNLE is usually attributed to urban middle-class speakers. Due to the settlement history, NIE is usually to be found on the Avalon Peninsula in the southeast of the island, and NBE occupies much of the remainder of coastal Newfoundland (Clarke 2010a: 16). The distinction between the two conservative varieties is also meaningful for the standard varieties spoken in the urban centers and major towns on the island. While SNLE in St. John’s shows several characteristics of Irish origin, this norm is not shared by SNLE speakers in urban centers off the Avalon Peninsula (Clarke 2010a: 17). Having stated that, the number of former Irish settlers off the Avalon Peninsula was substantially smaller, if not so close to zero that their influence on the language is negligible (cf. Clarke 2010a: 17), than the number of former English settlers on the Avalon Peninsula (cf. Clarke 1985a: 68, 2010a: 9), so that the origins of SNLE in St. John’s are relatively mixed, if not bleached.

The vowel inventory of all NLE varieties and that of standard Canadian English (CE) are identical. NLE’s vowel inventory is also identical with those varieties of General American English in which the low-back merger is present (cf. Section 2.3). In the same vein as most North American Englishes, SNLE does not differentiate between TRAP and BATH vowels (Clarke 2010a: 24). In summary, NLE is clearly a North American variety of English regarding the level of phonetics and phonology, despite the fact that the former Dominion did not join the Confederation until 1949 (Clarke 2010a: 19). In terms of Well’s (1982a) lexical sets, the vowel inventory of NLE’s stressed vowels is comprised of (cf. Clarke 2010a: 24-25):

- six lax vowels (KIT, DRESS, TRAP/BATH, LOT/THOUGHT, FOOT and STRUT),
- four tense vowels (FLEECE, FACE, GOAT and GOOSE/USE),
- three diphthongs (PRICE, MOUTH and CHOICE) and
3.2. The Sociolinguistic Situation

<table>
<thead>
<tr>
<th>Lexical Set</th>
<th>CE</th>
<th>SNLE</th>
<th>NIE</th>
<th>NBE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PALM</td>
<td>æː, ɔ(t)</td>
<td>æː, ɑː, æː</td>
<td>æː(ː)</td>
<td>æː(ː)</td>
</tr>
<tr>
<td>LOT/THOUGHT</td>
<td>ɑ(ː), ɒ(ː)</td>
<td>ʊ(ː), a(ː); ɑ(ː)</td>
<td>æ(ː), ɑ(ː)</td>
<td>æ(ː), ɑ(ː)</td>
</tr>
<tr>
<td>TRAP/BATH</td>
<td>æ(ː)</td>
<td>æ(ː), æ(ː)</td>
<td>ɛ</td>
<td>ɛ, ë, æ(ː)</td>
</tr>
<tr>
<td>DRESS</td>
<td>ɛ</td>
<td>ɛ, ë</td>
<td>i</td>
<td>ë, ɛ1, æ(ː)</td>
</tr>
<tr>
<td>KIT</td>
<td>ɪ</td>
<td>ɪ, ñ</td>
<td>ɪː; ʌ, ʊ(t)</td>
<td>ʌ, ʊ(t)</td>
</tr>
<tr>
<td>FOOT</td>
<td>0</td>
<td>0</td>
<td>ʊ, ʌ, ʊ(t)</td>
<td>ʌ, ʊ(t)</td>
</tr>
<tr>
<td>STRUT</td>
<td>ʌ, ʌ</td>
<td>ʌ, ʌ</td>
<td>ɔ, ʌ, ʊ</td>
<td>ʌ, ʊ</td>
</tr>
</tbody>
</table>

Table 3.1: Phonetic realizations of the NLE lax vowels. Table adapted from Clarke (2010a: 27). CE = Canadian English; SNLE = Standard NLE; NIE = conservative Irish-influenced English (additional pronunciations); NBE = conservative British-influenced English (additional pronunciations); angle square brackets (<>%) indicate recessive realizations by highly conservative speakers.

- six rhotic pre-/r/ vowels (NEAR, SQUARE, START, NORTH/FORCE, POOR/CURE and NURSE/letter).

Two lexical sets have not been part of Well’s (1982a) original classification, namely USE and POOR. I include the former to refer to those GOOSE vowels that are realized with an on-glide such as beautiful and few as opposed to those GOOSE vowels that are glideless (e.g. new; also cf. Clarke 2006). The latter emphasizes the contrast between those CURE vowel tokens that have a quality close to NURSE vowels such as sure and pure in many North American varieties of English and those that do not (e.g. tour). TRAP/BATH, LOT/THOUGHT and NORTH/FORCE are not differentiated in NLE (Clarke 2010a: 24).

Although phonologically identical, the vowels of NLE differ phonetically quite substantially from CE, particularly in the conservative varieties, NIE and NBE, but to some degree also in SNLE (Clarke 2010b: 76). The difference is further particularly evident in the lax vowel subsystem of NLE (Clarke 2010b: 76; Hollett 2006: 146) and even more so in the low lax vowels, TRAP/BATH, LOT/THOUGHT and STRUT (Clarke 2010a: 26). Since this dissertation is solely concerned with innovations in the phonetic quality of the lax vowels that have been associated with the Canadian Shift, LOT/THOUGHT, TRAP/BATH, DRESS and KIT (and STRUT; cf. Section 2.3), only the lax vowel subsystem of the NLE varieties will be discussed in detail in this section.

The considerable difference in the phonetic realizations of NLE lax vowels when compared to CE is, at least in the two conservative varieties, phonetically conditioned. Some of these variants are highly recessive today, that is only a very limited number of highly conservative speakers exhibits such realizations (Clarke 2010a: 26). Table 3.1 provides a general overview of the main phonetic variants that occur frequently in CE, SNLE, NIE and NBE, and, at the same time, the table provides a road map of the individual subsections to follow.
3.2.1 The Low-back Merger

In NLE, most speakers show a full merger of the LOT (⟩a⟩) and THOUGHT (⟩ɔ⟩) vowels with centralized vowel qualities of unrounded [a], [a] or occasionally [a], as compared to CE’s unrounded [a] (Clarke 2004b: 371, 2010a: 26, 2010b: 76; Hollett 2006: 146).28 The vowel quality is the most salient distinction between NLE and CE and is particularly evident in the regional vernaculars of the island (Clarke 2010a: 30, 2010b: 76). Perceptually, the lexical item cot in SNLE occupies a similar vowel space to that of cat in innovative CE, in which the TRAP/BATH vowel is retracted, so that misunderstandings are possible when speakers from the island visit mainland Canadian cities like Toronto (Clarke 2010a: 31; also cf. Clarke 2005). According to Labov, Ash and Boberg (2006b: 86-87), NLE’s quality of the merged vowel resembles the realization of the same vowel in the Northern Cities Shift in the United States, i.e. in a very front position (also cf. Clarke 2010a: 31). As Boberg (2010: 239) emphasizes, such a front position consequently hinders the retraction of TRAP/BATH in NLE, which would be the first stage of the Canadian vowel shift (Clarke, Elms and Youssef 1995: 212).

He further argues that for some NLE speakers, the low-back merger is variable, i.e. not fully completed (Boberg 2010: 239). As Clarke (2010a: 30) notes, a few speakers may have a length distinction between LOT and THOUGHT (also cf. Clarke 2004b: 371), but the THOUGHT vowel only rarely shows greater retraction than the LOT vowel, for instance in an NBE enclave variety in Conception Bay (Clarke 2010b: 76). In a few regions, the merged vowel may also be less rounded for the LOT set, with an unrounded and central quality of [e] (Clarke 2010a: 31, 2010b: 76-77). However, most speakers of NLE show a full merger of the two low back vowels, particularly on the Avalon Peninsula (Kirwin 1993: 74). In addition, “[... ] younger upwardly mobile speakers are tending to adopt more retracted CE-like variants [...]” (Clarke 2004b: 371).

All NLE speakers with a fully merged LOT/THOUGHT vowel show a retracted variant in the PALM-class (e.g. psalm and calm), which is typically a learned pronunciation (Clarke 2010a: 30). Since the PALM lexical set is very small and consists of very infrequently occurring lexical items, I exclude the PALM vowel from the analyses in this dissertation.

3.2.2 Trap and Bath

As mentioned in Section 3.2, TRAP and BATH are not differentiated in NLE. As in many other North American varieties of English, the vowel may be lengthened when it occurs before a fricative or a nasal plus consonant cluster such as in the lexical items laugh, last and dance (Clarke 2004b: 370, 2010a: 29), which was the first stage of a historical sound change that inconsistently affected the lexicon of southern British English varieties via lexical diffusion (Wells 1982a: 100-101, 232-233). In a second step, the lengthened vowels

28 Note that Scargill and Warkentyne (1972: 64) found the merger to be at a rate of 70% in Newfound-
land, based on a survey.
were retracted to a quality of modern RP’s bath vowels, a process termed the TRAP-bath split by Wells (1982a: 232). This historical sound change, however, is unlikely to have been transported to Newfoundland, since the settlers of West Country origin (cf. Section 3.1) never had a TRAP vowel ([æ]), but rather used a centralized [a] sound in the TRAP lexical set (Wells 1982b: 345). The handful of older educated St. John’s residents who do realize bath vowels with a retracted [a]-like quality do so in “[...] no doubt conscious imitation of the prestige British model [...]” (Clarke 2010a: 30). Yet, in an NBE enclave variety in Conception Bay, this realization occurs historically in the bath lexical set (Clarke 2010a: 30, 2010b: 76).

In all NLE varieties, /æ/ is generally more tensed, raised and fronted to an [ɛ]-like quality than its CE counterpart (Clarke 2004b: 370; Hollett 2006: 146), which is the most salient distinction between NLE and CE, particularly evident in regional vernaculars and to a lesser degree in innovative speech (Clarke 2010a: 29, 2010b: 76). The raising and fronting occurs in all linguistic environments, although it is most apparent before nasals (e.g. in land and lamb) and in a few lexical items such as catch (Clarke 2010a: 30, 2010b: 76). In this regard, NLE behaves precisely like many North American varieties of English, including CE (cf. e.g. Boberg 2000: 5 and Labov 1991: 5, 1994: 503). In fact, the Atlas of North American English (ANAE) found a substantial distance of up to 300 Hz between trap vowels before nasals and before other consonants for young speakers from St. John’s (Labov, Ash and Boberg 2006b: 176), which groups NLE with varieties of CE in the western provinces of Canada (Labov, Ash and Boberg 2006b: 223).

In conservative NBE, raised /æ/ may be diphthongized to [ɛɾ] before voiced velars, for instance in bag ([bɛɾg]) and plank ([plɛɾk]; Clarke 2010b: 76). In her (2004b) article, Clarke maintains that such diphthongization occurs more frequently before alveolars and alveopalatals, particularly /n/ as in dance, and among younger speakers in English-settled areas (2004b: 370). Another recent innovation Clarke outlines is the lowering and retraction of /æ/ towards [a], which is apparent among young urban females, particularly in St. John’s, and “reflect[s] the influence of the Canadian Shift” (Clarke 2004b: 371; also cf. Clarke 1991: 116, 2008a: 103, 2012: 514). D’Arcy (2005: 337) finds the same innovation among pre-adolescent and adolescent middle-class females from St. John’s, but emphasizes the phonological conditioning of this innovative variant as it occurs most frequently before /g/. This result is not very surprising, since all of the relevant social factors, age, gender and social class, which likely condition the innovation, are controlled for in her study. Boberg supports D’Arcy’s findings in (2008b: 136, 148), but revokes them in (2010: 204, 240). Hollett (2006: 146) cautions that the effect of lowered/retracted TRAP is not clear in NLE, i.e. it may not be an active feature in the speech community. However, I consider her findings somewhat inconsistent as I will detail in Section 4.1. Clarke

Note that the effects of linguistic environments will be detailed in Subsection 4.4.5.

D’Arcy’s (2005) study will be detailed in Section 4.1.
(2008a: 103) also emphasizes that the innovative, retracted TRAP variant is present in St. John’s, despite the findings of Hollett (2006, 2007) and Labov, Ash and Boberg (2006b).  

3.2.3 DRESS and Kit Raising

The DRESS and KIT vowels are most often realized as standard lax low-mid [ɛ] and high-mid [i]. In addition, they have traditionally been described as being variably raised in both conservative varieties of NLE, which is favored by a following stop in NIE and a following stop or affricate in NBE, in words such as wind, ten, get, peck and wedge, i.e. it is phonetically conditioned (Clarke 2010a: 28, 2010b: 77). In the latter variety, it may also be lexically conditioned (Clarke 2010a: 28).

The raised DRESS-variant can also be tensed among very conservative rural speakers on the island (e.g. dead [diːd]). In NBE, the non-raised variant [ɛ] may be tensed and diphthongized in a stressed syllable before voiced velars such as in keg [kʰɛɡ] (Noseworthy 1971: 54; also cf. Clarke 2004b: 370).

Like the raised DRESS-variant, standard lax [i] can be tensed to [iː] among some conservative speakers, mostly on the Irish-settled Avalon Peninsula, including St. John’s (Clarke 2004b: 369). In conservative NIE, tensing of [i] is typical of certain linguistic environments, whereas in NBE it is phonetically and perhaps lexically conditioned (Clarke 2004b: 370, 2010a: 28). In the latter, it occurs before alveopalatal (e.g. fish), alveolar nasals (e.g. in, wind), velars (e.g. big) and occasionally before labiodentals (e.g. skiff) and laterals (e.g. pill; Clarke 2010a: 28, 2010b: 76). In SNLE, the KIT vowel is usually not as tense, but rather raised or peripheralized ([iː]). This is particularly true for St. John’s, but not as much for other urban areas in the province (Clarke 2004b: 370, 2010b: 76). In all NLE varieties and even among younger speakers, [i]-tensing most frequently occurs in verbal -ing and in possessive his ([hɪs] as in he’s; Clarke 2004b: 370, 2010a: 28, 2010b: 76).

3.2.4 DRESS and Kit Lowering

Suddenly my friend halted to ask me if I had ever heard of the old saying ‘Wher dere’s a well dere’s a way’ [...] (Paddock 1982: 71)

Only in NBE, phonetic conditioning causes the DRESS and KIT vowels to be uttered with a lowered quality, most frequently when followed by /l/ as in will [wl] and yellow [jælou] (Clarke 2010b: 77). The lowering of KIT is additionally favored when the vowel is followed by a fricative (e.g. different or with) and occasionally by a nasal as in since (Clarke 2004b: 370). In terms of DRESS, lowering is additionally promoted by anterior

---

31 As I will outline in Subsection 3.2.6, Labov, Ash and Boberg (2006b: 130) attest that there is no Canadian Shift east of Montreal in the atlas data, which does, however, not per sé contradict the presence of an innovative retracted /w/ variant in St. John’s.
voiceless velars as in *breakfast* or *wreck* (Clarke 2004b: 370, 2010a: 28-29, 2010b: 77). In summary, the phonetic behavior of the two vowels indicates a tendency toward a partial merger in conservative NBE. This extensive overlap is, however, hindered by the effects of the standard English spelling system. Contemporary speakers are generally aware of the more formal variants of these two vowels and tend to use those variants in their more formal styles (Clarke 2010a: 29).

Merely for the DRESS vowel, Clarke (2004b: 370) states that a somewhat lowered and retracted variant is about to enter the speech community in a broad set of phonetic contexts. This variant is particularly evident in the speech of younger urban Newfoundlanders, which reflects the influence of the innovative CE tendency to lower this vowel as part of the Canadian Shift (2004b: 370). As for TRAP, this suggestion has first been supported by Boberg (2008b), but later revoked (Boberg 2010). Likewise, the findings of Hollett (2006, 2007) and Labov, Ash and Boberg (2006b) do not provide evidence for such an incoming innovation (Clarke 2010a,b, 2012 herself does also not maintain this suggestion).

### 3.2.5 Foot and Strut

In SNLE, both FOOT and STRUT are realized as in most North American varieties: STRUT as a lax unrounded low-mid back vowel [α], and FOOT as a lax rounded high-back vowel [o] (Clarke 2004b: 371, 2010a: 31). However, the quality of STRUT in SNLE is more retracted than in CE (Clarke 2010b: 77). Especially in Irish-settled areas of the island, the realization of STRUT is typically accompanied by lip-rounding, resulting in an [ɛ]-like or at least in an [a]-like quality (Clarke 2004b: 371, 2010a: 31, 2010b: 77). Similarly to the KIT vowel, STRUT may undergo raising and tensing in NIE, resulting in a quality similarly to the GOOSE vowel (Clarke 2010a: 31).

In the two conservative varieties, NIE and NBE, some lexical items such as *put*, *took* and *look* are realized in a quality perceptually similar to that of [α], although these items belong to the lexical set of FOOT in standard varieties of English (Clarke 2010a: 31-32, 2010b: 77). Irish-settled areas are home to old speakers who occasionally raise and tense [o]. This realization also occurs before voiced alveolars (e.g. *wood*) and voiceless velars (e.g. *brook*) in English-settled areas (Clarke 2004b: 371, 2010a: 31). In SNLE, on the other hand, the STRUT vowel in the affix *un-* is typically produced with the central [ɛ] vowel of lexical items belonging to the set of LOT, so that, for instance, *unsure* is perceptually identical to *onshore*. Particularly young urban speakers tend to produce more centralized variants of FOOT in certain lexical items (e.g. *good*), as elsewhere in North America (Clarke 2004b: 371).
3.2.6 The St. John’s Speech Community

The lax vowel subsystem outlined in the previous subsections was described for the island of Newfoundland as a whole. This subsection is concerned with a summary of Clarke’s (1985a, 1985b, 1991) “first wave” (Eckert 2005: 1), i.e. macro-sociolinguistic quantitative, study, which was carried out in a traditional Labovian (VS) framework (cf. Labov 1963, 1966, 1972b, 1982, 1994) solely in the speech community of St. John’s in the early 1980s (also cf. Clarke 2012: 505).32

Her study differs from the present one in several respects, which can be broadly divided into methodological and theoretical ones. In terms of methodology, Clarke included more informants (stratified random sample of 120; 1991: 111, 2012: 505), analyzed more phonetic variables (“some 20”; 1991: 112; or 25; 2012: 505), distributed her informants according to four age groups and five socioeconomic statuses (SES),33 included religion as a predictor variable, only used social variables as predictors (age, sex, SES, religion) and analyzed the phonetic features impressionistically/auditorily (cf. Clarke 1991: 111, 2012: 505-506). All of these methods are characteristics of traditional “first wave variation studies” (cf. Eckert 2005: 1), and the latter was commonplace for phonetic studies at that time (cf. e.g. Schmied 1991a,b; although they have not been conducted in a VS framework). With regard to the theoretical framework, Clarke was interested in rapid phonological change in the speech community as a result of its loss of isolation from the outside world through World War II and Newfoundland’s joining of the Confederation in 1949, which transformed its political, social and economic structure. Both of these major events increased contact with mainland North America, also for those Newfoundlanders who had never had any direct interaction with mainland Canadians, as a General American Speech variety and/or CE were disseminated via the nationwide network of the Canadian Broadcasting Cooperation from 1949 (Clarke 1991: 110).

As the capital of St. John’s contained a somewhat higher proportion of Irish than English residents in the 19th and early 20th centuries, Clarke predominantly investigated phonological features of Hiberno-English origin (1991: 110-111). Of the more than 20 features she analyzed, six are discussed in detail in her (1991) publication:

1. ‘Clear’ or fronted variants of post-vocalic /l/ as in *pill* or *pull*.
2. A slit fricative pronunciation of post-vocalic, non-preconsonantal /t/ in *bit* or *pity*.
3. Lowered and monophthongal variants of the tense mid vowels FACE /e/ and GOAT /o/ as in *made* or *go*.

32 Note that Clarke (2012) is in principle an updated and revised reprint of her original (1991) study, including findings of similar, more recent studies carried out in rural Newfoundland (e.g. Childs et al. 2010; Colbourne 1982; Lanari 1994; Newhook 2002; Paddock 1982; Van Herk, Childs and Thorburn 2007).
33 Note that Clarke (1991: 121) calculated the SES based on independent scores for each of the factors of income, occupation, father’s occupation, education and housing (also cf. Clarke 1985b: 76).
3.2. The Sociolinguistic Situation

4. A rounded and retracted pronunciation of the vowel /ʌ/ as in *strat*.

5. Neutralization of the oppositions /ai/ vs. /aɪ/ (e.g. *toy/tie*) and /or/ vs. /aɹ/ (*port* vs. *part*).


None of these features are of immediate relevance for this study, but Clarke also reports the lowering and retraction of TRAP among younger speakers in her sample. The relative frequency of use was 0.08 for teenagers in casual speech style and 0.04 for the other groups (p = 0.10; Clarke 1991: 116). However, the stylistic profile was on average relatively flat, particularly for the use of local features, indicating a lack of stylistic stratification. That is, age, gender and SES (in this order) exerted a much stronger influence on innovative feature use than style (Clarke 1991: 121). The lack of a stylistic profile in the speech community of St. John’s is strikingly similar to the incipient sound change, MOUTH-fronting, on Canada’s mainland, outlined in Section 2.1. In case of TRAP, the innovative CE variant was predominantly used by young females from the upper class (relative frequency of 0.29; Clarke 1991: 118). Again, this finding is in line with that of Esling and Warkentyne (1993: 240-242) for TRAP-retraction in Vancouver, i.e. on the Canadian mainland (cf. Section 2.3). Overall, Clarke’s (1991) findings indicate that of the four social variables included, age is the most important one, as it is “the only factor for which significant differences in language use are evident in the 120-subject sample” for every one of the phonological variables examined (1991: 112). Furthermore, younger generations of St. John’s residents tend to replace local variants with CE pronunciations.

In her (2012) article, which is based on her (1991) analysis of her data set and some more recent studies (Master’s Theses, except for Boberg 2010 and D’Arcy 2005), Clarke discusses merged LOT-THOUGHT in addition to non-differentiated TRAP-BATH (2012: 513-514). For the LOT/THOUGHT vowel, Clarke (2012: 514) reports a parallel finding to TRAP, namely that upper-class females showed most retracted mainland CE-like [a] variants in St. John’s. In terms of the Canadian Shift, she merely states that the lax vowel innovations, which she determined impressionistically, tend to be more conservative in NLE than in CE, based on Boberg’s (2010: 145-146) interregional mean formant one and two values of 86 participants from eight Canadian regions, including five female and one male McGill University students from Newfoundland. What she does not mention is Boberg’s earlier observation that the Canadian Shift seems to be “a pan-Canadian development, at least among middle-class youth, contrary to the report of [Labov, Ash and Boberg 2006b] that Atlantic Canada does not participate in it” (Boberg 2008b: 136), based on the exact same 86 participants, and his repetition of those exact words in Boberg (2010: 204). Having stated that, he revokes that statement when discussing

34 Boberg’s (2008b, 2010) studies will be detailed in Section 4.1.
the (Canadian Shift) vowels of his young female Newfoundland case study (Boberg 2010: 239).

Clarke’s view on and the interpretation of her data from the 1980s have recently changed somewhat, most likely due to the mixed evidence that has been brought forward since 1991 in terms of TRAP retraction, as she outlines herself (Clarke 2012: 514). Although the findings she reports for TRAP vowels that feature retraction in her recent article are by and large the same as the ones she reported in Clarke (1991), she adds that, first, the retraction is generally not as advanced as that of innovative CE variants, and second, that (suddenly) the greatest formal style use comes from the women aged 55 and older (2012: 513-514). The first addition is not necessarily informative with regard to the presence or absence of CE innovations, as Boberg’s (2010: 238) Jewish respondents from Montreal also had relatively front TRAP realizations and nevertheless showed a Canadian-Shift pattern. The second addition may well be in response to Hollett (2007: 52-53), who found the exact same in her Master’s thesis, using acoustic phonetic methods in her analysis of only word-list data. As I will outline in Section 4.1, Hollett’s (2006, 2007) number of respondents is quite small (12 females, including four speakers from Clarke’s 1991 data; cf. Hollett 2007: 33), the total amount of tokens for eight vowels in three generational groups is only 1156 (Hollett 2007: 36), and she did not indicate whether she analyzed TRAP tokens before nasals and /g/ separately from other following environments, although the vowel is usually raised and fronted in the former (Clarke 2010a: 30, 2010b: 76) – even in CE regions with an innovative retracted variant (Labov, Ash and Boberg 2006b: 222).

Although D’Arcy (2005) provides evidence that pre-adolescent and adolescent females from the middle class of St. John’s impressionistically retracted their TRAP tokens in up to 25% per cent of all tokens (2005: 340; also cf. Section 2.3), Clarke (2012: 514) emphasizes that this realization was more common in formal than in casual style and remained a minority variant. In fact, the innovation was significantly more common in casual style and among adolescent girls within the local parentage group of D’Arcy’s study; only within the non-local parentage group, the innovation was more common in formal style but never reached significance (D’Arcy 2005: 340). Recall that in 2008, Clarke emphasizes (2008a: 103), based on her (1991) and D’Arcy’s (2005) results, that TRAP retraction is an incipient change in St. John’s despite the findings of Hollett (2006, 2007) and Labov, Ash and Boberg (2006b). Between 2008 and 2012, no new data has been presented that would justify her claim that TRAP retraction is not increasingly incurring into NLE (Clarke 2012: 513). As further evidence against an increase in TRAP retraction in St. John’s since the 1990s, Clarke (2012: 514) cites a student paper (Reckling 2008), which replicated D’Arcy’s study on a much smaller scale and to some extent Hollett’s (2006, 2007) study. Contrary to D’Arcy, Reckling found, however, no retracted variants of ash for speakers within her non-local parentage group, and 10% retracted ashes for one male respondent in the local parentage group. If generalizable, this would in fact suggest
that young males with parents from St. John’s retract some of their TRAP tokens, but those with parents from the Canadian mainland would not. In addition, Reckling’s study was most likely impressionistic (cf. percentages of variant distribution 2008: 39), so that the same limitations as for D’Arcy’s study hold (cf. Subsection 4.1). In light of D’Arcy’s study, Reckling’s findings seems to be negligible.

As a final addition in some support of Clarke, which she does not mention in her recent article, Labov, Ash and Boberg (2006b: 60) state that St. John’s only has a merger of the low back vowels before nasals, but it is transitional because it was classified as the ‘same’ in perception by the respondents and as ‘close’ in production by the analysts (Labov, Ash and Boberg 2006b: 61). A few pages into the atlas, Labov, Ash and Boberg (2006b: 108-109) consider LOT and THOUGHT merged in St. John’s, but in a front position in the vowel space. In addition, Labov, Ash and Boberg (2006b: 130) emphasize that there is no Canadian Shift east of Montreal in the atlas data. This claim, based on 33 acoustically-analyzed Canadian respondents (Labov, Ash and Boberg 2006b: 217), has already been refuted by Boberg (2008b, 2010); Gardner and Childs (2011) and Sadlier-Brown and Tamminga (2008) for Nova Scotia. It has to be emphasized that Labov, Ash and Boberg’s observations concerning Newfoundland are based on two speakers, a 35-year-old male and a 24-year-old female (2006b: 217).

Overall, the picture that emerges from the recent studies that have been concerned with the Canadian Shift vowels in St. John’s is far from clear. Unlike Clarke (2012: 513), I do not see that the increasing incursion of innovative CE variants into NLE varieties is not supported by recent studies, but rather that these studies are hardly comparable and somewhat contradictory. As I will outline in Subsection 3.4, two groups of publications emerge: one that generally doubts CE innovations in (St. John’s and) Newfoundland, including the Canadian Shift, (e.g. Boberg 2010; Clarke 2012; Hollett 2006, 2007; Labov, Ash and Boberg 2006b) and one that does not (e.g. Boberg 2008b; Clarke 1991, 2004b, 2008a; D’Arcy 2005). Some of the studies in both groups have methodological problems (e.g. Hollett 2006, 2007), whereas others have a too limited data set for making valid generalizations about CE innovations in St. John’s (e.g. Boberg 2008b, 2010; Labov, Ash and Boberg 2006b), which is at the same time explicitly not the goal of these studies. An acoustic (socio-)phonetic study based on a representative judgment sample and linguistic as well as social predictor variables, including reading and spontaneous-speech styles, that could clarify the status of the Canadian Shift vowels in St. John’s does not exist.

3.3 Recent Developments

After I have outlined the lax vowel subsystem of Newfoundland as it has been described traditionally, and after the recent phonological findings for LOT/THOUGHT and TRAP/BATH in St. John’s have been discussed, I will now outline additional recent linguistic developments that go beyond the lax vowels usually associated with the Canadian
Shift. As I will show below, the speech communities of Newfoundland in general and of St. John’s in particular display a number of recent mainland CE innovations, on the phonological as well as the morphosyntactic level. If the partially contradictory findings with regard to the Canadian Shift in NLE vowels are pieced together with other innovations in NLE and in CE, the picture that emerges does not allow for a general and strict rejection of the presence of the Canadian Shift in St. John’s. From this picture, I will thus derive the overarching research question, including a number of sub-questions, to be answered in the present study.

In the first subsection, I will outline recent changes in the attitudes towards Newfoundland speech and in the linguistic identity of Newfoundlanders in general. In Subsection 3.3.2, recent phonological and morphosyntactic changes will be the focus of discussion in order to derive the research questions for this dissertation from the former two subsections, Sections 3.1 and 3.2, as well as Chapter 2.

### 3.3.1 Language Attitudes and Linguistic Identity

Until very recently, outsiders considered Newfoundlanders and Labradorians to be poor, simple, backwards and lazy but friendly people (Clarke 2010a: 132-133). Particularly the salient phonetic features of Hiberno-English origin such as interdental stopping, the unrounding of the /ɔi/ diphthong and a vague Irish cast to the vowels suggested to English Canadians that Newfoundlanders were uneducated fishermen (Pringle 1985: 186; also cf. Clarke 2010a: 134). In McKinnie and Dailey-O’Cain’s (2002) study conducted in the framework of perceptual dialectology, Ontarian and Albertan (Canada’s wealthiest provinces; cf. Clarke 2010a: 135) informants rated NLE negatively in the domains of ‘pleasantness’ and ‘correctness’ and labeled it with terms such as ‘drawl’, ‘Newfie talk’ and speech that sounded ‘extremely fast lower class’. The term Newfie or Newf is an ethnic label which was invented by American, British and Canadian military personnel in the 1940s. They used it as an ethnic slur (ethnophaulism) – similarly to the term Nigger. While most Newfoundlanders may still perceive the term in this way, some young Newfoundlanders take pride in being Newfies and others differentiate both meanings as a function of whether community outsiders or insiders use the term (King and Clarke 2002: 539-540). These negative views of Newfoundlanders and NLE were in part also caused by Newfoundland’s economic status as a ‘have-not’ province, i.e. it was heavily dependent on the Federal government’s annual equalization payments, designed to ensure a comparable standard of living throughout the nation. Needless to say, ‘have’ provinces consider residents of such ‘have-not’ provinces as Newfoundland and Labrador as ‘welfare bums’ (Clarke 2010a: 132).

Such negative stereotypes from the Canadian mainland have been largely internalized by Newfoundland residents. Although they overtly claim that they hold positive views of NLE, studies using indirect techniques of attitude evaluation have shown that Newfoundland-
3.3. Recent Developments

landers downgrade both social status and general competence of speakers of one of the local conservative outport varieties (cf. Clarke 2010a: 135). While this local speech was evaluated negatively by informants from St. John’s, they judged the conservative speakers to be socially attractive, as they considered them to be friendly, honest and likeable. This pattern does not hold for young residents of small outport communities, who downgrade themselves as conservative speakers: Colbourne (1982: 90), for instance, found that educated young male residents of a rural northeast coast Newfoundland community were aware that they did not speak proper English and expressed a negative attitude toward it. A similar pattern was found by Newhook (2002) for the Burnt Islands variety. She concludes that “[i]t appears that there is little pride with respect to the local variety […], and little concern about dialect preservation” (2002: 94). More recent findings by Childs, Van Herk and Thorburn (2007) and Van Herk, Childs and Thorburn (2007) in Petty Harbour add that salient rural local features (interdental stopping and non-standard verbal -s marking) remain an active part of the performed dialect of young speakers – and thus serve as totems of Newfoundland identity – and yet, young rural speakers are externally affiliated in contrast to their middle-aged and old internally-oriented neighbors.

The Canadian mainlanders who do not interact with Newfoundlanders directly are sometimes exposed to maybe exaggerated, highly vernacular NLE accents on television, as comedy groups from Newfoundland construct characters who use such accents (Clarke 2010a: 135). The same is true for local musicians such as the hip-hop group Gazeebow Unit. Their lyrics are characterized by a highly conservative feature use on the lexical, phonological and morphosyntactic levels (Clarke and Hiscock 2009: 248-251), but the members have also been publicly criticized for not being authentic urban working-class skeets35 (Clarke and Hiscock 2009: 253-254). Thus, Newfoundlanders themselves helped to reinforce the negative stereotypes held by mainlanders who do not necessarily have a chance to expose themselves to authentic urban Newfoundland speech (Clarke 2010a: 135). At the same time, Newfoundlanders react extremely negative when outsiders make fun of working-class or rural Newfoundland speech. As King and Wicks (2009: 266) note, a Nissan TV advertisement in 2006 used an actor from Nova Scotia whose imitation of the local accent was poor and accompanied by standard English subtitles, attempting to provoke humor through irony. Although some local reactions to the commercial, in an online forum, were positive, most were not. One viewer even observed that had “[…] they put a black man in the Newfoundland role, speaking in Ebonics, [… T]he end result would at least [have been] cries of racism and the commercial would [have been] pulled from the air” (King and Wicks 2009: 279).

35 The term skect refers to a person who in the United States is referred to as white trash or trailer trash and elsewhere in Canada as skid. Such people portray a life-style “[…] of disaffected male adolescents whose interests lie not in establishment-endorsed activities, but rather in cars, motorcycles, sex, drinking, smoking, drugs, petty crime, fighting with rival groups from other neighborhoods, and generally hanging out; whose language is laced with profanity; and whose clothing indicates a distinct lack of fashion sense” (Clarke and Hiscock 2009: 249).
In summary, it seems that young rural Newfoundlanders in particular overtly attribute prestige to salient non-local linguistic forms, especially when they intend to leave their outport communities in order to find economic perspectives or pursue careers in the urban centers of Newfoundland (cf. e.g. Colbourne 1982; Newhook 2002). Such migration patterns have increased particularly since the 1960s and again since the 1980s (cf. Section 3.1). The situation is more complex and different for rural Newfoundlanders from bedroom communities such as Petty Harbour (cf. Childs et al. 2010). For young urban Newfoundlanders, the situation is further complicated by the increased in-migration of rural Newfoundlanders from outport communities (cf. Section 3.1).

As Clarke (2010a: 132) notes, in November 2008, Newfoundland assumed the status of a ‘have’ province, independent of the equalization payments from the Federal government. This achievement was made possible by the new revenues from Newfoundland’s offshore oil – an industry that has silently started 20 years ago and been developed further ever since. The province that had formerly driven the Canadian economy, Ontario, was demoted to a ‘have-not’ province at the same time. This change was naturally much-celebrated in Newfoundland, especially since the province had been habitually seen as the poor cousin of the Canadian Confederation. At about the same time, St. John’s publicly debated in the *Telegram* – St. John’s daily newspaper – that Newfoundlanders should refocus on their ability to laugh about themselves and communicate the pride of having a local accent (also cf. King and Wicks 2009: 274). This pride is, for instance, visible today in that “summer dinner-theatres [catering] to the tourist market often involve portrayals of the ‘country bumpkin’ character, with, of course, a broad local accent” (Clarke 2010a: 136). In addition, all of the stereotypes outsiders nurture have been marketed by the locals as “Newfie-ism”, including Newfie joke books, Newfie mugs with handles on the inside, the Newfie rolling pin which is square, the Newfie revolver with the barrel pointing toward the shooter and making visitors ‘honorary Newfoundlanders’ via an invented ritual termed ‘screech-in’ (which I did not have to go through when I arrived in the community; cf. King and Wicks 2009: 264). Recently, a large part of Newfoundland identity is constituted by the image of hard drinking and hard partying, particularly on George Street in downtown St. John’s, which has the most bars per square foot in North America (which I had to visit and test for beer on a few Saturday nights; cf. Clarke 2010a: 137).

For most of the middle to upper-class residents of St. John’s, a refocusing on local identity and local speech does, however, not necessarily mean to speak a conservative variety, full of traditional and socially stigmatized linguistic features, because they are usually attributed to the speech of working-class skeets. They are also often attributed to rural Newfoundland speech (cf. Clarke and Hiscock 2009: 256). The desire to emphasize the social distinctiveness of St. John’s middle class from skeets and baymen is rooted in Newfoundland’s settlement history. That is, historically, the English merchants were the dominant social class, and the Irish servants were the inferior social class (cf. Section 3.1). Today, however, the ethnic distinction is largely watered down, but the social distinction
remains (along with the attitudes) in that dominant middle- to upper-class members are of mixed English/Irish heritage and the inferior working-class members are of mixed heritage as well. The desire to emphasize this social distinctiveness has regained public awareness through, for instance, Gazeebow Unit’s inauthentic public mockery of a skeet underclass (i.e. they were not part of the skeet underclass – community outsiders – and yet imitated their speech) which reproduces the “dominant ideologies of social class” (Clarke and Hiscock 2009: 256), and is reinforced through the increased in-migration of baymen into St. John’s since the 1960s and the 1980s. According to the Dictionary of Newfoundland English, the terms bayman and townie have been in use since the early 20th century and served predominantly for the residents of St. John’s to express their perceived superiority in terms of social status, wealth, being ‘cultivated’ and consequently linguistic choices (cf. DNE 1999a,b). The above-mentioned studies using indirect techniques of attitude evaluation support this picture: Most Newfoundlanders share the negative views of bayman speech, but openly stating or admitting it is a social taboo. While middle-class townies express social distinctiveness from skeets and baymen via using more standard speech, in-migration from the Canadian mainland, due to the development of the oil industry, has increased the direct contact between St. John’s residents and mainland Canadians, after the contact had initially been established during World War II. This was naturally accompanied by an increase in regular, often daily, contact with innovative standard CE features such as the Canadian Shift over a period of more than 20 years now.

It may even be the case for some, if not many, middle-class residents of St. John’s that after one generation of contact, St. John’s residents of mainland Canadian origin and their innovative speech patterns are now considered part of the social profile of modern St. John’s and its population, i.e. a rural Newfoundland origin may be perceived more negatively and distant than a mainland Canadian origin among residents-by-birth of St. John’s. A refocusing on local identity does consequently not exclude those St. John’s residents of mainland Canadian origin who, within the past generation, have become part of the social networks of residents-by-birth among young middle to upper-class residents of St. John’s. The refocusing may thus only become a conscious issue when off-island Canadians on the mainland – community outsiders – (publicly) mock Newfoundland’s culture and language.

The members of the St. John’s middle class thus take pride in being Newfoundlanders without using more of the conservative (skeet and baymen) features. Rural Newfoundlanders from outport communities, who moved to St. John’s for education and employment, 36

The insider-outsider effect in this context concerns the questions of who has the right to judge and who has the right to perform the vernacular (King and Wicks 2009: 280). This effect is not limited to Newfoundlanders, but generally found with groups that have been exposed to imbalance of power, i.e. segregation, oppression, racism, discrimination, etc., as, for instance, apparent in the African American context via terms such as nigger (an ethnic slur; used by whites), nigga (meaning brotha; used by blacks) and whigga (whites trying in vein to act black; used derogatorily by blacks; cf. Smitherman 2000).
attempt to use more standard variants. The standard are supra-local norms originally introduced via the profound contact with mainland Canadians on a regular basis, which today may be perceived as part of modern St. John’s social life by the locals. In addition to the constraints of the linguistic marketplace (cf. Ash 2013: 359; Chambers 2009: 190-191; Llamas 2007: 72; also cf. Bourdieu 1991), that is using more prestigious or standard variants that are socially legitimate in early adulthood due to increased chances of employment, young middle-class St. John’s residents exhibit higher usage rates of innovative (formerly) supra-local variants, especially when young rural Newfoundlanders who have migrated to St. John’s pose a threat to young middle-class residents of St. John’s in terms of competing for jobs.

### 3.3.2 Linguistic Change

As outlined previously, NLE loses many of its traditional features and displays general leveling tendencies. By and large, the studies conducted in different regions of Newfoundland (rural and urban) generally point to rapid linguistic change across generations. The younger speakers in general tend to favor norms of standard varieties in North America and of “trendy” mainland Canadian speech such as retracted version of TRAP as in path and rack (cf. Clarke 1991: 116, 2008a: 103, 2010a: 142; D’Arcy 2005), similarly to young mainland Canadians who favor U.S. speech patterns (Clarke, Elms and Youssef 1995: 224; also cf. Woods 1993; Zeller 1993). According to Clarke (2010a: 143), the reasons for the rapid linguistic change in Newfoundland are the negative attitudes to local language varieties, out-migration due to non-viability of local economies and the interest of NLE speakers in economic advantages and higher education (also cf. Subsection 3.3.1).

Gender plays a major role in these changes in that women tend to be ahead of men by one generation. This pattern is, however, not uniform across rural and urban NLE varieties. Females appear to be advancing those linguistic features that have connotations of non-local-ness and trendiness. Especially among rural females, there seems to be a greater attachment to salient linguistic features that signal local identity, most likely because they rarely travel outside their local regions and do not intend to change that (cf. Clarke 2010a: 144).

Particularly among rural varieties of NLE, traditional dialect features are in no danger of imminent disappearance, despite the rapid generational change (cf. e.g. Clarke 2004c: 254-256). Younger males’ usage rates for local feature tend to be quite similar to those of older males, indicating that younger speakers are not losing their linguistic roots, but become more bidialectal or proficient at manipulating standard supra-local pronunciations along with traditional community norms (cf. Clarke 2010a: 147-148).

Recent linguistic changes towards innovative CE forms that have been reported in the literature for St. John’s that are not concerning the lax vowel subsystem of NLE are quite numerous. Some of these changes concern the level of morphosyntax, but the
majority of them are phonological. In the **mouth** vowel, for instance, D’Arcy (2005: 336-337) found that among young middle-class residents of St. John’s, the nucleus /a/ is raised to a mid [a] (or mid-back [A]; cf. Clarke 2010a: 38) quality, sharing the same phonological conditioning reported for CE, e.g. [aʊt] in *out* or [pʊt] in *pout* (also cf. Section 2.1). While this feature is typically inherited from southern British Englishes in English-settled parts of the island, this is not the case for Irish-influenced St. John’s, according to Clarke (2010a: 38, 2010b: 79). As D’Arcy (2005: 336) points out, the same phonological conditioning as in CE suggests that it is a recent phonological change. Clarke (2004b: 373, 2010a: 37) adds that the nucleus in *price* is similarly raised in NLE varieties. Raising of the nuclei in *price* and *mouth* is commonly referred to as Canadian Raising, a feature typical of CE (e.g. Chambers 1980, 1989; Chambers and Hardwick 1986; Hung, Davison and Chambers 1993) that has thus served as one of the features in Labov, Ash and Boberg’s atlas to define Canada as a dialect region (2006b: 146).

A relatively recent innovation in CE is the deletion of on-glides before *goose* vowels in lexical items such as *news*, *tune*, *duke* and *student*. Most often, retention of the glide in CE has been interpreted within Canada as one of the salient markers of an Anglo-Canadian linguistic identity that seeks to be distinct from General American English (Clarke 2006: 226). Results by Clarke show that glided and glideless variants are no longer viewed by Canadians as carrying British or American affiliation, but rather new social meanings. As a result, both variants are prestige targets in CE and are thus used at different frequencies as a function of social class (Clarke 2006: 226). Although this feature is usually considered to be a change from below, females do not appear to be in the lead in Clarke’s study (2006: 235). The importance of her study for the present one lies in the fact that St. John’s residents display the same decline in using the glided variant in apparent time as mainland Canadians (2006: 232, 235, 237). Particularly in the local media, Newfoundland is much more innovative than mainland Canada in terms of glide deletion (2006: 239). Another phonetic phenomenon concerning the same vowel is the fronting of *goose* to a [u]-like quality, also a mainland CE feature (Labov, Ash and Boberg 2006b: 143-145). According to Clarke (2004b: 373), this feature is about to make its way inroads to the variety of upwardly mobile younger females in St. John’s (also cf. Clarke 2010a: 35), which “[...] may possibly represent a change in progress” (Clarke 2010a: 36).

Other recent phonological innovations among younger women have been attested in the domain of prosody (cf. Clarke 2008a: 104-105): The first is the use of a feature usually
referred to as high rise terminals, that is, declarative sentences uttered with high rising terminal contours typically associated with questions. The second feature is the variable use of vocal fry or laryngealization, a phonation type more commonly known as creaky voice (cf. Subsection 4.3.1). Both features constitute ongoing changes in North America and were attested by Clarke (2008a: 105) among younger upwardly mobile speakers of NLE.

A final example of recent innovations in the speech of young residents of St. John’s is a morphosyntactic one. Like many North American varieties of English (cf. e.g. Tagliamonte and D’Arcy 2007 among many others), young speakers tend to replace verbal choices such as *say* and *go* in reported speech or narratives with the innovative quotative *be like* in sentences such as *He was really mean to me and I was like* ‘you are such an idiot’ (my example; my emphasis). D’Arcy (2004: 323) found that this recent innovative variant in the quotative system of mainland CE is not only present among young females in St. John’s, but seems to be more advanced in its development than in mainland CE already.

### 3.4 Deriving the Research Questions

As mentioned in Subsection 3.2.6, a sociophonetic (i.e. acoustic phonetic) study of the lax vowel subsystem of St. John’s residents based on a judgment sample including linguistic as well as social independent variables, which allow for establishing an apparent-time trajectory and a stylistic stratification, has never been conducted: Clarke’s original study (1991) was impressionistic and its data collection was carried out almost a decade before the cod moratorium in 1992 and before the discovery of offshore oil, developing into an unprecedented wealth-generating, mainland-Canada oriented and contact-increasing industry in the province; D’Arcy’s (2005) study was impressionistic and only investigated /æ/ lowering and retraction (along with /au/ raising) in formal and casual speech of pre-adolescent and adolescent middle-class females from St. John’s, whose vernacular is still unstable (cf. Bailey 2002: 319; Chambers 2009: 207; Cukor-Avila and Bailey 2013: 246; and Section 4.2); Hollett’s (2006, 2007) acoustic phonetic study only investigated word-list data and has some methodological issues as well as somewhat contradictory findings (cf. Section 4.1). The other acoustic phonetic studies, Boberg (2008b, 2010) and Labov, Ash and Boberg (2006b), were interested in Canada as a dialect region from coast to coast and consequently only included a small number of respondents from St. John’s and/or Newfoundland: two in the atlas and six in Boberg’s data set. Both of these data sets from St. John’s and Newfoundland, respectively, cannot be considered as representative of NLE. Particularly Boberg does not elaborate where exactly his other four Newfoundland respondents, in addition to his one middle-class female case study from St. John’s (2010: 239) and the male one from Gander (2008b: 140), came from on

---

38 Most of these constraints are also true for Reckling’s (2008) replication of D’Arcy’s study.
the island, i.e. are they originally speakers of NLE, NIE or the SNLE variety spoken in the middle class of St. John’s? In addition, Boberg’s results in (2008b) concerning the Canadian Shift in Newfoundland are almost diametrically opposed to those in (2010), although they seem to be based on the same Newfoundland respondents.

As mentioned earlier, two groups emerge in terms of the studies’ results: one that generally doubts CE innovations in (St. John’s and) Newfoundland, including the Canadian Shift, (e.g. Boberg 2010; Clarke 2012; Hollett 2006, 2007; Labov, Ash and Boberg 2006b) and one that does not (e.g. Boberg 2008b; Clarke 1991, 2004b, 2008a; D’Arcy 2005), while most of Clarke’s recent publications (2004b, 2008a, 2010a, 2010b, 2012) are not based on new data, but in large part on Clarke (1991) and for some publications also on the more recent studies mentioned above (e.g. Boberg 2008b; D’Arcy 2005; Hollett 2006, 2007). While none of the studies is representative in terms of making claims about the contemporary lax vowel subsystem of middle-class St. John’s, some of those that do not find evidence for the Canadian Shift or its prerequisite(s) put forward considerably strong postulations. For instance, 1) “No Telsur Canadian city east of Montreal shows the Canadian Shift.” (Labov, Ash and Boberg 2006b: 130; my emphasis); 2) “As in the Nova Scotia system, this residual distinction [i.e. the absence of the low-back merger] has prevented the full development of the Canadian Shift [...].” (Boberg 2010: 239; discussing his case study’s vowel plot); and 3) “[... these apparent changes [including the Canadian Shift in NLE] largely represent ‘change from above’,” affecting the careful as opposed to the casual speech styles of socially mobile segments of the population” (Clarke 2012: 513). This remark implies that the Canadian Shift is usually considered to be a change from below the level of social awareness in the same fashion as the Northern Cities Shift (cf. e.g. Clarke, Elms and Youssef 1995: 220-224; Labov 1994: 262), so that the Canadian Shift would not enter the speech community of St. John’s. However, as I have outlined in Section 2.3, this implication cannot be easily maintained for the Canadian Shift in mainland-Canadian English, because the studies analyzing the shift have neither investigated and/or reported the effects of the full range of styles nor those of the full range of social classes. In fact, as Hall (fc.) shows in terms of the former, TRAP, DRESS and KIT are significantly more lowered in formal word-list style than in less formal reading-passage style; other studies’ results suggest that the Canadian Shift is a middle- to upper middle-class phenomenon in terms of the latter. If the Canadian Shift is indeed present in St. John’s, and if all the recent changes in St. John’s are changes from above, then the Canadian Shift in St. John’s might be a sound change that “will undoubtedly show many intermediate combinations” (Labov 1994: 542) of changes from above (lexical diffusion) and from below (neogrammarian change; cf. Section 2.2). It is further not clear whether the same can be said about the Canadian Shift in mainland Canadian English or not. However, this issue goes beyond the scope of this dissertation, as I have no data from mainland Canada to compare mine from St. John’s, Newfoundland with. In addition,
research suggests that the Canadian Shift is much more advanced on the mainland than it could possibly be in St. John’s (cf. Subsection 2.3.5).

The studies that consider the Canadian Shift to be possibly developing in St. John’s have only included merged LOT-THOUGHT and TRAP/BATH in their analyses, and thus are relatively tentative in the formulation of their claims. However, adding further results from other studies to this picture tones down the relative strength of the postulations outlined above to a certain degree. Contrary to Labov, Ash and Boberg’s postulation in 1) and Boberg’s statement in 2), Sadlier-Brown and Tamminga (2008) do find the Canadian Shift in Halifax, Nova Scotia. In addition, Boberg’s (2008b, 2010) research confirmed that of Sadlier-Brown and Tamminga (2008). In terms of 3), it is important to note that at the same time, Clarke, Elms and Youssef (1995: 224) stress “[...] that a good deal of ongoing phonological change in CE may be socially motivated by an increasing identification with U.S. speech patterns”. The same could be stated for the behavior of the middle-class residents of St. John’s before the very recent refocusing on local speech outlined in Subsection 3.3.1 and even more so for young rural Newfoundlanders having migrated from the outport communities to St. John’s. Additionally, glide deletion before GOOSE in lexical items such as news, which has been attested for St. John’s as well, is usually understood to be a regular sound change (from below; cf. Subsection 3.3.2). Moreover, bearing the other recent innovations from mainland CE attested for Newfoundland in general and St. John’s in particular in mind, it seems that the absence of the Canadian Shift in St. John’s is highly unlikely: Why would young middle-class speakers from St. John’s adopt virtually all innovative mainland CE features, except for the Canadian Shift?

In light of these findings, particularly due to relative strength of the claimed absence of the Canadian Shift in St. John’s and Newfoundland, I formulate the main research question of this dissertation as follows: Do young middle-class residents of St. John’s, Newfoundland, exhibit an innovative, mainland CE-like vowel pattern in apparent time that is usually referred to as the Canadian Shift? Based on the literature review outlined in Sections 3.1 and 3.2 as well as Chapter 2, I add the following supplementary questions: If so, a) Which of the lax vowels, STRUT, TRAP/BATH, DRESS and KIT, are involved? b) For those vowels that change in apparent time, how are the innovative vowel variants constrained by the social (particularly age and gender) and linguistic (particularly linguistic environment) variables and their interactions, respectively? c) Which stylistic profile is exhibited by the innovative vowel variants? More specifically, can the flat stylistic profile (no stylistic stratification) that Clarke (1991: 121) found, be confirmed?

Based on the same sections and chapter, my hypotheses are as follows:

1) The innovative mainland CE shift-like lax vowel pattern is present in the middle class of St. John’s, Newfoundland.
2) Speakers who participate in the innovations have a complete low-back merger (i.e. the merger is not a change in progress).

3) STRUT is stable in apparent time (i.e. it is not part of the innovation).

4) TRAP is retracting with accompanying lowering in apparent time.

5) Subsequently, DRESS is lowering with accompanying retraction in apparent time.

6) Finally, KIT is also lowering with accompanying retraction in apparent time (i.e. it is part of the innovation).

7) The social variables, age and gender, exert a stronger influence on the innovative vowel variants than the linguistic variables such as linguistic environment.

8) Young (middle-class) speakers in general and females in particular lead in the adoption of these innovative vowel variants.

9) The variants are stylistically stratified in such a way that the vowels in earlier stages (DRESS and KIT) of the change occur more frequently in the formal styles, while those in later stages (TRAP) occur more frequently in spontaneous speech.

Since my main research question contradicts the claims that have been brought forward by Boberg (2010); Clarke (2012); Hollett (2006, 2007) and Labov, Ash and Boberg (2006b), I will focus on the triangulation of my results with several different methods. These will be outlined in detail in Chapter 4. I decided to design the present analysis methodologically as similar as possible to the studies of Labov, Ash and Boberg (2006b) and to some degree Boberg (2008b, 2010), since I consider these acoustic phonetic studies to be the methodologically most elaborate, (versus those of D’Arcy 2005 and Hollett 2006, 2007; cf. Section 4.1), and since I principally intend to test their results or replicate their studies for the speech community of St. John’s, NL. My study is thus designed and conducted in a variationist sociolinguistic (VS) framework, similar to Boberg’s (2005) study on the Canadian Shift in Montreal, Quebec, Hoffman’s (2010) or more recently Hall’s (fc.) in Toronto, Ontario. This dissertation is consequently going to fill a gap in the research, because of its focus on St. John’s with acoustic phonetic methods in a state-of-the-art VS framework and its investigation of the role of style in the Canadian Shift.

39 ‘To some degree’ here refers to the fact that Boberg (2008b, 2010) conducted his studies methodologically similar to the atlas (Labov, Ash and Boberg 2006b), but they differ in some details, as I will outline in Section 4.1.

40 Note that I do not conduct the study in a traditional Labovian framework typical of first wave variation studies such as that in Labov (1963, 1966) and Labov, Yaeger and Steiner (1972).
As mentioned above, I triangulate my results with several different methods so that this chapter is the longest of my dissertation. Many if not most of the methodological decisions concerning my study on the Canadian Shift in the speech community of St. John’s, Newfoundland, are motivated by the comparability with earlier acoustic (socio-)phonetic studies on this shift in this community. Only three such studies exist, namely Boberg (2008b, 2010), Hollett (2006, 2007) and Labov, Ash and Boberg (2006b). While the former and the latter only include six and two respondents from St. John’s, Newfoundland, respectively, which is due to their focus on the English varieties of Canada/North America as a whole, Hollett’s (2006, 2007) study is exclusively concerned with St. John’s. As Section 4.1 will show, the methodology of Hollett’s study poses some problems such as the small number of tokens. It may well be that the methodological problems are responsible for the fact that she had great difficulties in interpreting her own data, as she admits. Boberg’s and Labov, Ash and Boberg’s methodologies are much more similar to one another than Hollett’s, but the former differs slightly in some important detail from the latter, such as the inclusion of following nasals in the analysis of KIT and DRESS. As a result, I will attempt to use a methodology as similar as possible to the one used in the Atlas of North American English (Labov, Ash and Boberg 2006b), which is, of course, in large part similar to that of Boberg (2008b, 2010). These similarities begin with the make-up of my sample of respondents, the data, and end with the point of measuring the vowels, the methodology.

In terms of study design, which I outline in Section 4.2, I placed most emphasis on a balanced sample of respondents and on including respondents between the age range of 18 and more than 65 years prior to my fieldwork. In addition, I aimed at including respondents from two social classes, working class and middle class. Moreover, I decided to slightly skew my sample towards younger respondents, which will be motivated in Section 4.2. Some unexpected problems I encountered during the fieldwork forced me to focus on middle-class respondents and exclude working-class participants altogether.
Subsection 4.2.1 focuses on the structure of the sociolinguistic interview I used during the fieldwork process and the motivation behind it, i.e. the choice of the lexical items on the word list (Subsection 4.2.1.1), of the reading passage (Subsection 4.2.1.2) and of the individual sections of the interview (4.2.1.3). The interview itself does not follow the guidelines and/or modules proposed by Labov (1984) directly, due to more recent research that contested some of his propositions, as I will discuss. A more recent collection of interview questions has been suggested by Tagliamonte (2006), which I followed in an adapted form.

In Section 4.2.2, I detail the process of my sampling procedure and the resulting data set of respondents. Unlike I had originally anticipated the recruitment process, I used a combination of a judgment sample and the ‘friend-of-a-friend’ approach. The sample comprises 34 participants from a middle-class background, stratified according to age and gender (16 males and 18 females). The thresholds I used for dividing my sample into three age groups are derived from studies on social networks and communities of practice, suggesting that speakers in their late teens and early twenties are linguistically influenced by their peers. Whenever speakers tend to change their peers or friends by leaving high school, attending college or graduating from university, the linguistic influence changes. This motivation for the age group thresholds works, however, only for those years in which virtually every respondent is faced with such a change in friends/peers. As soon as they enter their professions, each speaker’s biography differs to a degree which is unquantifiable. Another possibly uniform cut-off point in speakers’ biographies may be their retirement.

The social status of my respondents is indicated via their level of education and their occupation in Subsection 4.2.2.1. In addition to those two indicators of social status, I used neighborhood to control for middle-class membership: All of my participants have been raised and/or reside in downtown and west end St. John’s, typically middle-class areas. Based on the findings by D’Arcy (2005) for local versus non-local parentage among pre-adolescent and adolescent speakers in St. John’s, I developed an index score for localness (LItotal), which consists of several factors that contribute to being local for speakers older than 17 years. The individual components of LItotal are outlined in Subsection 4.2.2.2.

Before I begin to detail the methodological features concerning the acoustic measurements of the individual vowel tokens, I will briefly outline the basics of Acoustic Theory in Section 4.3. Taken together, these basics are essential in acoustic phonetic studies in order to understand and consequently specify the settings for formant trackers in spectrograms and/or of Linear Predictive Coding analyses correctly. The Source-Filter Theory, discussed in Subsection 4.3.1, is one way of conceptualizing the production of human speech and accounts for notions such as fundamental frequency, vowel stress levels and formants. Very generally, the source in this theory are the vocal folds and the filter is the vocal tract, which are independent of one another. The vocal tract can be schematized
by tube models, which I will summarize in Subsection 4.3.2. These hypothetical tubes change the resonant frequencies in the vocal tract when they change their shape, length and diameter, for instance through lip-rounding or tongue movement. Their resonant frequencies are infinite, but the lowest three suffice to determine the quality of vowels.

An alternative account of formant frequencies is offered by perturbation theory, which relies on pressure and velocity of air particles. In order to be able to determine formant frequencies in an acoustic phonetic analysis of vowels and other sounds, Linear Predictive Coding (LPC) was developed (Subsection 4.3.3). Broadly speaking, LPC can be thought of as the inverse of the process of speech production, accounted for by, for instance, the Source-Filter Theory.

The methodological choices concerning the vowel measurements in general are based on the literature and the fundamentals of Acoustic Theory. The most important decision is the number of measurement points, which I discuss in Subsection 4.4.1. The traditional approach has been to identify the target of a vowel, usually the midpoint, and take one reading of a formant pair at this point. More recent studies in acoustic phonetics proper (not necessarily sociophonetics) take readings from either three points (onset, offset and midpoint) of a vowel, or even more temporal locations, be they time-normalized or not. For reasons of comparability with all the studies I outlined in Section 2.3, I follow the traditional approach of carrying only one point of measurement forward to the statistical evaluation. A related problem is the quantification of the single-point measurements Labov, Ash and Boberg (2006b) used in the atlas. Their technique measured each of the vowels at their formant one or formant two maximum, depending on the subsystem the vowel belongs to. I base my automatic formant readings via a Praat script (Boersma and Weenink 2011) on the findings by Evanini (2009), who evaluated alternative single-point measurements on the basis of the formant values derived from manual measurements in the atlas (Subsection 4.4.2). In addition to this single-point measurement, I include several other time-normalized temporal locations in order to clean the data (see below) but not to statistically evaluate the additional information about the vowels.

The stress level of a vowel has several acoustic correlates, which, however, do not form a uniformly predictable pattern for each vowel in each sentence, clause or point in time during connected speech. I will discuss these correlates in Subsection 4.4.3, including fundamental frequency (Subsection 4.4.3.1) and vowel duration (Subsection 4.4.3.2). Ultimately, I rely on an impressionistic determination of a vowel’s stress level, as did Labov, Ash and Boberg (2006b). In the subsection to follow (4.4.4), I provide the details underlying spectrograms and their settings I used for my analyses (Subsection 4.4.4.1) as well as those of the automatic formant tracker in Praat (Subsection 4.4.4.2). For the latter, I use an example from one of my speakers in order to outline the differences in formant readings for different formant tracker settings, i.e. LPC analysis settings. The most important issue is consistence within the analysis of one speaker, as the anatomical
characteristics of the vocal tract of one speaker do not change during the course of one interview.

In Subsection 4.4.5, I discuss the role of linguistic environments and their effects on vowel formant measurements. Some of these effects can be accounted for via choosing a certain point of measurement, but others have to be excluded or treated as a separate analytical category. With these prerequisites, I outline the vowel tokens I extracted from my sample of respondents in Subsection 4.4.6.

After I extracted all of the vowel tokens for statistical evaluation, I cleaned the data set before I normalized the vowels via NORM (Thomas and Kendall 2007), as I show in Section 4.5. Data cleaning does not only include an impressionistic or auditory confirmation of the formant measurements for each vowel, but more importantly the investigation of each of the formant bandwidths. As I will outline in Subsection 4.5.1, high bandwidths are indicative of several issues, which include, for instance, low amplitudes of the formants. A visual inspection of the spreadsheet containing the formant readings from several temporal locations within the same vowel token showed that formants with high bandwidths differed up to 150 Hz from those with small bandwidths. I consider this as an artifact of the low amplitude rather than as socially meaningful differences in vowel quality, so that I manually remeasured or excluded these tokens. The cleaned vowel measurements were normalized using the same speaker-extrinsic, Nearey-based algorithm that Labov, Ash and Boberg (2006b) used in their atlas (Subsection 4.5.2).

The final section of this chapter (4.6) is concerned with the statistical analysis of the normalized vowel tokens. Before I discuss the reasons for selecting the different methods for triangulating my results, I review the assumptions underlying the data exploration methods (Subsection 4.6.1) and statistical tests in general (Subsection 4.6.2), including linear models (Subsection 4.6.2.4) as well as generalized linear mixed models in particular (Subsection 4.6.2.5). The triangulation of findings as I intend it (Subsection 4.6.3) results in using different statistical tests for the two categories of merged vowels and shifted vowels. For the former, I use simple t-tests, ANOVAs of Euclidean Distances between two merged vowels and Pillai’s traces (Subsection 4.6.3.1). For the latter, I use a MANOVA based on Wilks’ λ in order to determine the stability of STRUT in apparent time, ANOVAs of Euclidean Distances of the front lax vowels in relation to a stable anchor or reference point in the vowel plane and significance tests for Kendall’s τ and/or Pearson’s r (Subsection 4.6.3.2). These tests statistically assess mergers and vowel shifts in apparent time only. In the t-tests and the ANOVAs for merged vowels, as well as in the Euclidean Distances for the shifted vowels, age is modeled in groups. In Pillai’s traces for merged vowels, as well as correlation coefficients for the shifted vowels, age is investigated continuously. Thus, within each vowel category, these tests serve to triangulate the findings yielded from a categorical modeling of age with those yielded from a continuous investigation of age. Moreover, both vowel categories are investigated with different statistical methods, because all have been used in the literature and each
of them has different shortcomings. After these tests have been conducted (Subsection 4.6.3.3), I use multivariate regression analysis in order to triangulate the findings of these tests concerning age once more, but more importantly, to model the other independent variables simultaneously. For the latter, the results derived from a linear mixed-effects model are triangulated with those of a logistic one. In the final subsection (4.6.3.4), I will outline how I fitted each of the regression models in terms of the assumptions these models make.

4.1 Comparability with Earlier Studies

As mentioned in the introduction to this chapter, only three acoustic (socio-)phonetic studies on the Canadian Shift in St. John’s exist: Boberg (2010), Hollett (2006, 2007) and Labov, Ash and Boberg (2006b). Hollett’s (2006, 2007) study is comprised of twelve middle-class, female respondents drawn from a judgment sample (2006: 148, 2007: 34). Out of the twelve speakers, four were recorded in conjunction with Clarke’s (1985b) sociolinguistic survey of St. John’s English in 1981/1982, and the remaining eight were recorded in 2003 (Hollett 2007: 30). The participants selected in 2003 were matched with those recorded in the early 1980s in terms of social class and age, so that Hollett arrived at three generational groups: 1) young females recorded in the 1980s (20 to 25 years of age), 2) young females recorded in 2003 (23 years of age) and 3) old females (45 to 48 years of age) recorded in 2003 that represent the same age cohort in real time as the speakers in the first generational group (Hollett 2006: 148, 2007: 30-31). The years of birth of her old speakers (groups one and three) thus range from 1955-1960 and those of the young speakers (group two) likely range from 1980-1981. The respondents I included in my data sample were born in the very same range of years, just that in 2011 the interviewees born between 1955-1960 were 56-51 years old (my old speakers) and those born in 1980 were 31 years old (my middle-aged speakers).

Unlike my acoustic analysis, Hollett’s is based on 240 word-list items only, from which she derived 868 tokens in mainly monosyllabic lexical items for kit, dress, trap, lot, strut and foot (/i, e, æ, a, ʌ, u/; 2006: 150). Except foot, these vowels were originally identified to participate in the Canadian Shift in Ontario by Clarke, Elms and Youssef (1995: 212). For the former four vowels, Hollett arrives at roughly 168 tokens per vowel and 56 tokens per vowel and generational group (2006: 150). She recorded her “traditional Labovian sociolinguistic interviews” on a Sony MD in 2003, while the interviews conducted in the 1980s were recorded on analog cassette via a Sony TC-142 (2006: 150). Although she thus compares two data sets stemming from two different recording sources, Hollett (2006: 150, 2007: 34) does not mention whether the data

---

recorded on cassette tape in the early 1980s differed in any way from the digitally recorded ones in 2003 regarding high vowels. Labov, Ash and Boberg (2006b: 36), for instance, state that normalized high vowels were lower by a difference of 30 to 50 Hz between reel-to-reel tape recordings (on a Nagra IV, a Nagra E, or a Tandberg Model 9021) and the later digital cassette tape recordings (on digital cassette tapes [DAT] using SONY TCD-D8 DAT recorders).

In terms of measurements and linguistic context (cf. Section 4.4), Hollett chose a traditional single-point measurement at the midpoint of the steady state of the nucleus, excluding only those vowels which occur before /l/ and /r/. As I will show in Subsection 4.4.5, vowels before and after glides and before nasals tend to be affected by coarticulation and/or undershoot (cf. e.g. Harrington and Cassidy 1999: 72-73; Labov 1994: 197; Labov, Ash and Boberg 2006b: 174; Thomas 2001: 52). In order to minimize such effects, vowels preceding such linguistic contexts should be excluded.

The statistical assessment of her measured vowels was conducted in SPSS 12, using ANOVAs and Tukey’s HSD post-hoc tests. The dependent variables were formants one and two; the independent variables were generational group, place, manner and glottal state of the preceding and following linguistic environment in two separate ANOVAs. Since different linguistic environments require measurements of more than one token per vowel phoneme (or observation) from the same speaker, including the latter three variables suggests a violation of the ANOVAs’ assumption of independence of observations (cf. Subsection 4.6.2.3). In addition, the large number of predictors in relation to 168 observations per vowel is very likely to reduce the validity of the results (cf. Kleinbaum et al. 2014: 485).

It may be due to these issues that Hollett’s findings are somewhat inconclusive in terms of the Canadian Shift: With regard to the low-back merger, the old females recorded in 2003 show significantly lower and more fronted realizations than both of the young female cohorts (2006: 153). This finding supports more innovative, mainland Canadian-like realizations of the merger for both of the young speaker cohorts, while at the same time it suggests that her young females recorded in the 1980s have shifted from an innovative back position of the merger towards a conservative low and fronted position in 2003. In terms of TRAP, her finding suggests a pattern opposite to what would be expected: The old females show the lowest and most retracted realizations of that vowel, i.e. the most innovative speech (2006: 155). Both of the young female cohorts only exhibit a significant lowering of ash, but no retraction (Hollett 2006: 154). Similarly, the old females lead in lowering of the DRESS vowel. Retraction of epsilon is not significantly different between any of the three generational groups (Hollett 2006: 155). KIT lowering is significant for the young and old generational groups recorded in 2003, while KIT retraction is significant for the old females recorded in 2003, i.e. the older females show significantly more retracted KIT vowels than both of the young generational groups (Hollett 2006: 156). In summary, both of the young female cohorts behave innovatively in terms of a back position of
4.1. Comparability with Earlier Studies

the low-back merger. Old females behave innovatively with regard to TRAP retraction and DRESS lowering. The old and young females recorded in 2003 behave innovatively concerning lowering of KIT.

Although Hollett tries to explain the linguistic behavior of old females in light of Boberg’s (2004b) late adoption hypothesis, her findings are in this respect inconsistent: In Boberg’s (2004b: 257) hypothesis, older speakers’ adoption of innovative features accelerates communal change in progress, “because the overall use of the innovative feature at the community level rises more quickly than would be inferred from real-time data in an apparent-time, postacquisition-stability model” (2004b: 258). That is, young speakers start using innovative variants at a certain point in time, while their parental generation does not yet do so. 25 years later, results from another apparent-time study in the same community may suggest that the new generation of young speakers still leads the adoption of new forms, but some of their parents and grandparents, who originally used the conservative forms, have joined them in the ongoing change since the original data were collected a generation ago (Boberg 2004b: 258). As Hollett outlines herself, “[e]ven Boberg’s [2004b] model of late adoption does not fit the data”, as it “does not provide an explanation as to why older speakers have the most innovative forms” (Hollett 2006: 155).

Hollett’s findings are also contrary to those of D’Arcy (2005), who analyzed the retraction of TRAP among adolescents and pre-adolescents in St. John’s impressionistically (cf. Section 2.3). She divided her interviewees into two groups: of local and non-local parental origin (2005: 332). In order to make comparability to the local status of the parents possible, I calculate a local-ness index for each speaker, as I will outline in Subsection 4.2.2.2 below, with some modifications of D’Arcy’s simple dichotomy, because the linguistic influence of the parents is only arguably stronger than that of their peers for 20-year-olds. Each of D’Arcy’s (2005: 332) groups consisted of four females, yielding a total of 16 respondents from whom she obtained 801 vowel tokens in TRAP words. She recorded her 8- to 11-year-olds and 16- to 17-year-olds during the winter of 2000, “using a ‘friend of a friend’ approach to sampling” (2005: 331). Her former age group corresponds to my young speakers (19- to 22-year-olds) and her second age group to my middle-aged ones (27- to 28-year-olds). As D’Arcy conducted her analysis impressionistically only, her findings are not directly comparable in terms of the mean formant values I use to determine shifted vowels. In her statistical analysis (stepwise logistic regression in GoldVarb), she excludes nasals, as her tokens categorically showed neither retraction nor lowering in this environment. However, D’Arcy (2005: 338) finds many retracted ashes before /g/, unlike suggested by Boberg (2010: 146) and Labov, Ash and Boberg (2006b: 223), so that her findings are incomparable to their analyses in terms of following phonological context as well.

Due to the issues mentioned with regard to Hollett’s results and the impressionistic character of D’Arcy’s study, I decided to design my study in a fashion similar to Boberg
(2008b, 2010) and Labov, Ash and Boberg (2006b), although both studies focus on supra-regional speech patterns instead of a single speech community. As outlined in Section 2.3, Boberg’s (2008b) article (and partially his 2010 book) is based on a recent study referred to as the *Phonetics of Canadian English* (PCE) and conducted at McGill University in Montreal, Quebec, Canada from 1999 to 2005 (2008b: 132, 2010: 146). This study is based on 86 participants from regions all over Canada, including six participants from Newfoundland (2008b: 133), and uses “[...] the same equipment and method of analysis as in the *Atlas* [Labov, Ash and Boberg 2006b: 36-40]” (Boberg 2008b: 134), as well as the same method of acoustic measurement (Boberg 2010: 144).

As Hollett’s, all of Boberg’s studies rely exclusively on word-list tokens in the acoustic analysis, although spontaneous speech is also part of both authors’ sociolinguistic interviews (Boberg 2010: 144, Hollett 2006: 150). For his PCE study, which focuses on the phonetic characteristics of Canada from coast to coast, Boberg outlines that word-list tokens are particularly useful for interregional analyses, because their use eliminates phonetic, prosodic, lexical and other linguistic variables from consideration (2008b: 132). The data set from each participant is thus uniform, as the crucial allophonic environment is ensured to be identical for all speakers (2010: 200).

While it is true that the linguistic context in which the vowels under analysis occur is in theory controlled for in word lists, the idea is yet too idealistic, especially when the interviewees are recorded in their homes, as Boberg outlines for his data from the *Phonetics of Montreal English* (PME) study (2005: 138).\(^{43}\) The setup of each home differs in terms of size, structure, place of recording and people present. For instance, large rooms may cause echoing effects which are recorded in addition to the sound waves of speech one is interested in. Large windows in living rooms may alter the sound waves via their own vibrations, a process which may interfere with the recording of speech. Noisy environments may be created by air conditioning, the refrigerator or other devices, if the interview is conducted in a living room with an integrated kitchen or in a kitchen. In addition, family members or other people present may introduce unexpected background noises which are being recorded simultaneously with the speech signal. Likewise, traffic passing the homes during the interview may introduce additional noises to the recording. All of these interferences may cause the recorded signal to include noises that eradicate formant measurements. These ‘dirty’ data have to be excluded from acoustic analysis (cf. Section 4.5), so that in reality using a word list does not necessarily ensure a uniform data set from each speaker regarding linguistic environments. Additionally, the more of these interferences are present during any single reading of a word list, the more tokens from that word list may have to be excluded.

\(^{43}\) Boberg (2008b, 2010) does not state where his PCE data was recorded. He does, however, mention that the full methodological details for his PCE study can be found in Boberg (2005, 2008b), which suggests that the PCE respondents may also have been recorded in their homes (2010: 144).
Another drawback of using word lists only is that the acoustic analysis does not approximate the ultimate object of description, namely vernacular speech (Boberg 2010: 200). In addition to producing ‘artificial speech’, reading words in their citation form can be confounded by a lot of other effects such as reading difficulties, resistance to reading while being recorded, hypercorrection, priming, reading speed, overemphasis, etc. Eliciting vernacular speech as used in conversations with close friends and relatives is also an idealized goal in sociolinguistic interviews; spontaneous speech (i.e. the combination of casual and careful speech style; cf. Labov 1989: 11 and Subsection 0.3.4) is, however, more than realistic to elicit in interviews and precisely the data on which Labov, Ash and Boberg’s (2006b) Atlas relies (Boberg 2010: 200). Spontaneous speech close to the vernacular speech of the respondents can be more easily elicited when the respondents are recorded in an environment with which they are familiar, e.g. their homes.

In both of his articles, Boberg included the six lax vowels of Canadian English, /i, e, æ, ə, a, u/, in his acoustic analyses (2005: 138, 2008b: 135). “They were recorded on Type II (CrO2) analog cassette tapes using Marantz PMD 221 cassette recorders and Audiotechnica AT 803b omnidirectional lavalier microphones” (Boberg 2008b: 133). The word lists include these vowels at least once before /d/, /t/, /r/, /l/ and /n/ in his PCE data and add up to a total of 145 vowel productions per speaker (Boberg 2008b: 133). Although he explicitly states that his method of analysis is the same as that of Labov, Ash and Boberg (2006b: 36-40), he does not state whether he excluded vowels before nasals (for KIT, DRESS and TRAP) and liquids as well as after glides and liquids as Labov, Ash and Boberg (2006b: 77) did.

For his PME data, he states that over one thousand of the six lax vowels as produced by 35 speakers were acoustically analyzed. According to Boberg’s footnote nine in his (2005) article, the vowels occurred at least once before /d/, /t/, /r/ and /l/; except for the FOOT vowel, each also occurred once before /n/. TRAP before nasals and /g/ as well as all the vowels before /r/ were excluded from the analysis, because the former is often raised and fronted in North American English and the latter form a subsystem separate from the main, non-pre-rhotic system in North American English. In terms of the main vowels of interest regarding the Canadian Shift – KIT, DRESS, TRAP and LOT (merged LOT-THOUGHT) – he additionally included tokens before /k/ and /p/. In total, Boberg thus analyzed six tokens of KIT, DRESS and LOT, five tokens of TRAP, four of STRUT and three of FOOT (2005: 152-153). His method of analysis thus differs from that of Labov, Ash and Boberg (2006b: 77), because Boberg did most likely not exclude KIT and DRESS in pre-nasal environments. As outlined above, vowels before nasals may be affected by coarticulation and should be excluded (cf. e.g. Harrington and Cassidy 1999: 72-73; Labov 1994: 197; Thomas 2001: 52).

The pre-nasal position of KIT and DRESS vowels is additionally problematic, as in some North American regions such positions condition a merger between the two vowels. Although the pin-pen merger has been predominantly attributed to the southern dialect
areas of the United States, Labov, Ash and Boberg (2006b: 220) state that it is continually spreading northwards by diffusion. The northermost attestation of the merger was made for Saint John, New Brunswick, and Halifax, Nova Scotia (2006b: 220). Since Boberg does not include following linguistic environment as a predictor in his statistical analysis of his PCE data (MANCOVAs in SPSS; 2008b: 134, 2010: 202), he can thus not exclude any possible effects of the merger on those vowels, although his PCE data includes 16 respondents from those two regions (Boberg 2008b: 133).

Apart from these minor, but important, differences between Boberg’s methodology and that of Labov, Ash and Boberg (2006b), his is quite similar to the latter. In light of both their analyses, Hollett’s (2006, 2007) methodological approach seems to produce results which are much less reliable. For this reason, I decided to use a methodology similar to Labov, Ash and Boberg’s (2006b), which I will outline in the remaining sections of this chapter.

4.2 Study Design

I designed my study based on the insights I gained from the literature and the pilot study I carried out prior to my fieldwork in St. John’s, Newfoundland, in 2011. The pilot study data was gathered by interviewers who were well-known and appreciated members of the communities of St. John’s and Pouch Cove. The latter is situated north of St. John’s and used to be a small, traditional fishing town. Since the cod moratorium in the 1990s, it has developed into a small bedroom community of St. John’s with approximately 2000 inhabitants (cf. Statistics Canada 2012: n.p.). The community members interviewed respondents from the two locales in their homes in 2006 as part of a local-heritage project conducted by the Department of Folklore at Memorial University of Newfoundland. In 2011, I contacted several members of the project in order to inquire about access to some of the interviews of St. John’s residents. One of the members granted me access to the collection of interviews, so that I was able to extract a subsample from the more than 80 interviews conducted in St. John’s for my pilot study prior to the fieldwork. I extracted three young females, one middle-aged female and one old female and analyzed a total of 643 vowels in connected speech from these approximately 40-minutes interviews. They did not follow the make-up of traditional sociolinguistic interviews, but they had the great advantage of providing me with valuable cultural and social insights into the speech community.

Although the quality of the recordings was rather poor for acoustic analyses, it was better than the quality of the recorded telephone interviews Labov, Ash and Boberg (2006b: 36) used for their atlas. For instance, the folklore interviews were recorded on a Sony ICD-P210 digital recorder using a Sony ECM-DS 70P unidirectional, electret condenser microphone. The sampling frequency of the recording device is 8000 Hz, which corresponds to a Nyquist-frequency of 4000 Hz. The telephone lines permit a frequency
4.2. Study Design

<table>
<thead>
<tr>
<th>Age</th>
<th>18-25</th>
<th>25-60</th>
<th>65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Middle Class</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Working Class</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Total per Age Group</td>
<td>20</td>
<td>20</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 4.1: Age, sex and social class of individual participants in the planning stage of the fieldwork (n = 52).

A bandwidth of 300 to 3000 Hz (Labov, Ash and Boberg 2006b: 36). A Nyquist frequency of maximally 3000 Hz may not be able to capture second formant values for high front vowels.

Prior to my arrival at the fieldwork site, I made arrangements with one of the interviewers of the heritage project to introduce me to her respondents as a starting point for further recruitment of participants for my study. In the first week of the six-week fieldwork period, we met in order to discuss the details of a collective initial recruitment phase and the structure of my sociolinguistic interviews, which are relatively fixed compared to her open interviews. She readily offered me to conduct co-interviews with her respondents in order to introduce me to the speech community. In this planning phase, I designed a speaker matrix as outlined in Table 4.1. I considered it realistic to conduct 52 interviews in six weeks, provided that most of the interviewees are recruited by my co-interviewer. Unfortunately, a week later on a Saturday morning when we had an appointment to meet some of the first respondents, she denied me any access to her interviewees without any reason: “I am sorry, I do not know enough people to be of any help to you”.

This new situation in the beginning of my fieldwork made the prospect of interviewing more than 50 respondents rather unlikely, so that I decided to focus on those respondents whose social profile is most important for my study. Age is the most crucial variables in my analysis since the literature largely suggests absence of the Canadian Shift in the speech community of St. John’s (cf. Chapter 3). That is, if the members of the community do participate in the shift today, given the social and demographic changes since the late 1990s (after Clarke conducted her St. John’s survey in 1991), the shift should be found predominantly among young speakers. Within this group, the females from the interior groups of the socioeconomic hierarchy are crucial (cf. Labov 1994: 78; 300, and more recently Baranowski 2013a: 275 on mergers): As mentioned in Sections 3.3 and 4.1, Clarke (1991, 2004b) and D’Arcy (2005) suggest that the young urban females in general and the pre-adolescent as well as the adolescent females in St. John’s in particular showed retracted TRAP (and lowered DRESS) vowels, which may be due to the influence of mainland Canadian English (the Canadian Shift). Hollett (2006, 2007) puts emphasis on two young middle-class female cohorts, one recorded in the 1980s and another recorded
in 2003. In addition, Labov, Ash and Boberg’s (2006b) data for the atlas are skewed towards participants from the middle to upper middle class (cf. Section 2.3).

Labov’s observation (1994: 78, 300) that young middle-class females lead in the adoption of prestige variants holds predominantly for changes from below the level of social awareness. Clarke (1991: 118, 2012: 514), however, emphasizes that the innovative mainland Canadian forms may enter the linguistic system via the formal styles, i.e. there is a social awareness of the linguistic innovations in the speech community of St. John’s. Additionally, Clarke states that, for instance, innovative TRAP retraction is introduced by the upper-class females (2012: 514). Yet, as outlined above, all of the acoustic phonetic studies including respondents from St. John’s are skewed towards young urban females from the middle class, and one of them suggests the presence of the Canadian Shift in St. John’s: Boberg (2008b).

It should be apparent from the studies on the Canadian Shift presented in Section 2.3 that the role of age is undisputed and that of social class ignored. The role of sex, however, should be considered as contested, since those studies that include females exclusively can of course not establish an effect of gender (e.g. Hollett 2006, 2007). I thus decided to skew the collection of informants towards young speakers from the middle of the socioeconomic hierarchy, but to aim at a balanced gender distribution. In terms of the youngest age group, the time span for the twenty speakers from the lower middle and upper working class of the envisioned sample covers only seven years (cf. Table 4.1). The time span for the twenty middle-aged speakers covers 35 years, which corresponds to a smaller net amount of speakers in relation to the ages they are supposed to represent. In the oldest age group, I planned to include the fewest speakers, since this cohort primarily balances the other two cohorts regarding the non-participation in the Canadian Shift. Put differently, if the Canadian Shift has very recently begun to enter the speech community (despite the claims made in the literature), it would most likely be found among the youngest speakers of the social matrix, but not the middle-aged ones, so that the oldest speakers would be redundant. I decided that the risk of excluding the oldest speaker cohort completely is too great to take and consequently include fewer old speakers per cell (3 versus 5 in the other two cohorts). This decision was additionally governed by the monetary and temporal constraints under which I would conduct my fieldwork.

Generally, the categorization of the participants into the respective age groups is based on the literature (cf. e.g. Cheshire 2004: 1553; Llamas 2007: 71; Labov 2001a: 170; and cf. Subsection 0.3.6). With regard to the youngest age cohort, my reasoning follows the contributions made, for example, by Eckert (1998). She maintains that adolescents construct their identity independently of their elders (1998: 163). Her work at Belten High has shown that teenagers in particular conform in their linguistic behavior and social identity to the fellow members of their community of practice (Eckert 2000).

It thus seems plausible to categorize young speakers in particular according to their social environments/networks. The circles of friends young speakers have vary according
4.2. Study Design

to the developmental stage they are in. For instance, 15- to 18-year-olds have relatively stable communities of practice or social networks when they are in high school. Their circle of friends usually radically changes when they move to the communities in which they attend universities, roughly at the age of 19 to 25. While adolescence (e.g. high school students) is usually considered to be the focal point of linguistic innovation and change (Chambers 2009: 189), I exclude speakers below the age of 18, because the complete familiarity with local speech norms is not yet acquired at such an age (Labov 1972b: 138). Arguably contradictory, Labov maintains that the phonological system of respondents younger than 20 years of age is stable (1994: 111-112). In fact, Bailey (2002: 319) and Cukor-Avila and Bailey (2013: 246) have shown in a large quantitative study comprising more than 1000 participants that the vernaculars of respondents below the age of 20 are not stable. They are thus unsuitable for apparent-time studies, as the differences between their speech patterns in comparison to those of older respondents do not allow inference of being temporally analogous, i.e. the linguistic behavior of subjects younger than 20 does not necessarily indicate communal change, but unstable vernaculars (cf. Chambers 2002a: 360-370, 2009: 207, 2013a: 310-312).

Since every teenager has to attend school and the majority of middle-class adolescents attend universities, such a categorization is quantifiable. However, the careers of each university alumni differ on an individual basis, i.e. one person may find a career job for a long period of time with relatively stable social networks, another may have to apply for a new job every other year, resulting in variable social networks and thus peer effects. As Eckert puts it, adults move through their life trajectories as individuals and at the same time as part of an age cohort (1998: 151). For the relatively long time span of adulthood, a quantifiable categorization into smaller groups is not feasible, although the observed linguistic variation across large age cohorts may not be easily explainable (Llamas 2007: 73).

In a similar way, the majority of old speakers in North America does not retire at the age of 65 in a uniform fashion as is the case in, for instance, Germany at the age of 67, so that their retirement from constraints of the linguistic marketplace can only arguably be quantified. The concept of the linguistic marketplace states that the perception of the prestigious variants adults use as socially legitimate exerts the strongest effect on their speech (e.g. Ash 2013: 359; Chambers 2009: 190-191). In their retirement, old speakers slightly alter their speech towards linguistic variants that are either local or carry covert prestige as a sign of retirement from the marketplace, but remain largely stable or consistent in the choices they have acquired during their formative years (Chambers 2009: 197). Especially in the case of Newfoundlanders, linguistic marketplace pressures on speakers toward standardization of their linguistic choices are extremely high – at least they have been experienced as such by the majority of my respondents. In the planning stage, I chose the age of 65 as a marker of the end of the social constraints of the linguistic
marketplace, as I thought my co-interviewer can provide me with access to respondents of that age and older.

While all of these contemplations remain unaffected, I was not able to recruit respondents in exactly this fashion after I was denied aid from my co-interviewer. As I will outline in Subsection 4.2.2, I used a judgment sample and a friend-of-a-friend approach (cf. e.g. D’Arcy 2005: 331) to recruit respondents who matched the initial conception of the social matrix as I outlined here.

4.2.1 The Sociolinguistic Interview

The order of the parts of a traditional sociolinguistic interview has mistakenly been understood as fixed by some scholars (e.g. Trudgill 1974): in the beginning a casual interview, at the end a formal word list and a less formal reading passage (also cf. Rickford 1986). This understanding might be based on the perception of interviewees as reactive to the interview situation. Contemporary research in Variationist Sociolinguistics (VS) has shown that interviewees interpret each of the situations individually and attribute meaning to them, i.e. each of the styles they perceive to be appropriate at any point in time in an interview is performed regardless of the interview’s structure (cf. Milroy and Gordon 2003: 50; Schilling 2013: 332; Schilling-Estes 1998: 75, 2008: 971; and Subsection 0.3.4).

If speakers do define the formality in an interview situation proactively for themselves, depending on the relationship to the interviewer (e.g. Rickford and McNair-Knox 1994) and the task in the interview (e.g. Cedergren 1973; Sankoff 1974), they will have a range of stylistic variation in their repertoire and decide for themselves how much of this range they are willing to use. The interviewer can increase and decrease this range by asking the interviewee to do or say different things: When they are asked to mimic the typical speech of a rural Newfoundland ‘bayman’ or working-class ‘townie’ and decide to do so, they will not employ their most formal style; when they agree to read a word list, they are not very likely to employ their most informal style. If no such favors are asked, their stylistic range is less affected by the tasks of the interview, but is likely to be more affected by the interpersonal relationship of interviewer and interviewee. Since it is a VS commonplace to assume that speakers use their most vernacular styles with close friends and family, but not necessarily with strangers, it is highly likely that they will place themselves at a distance within the unfamiliar situation of talking to an unknown interviewer via use of more formal linguistic styles. Depending on the development of the interview with regard to this interpersonal relationship, they might feel more familiar with the interviewer, build up trust to share personal details and may even enjoy the mere fact of having someone to talk to after some interview time has passed.

This assumption is supported by, for instance, Hall-Lew (2009: 132), Hoffman and Walker (2010: 53) and Van Herk and Knee (2013: 34), who all began their acoustic analysis after the first 15 minutes of the interviews, and by Labov’s “danger-of-death”
question (cf. Labov 1984: 35), in which the emotional involvement in personal life-threatening events was thought to distract the interviewees from the interview situation so that they use their vernacular to describe such events. The assumption is also supported by the personal narratives (cf. Labov 2010) my interviewees chose to tell me towards the middle of the interview: one talked in great detail about how she came to donate eggs to a friend who could not conceive; another revealed that she carried a child for a befriended couple who had the same problem; others told me about their open marriage and how “easy” Newfoundland women were, even when they have children; still others talked about their continuous misuse of drugs and their bi- or homosexual orientation; another interviewee revealed thoughts and experiences about the way religion formed their life; older interviewees told how they were forced by teachers to learn “standard” English in school and how they were punished and excluded from social life at school when they did not speak the putative standard or not as well as others – clearly no conversational topics for indicating interpersonal distance.

In other words, if the formality of the interview situation is in part a function of time and of the tasks of the interview, then in the opposite direction as implied by Trudgill (1974). This means that an interview is more formal in the beginning and less so at the end (or: interviewees accommodate throughout the interview), because of the interpersonal distance of interviewee and interviewer, and because of the unfamiliarity of the interviewee to the interview situation. This suggests that tasks designed to elicit formal styles can also be conducted in the beginning of the interview, not solely at the end. Further, if interviewees can switch in and out of performance style (i.e. performing any style they consider appropriate in a certain situation, including authentic imitations of local dialects) at any point in the interview, as empirically tested and suggested by Schilling-Estes (1998) and others (e.g. Bigham 2013), the order of the tasks in an interview does not matter at all. This means that comparability with other studies is not hindered, particularly in light of Boberg’s procedure in which he had his respondents read the word lists before the conversational parts of the interviews (2008b: 133, 2010: 144).

Due to these findings and assumptions, I decided to begin the interview tasks with the word-list section, followed by the reading passage and other tasks designed to elicit gradually less formal speech. The intermediate-to-last sections of the interview contained open questions concerning family, friends and personal lives of the interviewees.

4.2.1.1 The Word List

The conceptualization of the word list was governed by several restrictions. My acoustic phonetic study on those vowels shown to participate in the Canadian Shift elsewhere in Canada is primarily based on vernacular speech, or more accurately on spontaneous speech, since such a study has not yet been carried out in this speech community (Hollett 2006: 149). The St. John’s sociolinguistic survey, conducted by Clarke (1985a,b, 1991) from 1981 to 1982, also relied predominantly on casual or free conversation style
(1991: 111), but was conducted prior to the cod moratorium in the 1990s that caused profound social and economic changes in the speech community (cf. Subsection 3.2.6). In addition, Clarke, also concerned with phonological change toward mainland Canadian English heteronomy (1991: 113), analyzed her phonological variables impressionistically, not acoustically (1991: 112). In Eckert’s terminology (2005: 1-3), it was a typical ‘first wave’ study, focusing on a stratified random sample of 120 subjects, representing different ages, sexes, socioeconomic statuses and religions, and on a range of styles, namely minimal pair, word list, reading passage and free conversation (Clarke 1991: 111).

Since I am primarily interested in spontaneous speech, I decided against differentiating the most formal style (Labov 1972b: 84) even further into elicited versus read word lists versus minimal pairs. Labov, Ash and Boberg (2006b: 32) relied, for instance, on elicitation of a random list of clothes, breakfast items, farm animals, days of the week and numbers from one to ten. The effort of eliciting items of interest in such general categories as the former three examples would have taken too much time in relation to the time I wanted to invest in free conversation. In addition, such elicitation does not yield vowels in the same lexical items, i.e. linguistic environment, for all speakers, which should be one of the advantages of word lists over connected speech (cf. Boberg 2008b: 132, 2010: 200) – at least in theory (cf. Section 4.1). In case of the latter two examples, days of the week would have included one stressed vowel of interest for my study (Saturday) and counting from one to ten would have yielded only two tokens of interest (six and seven). I consider the net amount of three vowel tokens in relation to 17 tokens in total as too inefficient. Furthermore, combining all of the tokens of interest from elicited word lists (an unknown number in the former cases with three in the latter) would have yielded a total number too small to contrast with a read word list, in which several tokens per vowel in different linguistic environments are included.

Due to these contemplations, I decided to have my respondents only read a word list, including minimal pairs, and a reading passage – also an established tool to cause stylistic differences in vowel productions – in order to introduce a potential stylistic profile into my analysis (cf. Clarke 1991: 111). In addition to the time constraints mentioned above, the intended style continuum also poses some restrictions to the word list. From a scientifically sound perspective, a direct comparison of the three lax Canadian Shift vowels requires word lists that contain (in)frequent, (but) comparable minimal sets (e.g. bad, bed, bid). The minimally three vowel tokens in the same lexical items (for vowels not undergoing change), as generally included in word lists (Di Paolo, Yaeger-Dror and Wassink 2011: 89), cause an imbalance in token number between word-list, reading-passage and interview style, with most tokens in the latter (but governed by lexical frequency). In order to allow for a direct, balanced comparison between styles and for generalizable results (e.g. for a subset of shifted vowels in monosyllabic lexical items only), a statistical analysis could include, for instance, 500 vowels in bad, bed and bid in word-list style, 500 vowels in bad, bed and bid in reading-passage style and 500 vowels in bad, bed and bid in interview style.
4.2. Study Design

For arriving at a total of 1500 of these vowels in the analysis, each of my respondents (34) would have to read at least 45 word-list items.

While this number is reasonable, it poses the following problems: (1) The word list would thus include only front vowels before /d/; adding other linguistic contexts such as /t/, /b/, /p/, /g/, /k/, /z/, /s/, /tS/ and /dS/ would add an additional 495 lexical items to list. Moreover, the tokens in similar environments of the other lax (and tense) vowels and the tokens in dissimilar environments needed to establish the vowel space would result in a total number of lexical items on the word list exceeding 1000 by far. (2) Such a word list would not include bisyllabic lexical items “with stress on the vowel to be measured” (Di Paolo, Yaeger-Dror and Wassink 2011: 89). (3) Controlling the preceding linguistic environment in a fashion similar to the following one does not necessarily introduce more lexical tokens to the word list, but further problems regarding the suitability of any corpus in terms of predictability of lexical frequency and thus of occurrence of lexical items in spontaneous speech (cf. Footnote 44 below for an example). This is also problematic regarding following context only, as outlined above, since the balanced sample does not reflect the difference in natural occurrences (lexical frequency) between these three lexical items (bad, bed, and bid). The suitability of a corpus largely depends on its representativeness and balance of included texts and thus to some extent on its size: For instance, the English portion of the CELEX (Centre for Lexical Information), as suggested by Phillips (2011: 180), consists of 1.3 million spoken words. Since it is not clear which conversational styles/levels of formality are included in this small corpus, it seems to be a doubtful source regarding lexical frequency information in general (also in comparison to the spoken data in large reference corpora such as COCA). Similarly, tables of consonant and vowel frequencies such as those offered by Crystal (1995: 239, 242) for conversational RP are useless if the the type of conversation (small talk in a bar versus church, etc.) and the transferability to St. John’s, Newfoundland, is not clear or testable. Using lexical frequency for determining suitable lexical items for the word list (rather than relying on one’s subjective assessment) is further complicated by the differentiation of linguistic context and the function of lexical items in spontaneous speech. Function words are usually much more frequent in connected speech than content words, but the former are either not or only emphatically stressed. (4) I could not find a reading passage that contained 15 stressed vowels of the same lexical items, not to mention the remaining contexts or vowels. (5) Although bad is a common monosyllabic word, it occurred only 14 times in my interviews, which are on average 90 minutes in

\footnote{A COCA search of \textit{bad} as an adjective, conducted on March 29, 2014, returned 292 occurrences per one million words in the spoken component (Davies 2008-). The lexical item \textit{bed} (as a noun, excluding the plural form) occurred 56 times per one million words, and \textit{bid} (as a verb and noun, excluding plural and third person singular -s) occurred less than 14 times per one million words. That is, \textit{bad} is 20.9 times more frequent than \textit{bid}, and \textit{bed} is four times more frequent than \textit{bid}. While this suggests a certain ratio for including a respective amount of these lexical items in the word list, matters are complicated by the make-up of COCA’s spoken section (i.e. it is not based on sociolinguistic interviews with topics similar to the ones in my data) and the suitability of COCA}
length. *Bed* and *bid* are even less frequent and consequently occur even less often in my interview data. Furthermore, the phonological environments in which /æ/ and /ɪ/ occur in interview data in general are not exactly the same, and neither is the ratio between these environments across the vowels, provided the same environments exist: *pad, ped* (uncommon), *pid*.

If the reading passage and interview data do not contain (a similar number of) lexical items similar to the word list, the formant frequencies measured in the latter are not contextualized. The difference in formant frequencies in cleaned vowel data can thus not be attributed directly to a stylistic difference. Word-list data may additionally be confounded, for instance, by effects of priming, reading difficulties, emphatic stress, etc. Having respondents read more than 1000 words plus a reading passage most likely results in not contributing much more speech in conversational style.

Although Di Paolo, Yaeger-Dror and Wassink also suggest to “[... ] either limit [one’s] study to just one clearly defined style [...], or select the same amount of tokens for each word class in each style represented in [one’s] recordings [... ]” (2011: 89; my emphasis), their proposition is problematic in addition to being impractical as outlined above. None of the word lists that were used to study the Canadian Shift included articles or infinitival markers, so that the analyst does not know whether they can select, for instance, the lexical item *dress* in interview style in a phrase along the lines of ‘get dressed’ or in a phrase like ‘got a new dress’ to represent the lexical item *dress* in the word list. Or alternatively, the analyst has to look for identical amounts of epsilon in both of these types of phrases, i.e. the analyst has to include “the same number of” /ɛ/ vowel tokens in *dress* as a verb and as a noun.

Since aiming at such a stylistic continuum is impractical for my study due to the reasons outlined above, I designed my word list primarily in order to contain each of the vowels of North American English in different, near-randomly-sampled, common linguistic environments for normalization (complete vowel space, except for diphthongs) and only secondarily for a stylistic comparison. For the latter purpose, I refer additionally to and primarily rely on the findings of other studies such as Clarke’s (1991, 2012) ‘first wave’ study. Furthermore, inclusion of linguistic environment as a fixed effect and lexical item as a random effect in the statistical analyses accounts for the correlation of vowel quality and lexical item (cf. Subsection 4.6.3; Johnson 2009: 373).

My word list thus consists of 125 lexical items, a reasonable number of words to read within a sociolinguistic interview focusing on spontaneous speech. In order to include phonological processes which may influence the St. John’s vowel system, I rely on Clarke’s observations (1991, 2004b, 2010a, 2010b) outlined in Section 3.2 and Labov, Ash and Boberg’s atlas (2006b). In terms of lexical sets to include in the word list (and later in

---

for determining the lexical frequency of these items in a speech community inhabited by former Irish and English settlers. Additionally, compilation of the BNC ceased in 1993, so that it may have become inaccurate within the past 20 years.
the analysis), I use an adapted version of the vowel space protocol for North American English proposed by Di Paolo, Yaeger-Dror and Wassink (2011: 88), which can be found in Appendix A.2. They also suggest to use vowels preceded by the voiceless glottal fricative /h/, since the articulators (e.g. lips, tongue and teeth) are in a neutral position for this segment (2011: 88), i.e. they are in position for realization of the following vowel (Ladefoged 1996: 112). However, since such lexical items virtually never occurred in the free speech of the respondents in the pilot study, I include some consonants produced in alveolar position instead (Di Paolo, Yaeger-Dror and Wassink 2011: 88).

Like Labov, Ash and Boberg (2006b: 77), I exclude vowels in unfavorable linguistic environments, namely vowels before nasals (for KIT, DRESS and TRAP), voiced velars (for TRAP) and liquids as well as vowels after glides and liquids from the analysis (cf. Subsection 4.4.5 for details). I include them, however, in the word list for the purpose of normalization (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 88). In addition, vowels in multisyllabic words (more than two syllables) are not controlled for (excluded) in the word list as suggested by Di Paolo, Yaeger-Dror and Wassink (2011: 89), because I include number of following syllables as a predictor variable in my statistical analyses (cf. Labov, Ash and Boberg 2006b: 220). Since the literature on the speech community of St. John’s largely confirmed the presence of the low-back merger (e.g. Clarke 2010a: 30; Hollett 2006: 153; Labov, Ash and Boberg 2006b: 217) and the retraction of TRAP (e.g. Clarke 2012: 514; D’Arcy 2005: 338), emphasis is placed on the remaining two front lax vowels participating in the Canadian Shift (KIT and DRESS) and the back lax vowel STRUT. The lexical items containing all of these vowels were determined via Kenyon and Knott’s (1953) A Pronouncing Dictionary of American English (cf. Phillips 2011: 180).

The choice of tokens on the word list was additionally constrained by the fact that I was an affiliate of the Sociolinguistics Laboratory at Memorial University of Newfoundland, which ultimately meant that I would recruit most of my young respondents from the Sociolinguistics Lab and the Department of Linguistics. That is, some of my respondents are Master and Ph.D. students of linguistics and thus familiar with the type of interview and to a small degree with the literature on mainland Canadian English (phonology). In order to distract them from the core phonological features I am interested in, I use the lexical items including vowels before /r/ (in particular SQUARE) as tokens to distract them from those that participate in the Canadian Shift. Additionally, I did not reveal details about the nature of my study, regardless of the participants’ education (cf. Schilling-Estes 2007: 178).

In terms of the main vowels of interest in the Canadian Shift and STRUT, the word list contains six tokens of /ɔ/, seven tokens of /a/, ten tokens of /ʌ/, nine tokens of /æ/, 14 tokens of /ɛ/ and 14 tokens of /ɪ/ in mono- and multisyllabic lexical items. Each of these vowels occurs predominantly before /t/ and /d/, but also before /f/, /v/, /k/, /g/, /p/, /s/, /ʃ/, /ʃ/, /θ/ and nasals. KIT and DRESS also occur before /l/. These vowels constituted 60 tokens in total, almost 50 per cent of all tokens.
Since the pilot study indicated that FLEECE is stable in apparent time, and since the vowel is realized in the corner of the vowel space (Di Paolo, Yaeger-Dror and Wassink 2011: 88), I additionally included eleven tokens of /i/ (/iy/) in the word list. With the putative stability of this vowel in apparent time, I intended to use it to calculate the Euclidean Distance metrics (but cf. Subsection 4.6.3.2). Although of no direct concern for this thesis, I also included ten tokens for GOOSE and five for GOAT, since, among other features, Labov, Ash and Boberg (2006b: 146) define mainland Canada as a dialect region by the fronting of GOOSE, while GOAT is retained in a back position. In addition, the word list contains three to five tokens for the remaining vowels, except diphthongs. The full word list can be found in Appendix A.1.

4.2.1.2 The Reading Passage

Reading style differs with the choice of the reading passage included in a sociolinguistic interview, since such passages can stem from numerous different genres ranging from news reports to poetry. The aim of a reading passage in the sociolinguistic interview is “to close the stylistic gap between speech and reading by writing texts that are more animated and colloquial” (Labov 2006: 60), and “to standardize the context towards the informal end of the possible range [of reading styles]” (Labov 2006: 61). In order to achieve this informality in reading, sociolinguistic studies usually include the reading of a funny fictional passage (Di Paolo and Yaeger-Dror 2011a: 12).

Labov describes the traditional reading passages devised by dialectologists, such as Grip the Rat, Arthur the Rat and The North Wind and the Sun as “rather painful assemblages of words of interest and evoke the most formal of reading styles” (2006: 60). Because of this and because it contains a relatively large number of the lax front vowels KIT and DRESS, I chose Comma gets a Cure (Honorof, McCullough and Somerville 2000). The reading passage has been used in a number of recent publications, including acoustic phonetic ones (cf. e.g. Kanjee et al. 2010; Scobbie et al. 2013; Turk et al. 2010) and, according to Honorof (2003: 120), was devised to include all of the vowels of General American English in several linguistic environments. It revolves around the anxious behavior of a goose named Comma at the veterinary’s office.

The reading passage is comprised of eight tokens of /æ/, eleven tokens of /a/, three tokens of /ʌ/, twelve tokens of /æ/, 20 tokens of /ɛ/ and 15 tokens of /I/ in mono- and multisyllabic lexical items. The following linguistic environment of these vowels is similar to the ones outlined for the word list above. In addition to those, one token of /æ/ also occurs before a glide, and one token of /æ/ is followed by /d3/. Only in case of lax front /I/, a following lateral is included in the reading passage. All of the Canadian Shift vowels occur at least once before a nasal. In addition, five lexical items are (emphatically) stressed function words.

The ratio between lexical items of the three front lax vowels, TRAP, DRESS and KIT, and other vowels is particularly in favor of epsilon in this reading passage. Other vowels I
included for the purpose of normalization are those outlined above for the word list. The emphasis or (sentence) stress readers placed on the individual items naturally varied on an inter-speaker level, so that I could not extract a uniform amount of vowels for all of the readers included in the analyses (cf. Subsection 4.4.3). The ratio between the extracted vowels is, however, similar per speaker. The full text of the reading passage can be found in Appendix A.3.

4.2.1.3 The Interview

For the planning of the interview questions I did not primarily rely on Labov’s (1984), Milroy’s (1987c) and Sankoff’s (1974) field work descriptions due to their partly outdated character (cf. Tagliamonte 2006: 17). For instance, the original *sine qua non* of the sociolinguistic interview was to elicit the most vernacular speech style from the respondents (cf. Subsections 0.3.4 and 0.3.7). It is thus the same speech style in which the influence of monitoring one’s speech is minimal (Labov 1972b: 208), and which is only minimally affected by a prestigious supra-regional standard, possibly imperfectly learned in literacy-inducing institutions such as schools (Milroy 1987b: 58). The vernacular is the natural speech respondents would use when speaking to very close friends and family in their most casual manner (cf. Kerswill 2004: 27). Consequently, the goal of the interview was to employ means to get past more careful speech styles such as answering interview questions. This objective is also referred to as the Observer’s Paradox (Labov 1972b: 209; more generally as the Hawthorne effect; cf. Chambers 2009: 19), which states that speech linguists observe in such interviews should be the one respondents use when they are not being observed. According to Tagliamonte (2006: 37), referring to the sociolinguistic interview as ‘interview’ is thus a misnomer, because it should be anything but an interview (i.e. the respondents should simply talk continuously; Tagliamonte 2006: 46).

As I have discussed in Subsection 0.3.4, one of the means Labov suggested to use as indicators of casual speech were channel cues. These included, for instance, laughter or varied pitches and were interpreted as signs of relaxation and thus natural linguistic behavior on the part of the respondent. Contemporaries of Labov (e.g. Wolfram 1969; Macaulay 1977) rejected the idea of channel cues as identifiers of casual speech, because their use was indeterminate (Wolfram 1969: 58-59). That is, laughter cannot be applied in an objective and reliable manner across different studies, because it can mean numerous things apart from relaxation. In fact, it can be indicative of the opposite of relaxation, namely nervousness or stress. Apart from their questionable meaning, channel cues are also impractical when used in such a way, as the immediacy of the vicinity of linguistic feature and respective channel cue is far from determined. In addition, using only combinations of channel cue and linguistic feature most likely results in an amount of tokens too small to invoke statistical measures (cf. Rickford and McNair-Knox 1994: 238; Schilling 2013: 332; Wolfram 1969: 58-59).
Another means to elicit the most casual speech of the interviewee is the use of modules which consist of such questions that would move the analyst gradually closer towards the vernacular. Labov (1984: 33-34) defines the interview as a series of hierarchically structured sets (modules) of questions designed to progress from general, impersonal, non-specific topics and/or questions to more specific, personal ones (Tagliamonte 2006: 38). The modules cover several different topics, ranging from demography (Module 1) to more informal topics such as family, friends, school/work, religion, danger of death (Modules 9, 11, 15/16, 10, 6) and back to the more formal topic language (Module 20; Labov 1984: 35).

While the danger-of-death question and similar ones Labov used in his interviews to elicit “narratives of personal experience” (Labov 2010) are still considered to be ‘optimal’ by some (e.g. Tagliamonte 2006: 38), research in different speech communities by others such as Chambers (1980), Feagin (1979), Macaulay (1977), Milroy (1980) and Trudgill (1974) has shown the failure of this question for different reasons (cf. Feagin 2002: 30, 2013: 30; Gal 1979: 94; Macaulay 1999: 13). Gal (1979: 94) and Macaulay (1999: 13), for instance, found that if such questions did elicit narratives at all, they were no different in style than questions about the first job of the interviewees. Gal (1979: 94) even suggests that instead of provoking casual speech, such “emotion-laden” questions caused her interviewees to use even more standard speech in order to convey the seriousness and to ensure the interviewer understands the contents of the narrative (this observation supports the notion of proactive interviewees rather than being reactive solely to the emotion-laden topic). Feagin stresses that danger-of-death questions may rather cause feelings of irreverence for some, especially elderly, respondents (2002: 30, 2013: 30). Such feelings inherently carry the potential of harming the respondents psychologically. Since my study had to be approved by the Interdisciplinary Committee on Ethics in Human Research at Memorial University of Newfoundland (MUN) prior to the beginning of my fieldwork, I could not and did not want to include such questions per their guidelines.

The following excerpt is taken from Module 6 (danger of death) of Labov’s generalized set, Q-GEN-II, from which the interviewer is to construct an interview schedule (1984: 33):

**Have you ever been in a situation where you were in serious danger of getting killed (where you said to yourself, “This is it!”). What happened?

**Some people say in a situation like that, “Whatever is going to happen is going to happen”. What do you think?

**In most families, there’s someone who gets a feeling that something is going to happen, and it does happen. Is there anyone like that in your family? Do you remember anything like that in your family?

Was there ever anything that happened when you were growing up that you couldn’t explain? Were there any spooky places you wouldn’t go at night? Does it bother you when people talk about ghosts?

Have you ever been somewhere new and know that you’ve been there before?
4.2. Study Design

What was the longest streak of luck you ever had? Do people feel the same way as they used to? What about bad luck? Are you lucky at cards? With women [men]?

(Labov 2004, Q-GEN-II, Module 6; note that double asterisks indicate usage of the exact wording of the question; Labov 1984: 34)

Feagin maintains that in her fieldwork, most respondents would answer the first question with a single “No.” after a pause (2002: 30, 2013: 30; also cf. Schilling-Estes 2007: 173). In the interview trials I ran before the fieldwork period began, colleagues and friends reacted similarly. Some stated that they could not remember the details but they were certain that, for instance, something happened when they were growing up that they could not explain; as in: I am sure everyone could tell you stories like that. Most respondents’ stories, however, were quite short and superficial. I interpreted them as signs of politeness concerning the provision of an answer to a question rather than reminiscent and emotional engagement in a personal narrative. While he did answer, one Indian-Canadian respondent from Toronto could not help but laugh and inquired about the seriousness concerning the nature of those questions. Instead, he suggested to talk some more about something else such as his hobbies and time spent with friends during summer.

Since the first question in the above excerpt has been found to be unsuccessful in communities other than New York, community-specific substitute questions have been suggested instead (e.g. Feagin 2002: 30; Tagliamonte 2006: 38). For Canada in particular, the question “People keep saying we’re getting more and more American. Do you think that’s true?” has been suggested (Feagin 2002: 30, 2013: 30). For St. John’s, I include community-specific questions that are supposed to serve the same purpose such as: “Where were you during the flood in 2006?”, “What did you do?”, “How did it affect you and your friends/family?” or “Do you feel as a Newfoundlander or as a Canadian?”, “Why?” (cf. Tagliamonte 2006: 38). As in the trial interviews, most of my respondents, however, engaged only in very short narratives of one to three minutes and did not show obvious signs of emotional investment.

In 1989, Labov subsumed careful and casual speech under spontaneous speech (i.e. conversational, but real and natural speech; 1989: 11), because such a definition was sufficient to distinguish oneself as a sociolinguist theoretically from those who rely on introspective data (1989: 51). Although the design of my interview also follows the ultimate goal of recording casual style, I am not as rigidly concerned with the differentiation of casual (vernacular) and careful style. This is particularly important because the existence of a homogeneous vernacular, independent of situational circumstances and interpersonal relationships (e.g. talking to one’s brother versus one’s close friend), is highly doubtful (cf. Eckert 2000: 78-82; Milroy and Gordon 2003: 49-51; Schilling-Estes 2007: 173). I will thus refer to the style recorded in the spontaneous speech section of my interview as interview style or spontaneous/connected speech. In addition, I decided not to use the questions or make-up of Labov’s modules for the sociolinguistic interview, but to rely on
more open (predominantly *wh-*) questions such as those suggested by Labov, Ash and Boberg (2006b: 32-35) and Tagliamonte (2006: 38-49, Appendix B) and on my university training in social and pedagogical psychology as part of my Master’s Degree in order to achieve the same goals as outlined by Labov (1984: 32-33).

I planned the spontaneous speech section of the interview to be one hour to one and a half hours in length (Tagliamonte 2006: 37) and to consist of several, variably ordered parts. After the respondents read the word list and the reading passage, I included a section on their demographic data, followed by spontaneous conversation with me as their interviewer (cf. Boberg 2010: 144; Labov 1984: 32) and by a section on language. Since the information on the demography of the speakers is among the most formal ones in a sociolinguistic interview (goal 2; Labov 1984: 33), I used this section primarily as means to measure the extent to which the interviewees are truly representatives of the speech community (cf. Appendix A.4.1). At the same time, I used it to ease the interviewee’s way into the more informal and specific parts of the interview to follow, by testing what kind of stories/topics they were (or were not) interested in talking about. The results of this section are quantified as an index of local-ness to St. John’s (cf. Subsection 4.2.2.2).

According to Labov (1984: 32), goals four to six are the ones that are directly concerned with eliciting the vernacular: four with personal narratives, five with recording conversation between speakers present, not addressed to the interviewer, and six with letting the interviewee define the topic of conversation. The fifth goal was difficult to achieve as many respondents were alone when I interviewed them. I left it to the interviewee to determine where they want to be recorded, as long as the recording site was in a rather small room without sources of noise such as an electric fan, the refrigerator, extensive traffic, etc. (cf. Labov 1984: 41; Tagliamonte 2006: 45), in order to achieve recordings of reasonable fidelity (goal 1; Labov 1984: 32). The recording sites I ultimately used were either a near-soundproof room in the Department of Linguistics at MUN (mostly but not exclusively students) or the living rooms of middle-aged and old respondents. Other family members and friends were rarely present in the living rooms or houses during the recording of the interview across my sample. In addition to the inconsistency, the very short length of such conversational interaction among the respondents made it incomparable on an inter-speaker level.

The introductory questions I asked for each new topic throughout the interviews in general and all questions concerned with goals four and six in particular were practiced, memorized and phrased in my personal colloquial style in order to be short and ideally inhibit more standard registers (cf. Labov 1984: 34). Per conversational topic, I only noted and memorized the initial, introductory questions which are formulated from an outsiders’ point of view (cf. Labov 1984: 34) and used the answers to elicit more speech from topics/stories the interviewees provided (cf. Labov 1984: 37). I considered this questioning technique particularly useful to avoid simple “No.” answers and other problems I had in the interview trials as outlined above. The print-out of the questions I used
4.2. Study Design
during the interview differed from the memorized questions in that it only contained a
summarized version of the most important questions per section of the interview as a cue
and the additional material I needed for the individual sections (e.g. examples of rhymes,
a map of Newfoundland, etc.). The questions I practiced can be found in Appendix A.4.2.

When formulating the spontaneous questions, I attempted to minimize the potential
of raising certain expectations of answers on the part of the interviewee. I particularly
emphasized my role as an outsider to the speech community in the sense that I was curious
to learn about the respondents’ family/friends, neighborhood, upbringing, hobbies, work,
etc (cf. Schilling-Estes 2007: 173). Only during the demography section, interviewees
were asking me questions of whether I needed other or additional information on their
education when they would think of something I did not (yet) ask for. I usually attempted
to reject any authority in this situation via replies such as “Oh, if it is OK with you, I’d
love to hear everything”. More generally, I attempted to minimize my authority via
colloquial speech, politeness, apologies, empathy, honesty and emphasis of my position as
a/an learner/equal (after I told the interviewees about the requirements the recording of
the interview had to meet; cf. Labov 1984: 40), while being authentic at the same time.
These strategies resulted on the one hand in very different stories of personal interest to
the respondents and on the other hand in personal narratives of 10 to 15 minutes without
me asking any questions.

For instance, some middle-aged females got so involved when talking about their
children’s parties with friends from school at their house that they unclipped the lavaliere
microphones multiple times through wild gestures. One old respondent asked me if I
knew how to construct a bear trap with utensils found in the forests around St. John’s.
After I told him “No, sounds dangerous though”, he also forgot about the microphone
while getting up to fetch his photo album in order to show me pictures and tell me in
detail about hunting trips with his friends (for roughly 30 minutes). Another middle-
aged respondent offered me beer and tried to give me guitar lessons after talking about
band practice and relating on-stage-performance stories. One young female respondent
continuously returned to experiences with her figure-skating team, regardless of the topic
she decided to talk about (e.g. differences between Canadians and Newfoundlanders,
closest friends, her work, etc.). Other respondents were less relaxed characters who rather
liked to talk about serious topics such as socioeconomic change in St. John’s, separation
of the province from mainland Canada, etc.

Since I could not plan how much time would pass once the respondents commenced
talking about their stories/experiences, I ultimately did not ask all of the questions in
every interview. Once the interview had lasted for two hours, I did not initiate new topics,
but also did not end the interview if the respondents continued to talk, in order to maxi-
mize the natural, conversation-like character of the interview (cf. Tagliamonte 2006: 46).
I also excluded questions that could potentially violate the guidelines of the Ethics Com-
mittee, such as housing value, salary, sex, sensitive information on marriage/relationships
and violence. Especially the latter three of these topics were brought up by some of my respondents, but I decided to simply listen instead of reacting in any way that would indicate that they could confide in me (cf. Tagliamonte 2006: 43). I excluded questions about the former two topics because I consider them inappropriate (cf. Tagliamonte 2006: 43), especially for Newfoundland. For instance, the Newfoundlanders I engaged with socially after work generally had very strong political views on the financial situation between the province’s and the federal government, especially since the cod moratorium in the 1990s, although none of them was personally affected.

After this interview section, concerned primarily with the sentiments and experiences of my respondents, I included another section on the language and the perception of linguistic features of Newfoundland English in general. For the latter, I used the method of drawing a map from perceptual dialectology (e.g. cf. Bucholtz et al. 2007; Preston and Fought 1993) without its original purpose: instead of quantifying perceptual isoglosses from such maps, I used them as a trigger for my respondents in order to assess how much salience local features have to them in general, which local features are salient and how well they can express them (cf. Part I in Appendix A.4.3). I theorized that (except for the linguistics students) they lack the technical register to express linguistic features, especially regarding vowels, although they are aware of them. One reasonable way out of this dilemma should be the usage of rhymes to express the quality of vowel sounds. After the map-drawing task, I provided respondents with multiple choice questions for some vowels. As in the word list (cf. Subsection 4.2.1.1), I used more vowels than of interest for this thesis (TRAP before /g/ and the low-back merger) in order to distract linguistics students. A detailed evaluation of the map-drawing task will be part of a future project.

After I raised the respondents’ awareness for linguistic features, I asked more general questions on their perceptions of the ‘townie’ versus ‘bayman’ divide, a known and salient distinction between people from the only conurbation in Newfoundland (primarily St. John’s, but also Conception Bay South and Mount Pearl) and all the other people from small towns near the coast, scattered across the island off the Avalon Peninsula. In addition, I included questions concerning the local dialect in St. John’s in contrast to the prestigious, supra-regional standard from mainland Canada, if the respondents did not bring it up by themselves. The questions I practiced before the interview for this section can be found in Appendix A.4.3.

I designed this section primarily in this fashion because I intended to create a similar atmosphere as in the section eliciting the vernacular of the respondents, namely to position me as the learner (despite its stylistically formal framework; see below). In addition, I wanted to elicit the sentiments of respondents from St. John’s towards the incoming standard and the non-urban speakers from the island, as well as towards the role of the incoming standard in their everyday lives. Since the language section is among the more formal ones in the interview in order to arrive at comparable results for all respondents (Labov 1984: 32), I use this section again to determine the extent to which all of the
4.2. Study Design

Respondents are exposed to the mainland Canadian standard of their peers. Virtually all of the speakers expressed very positive evaluations of the local ‘baymen’ varieties while at the same time maintaining that they do not have a very strong Newfoundland accent due to their work (except for three speakers). All of the respondents have mainland Canadian co-workers, “work friends” and/or friends who they spend and/or have spent extended periods of time with. Due to this categorical result, I do not include a per-speaker index for exposure to mainland Canadian features in the analyses.

As mentioned in Subsection 4.2.1.1, this thesis is not primarily concerned with an accurate stylistic continuum in the sense of attention paid to speech. Moreover, I understand the style respondents can shift into and out of as performance styles, which are not solely determined by the task they are asked to do (reading word lists versus reading a text versus spontaneous speech; cf. Subsection 4.2.1). In order to ensure comparability to the atlas by Labov, Ash and Boberg (2006b), I extracted the tokens for acoustic analysis only from informal, spontaneous, natural dialogue in the interview (cf. Tagliamonte 2006: 46).

4.2.2 Sampling and Data Collection

In his early work in New York City, Labov (1966) attempted to recruit participants via random sampling (Milroy and Gordon 2003: 25). This is a necessary prerequisite in any statistical analysis if the sample is to be representative in the sense that “[... ] its relationship to the population can be precisely specified [... ]” (Milroy and Gordon 2003: 26). Strict statistical representativeness as sometimes required in other social sciences (cf. e.g. Bortz and Döring 2006: 395) is rarely the intention of sociolinguistic research today (Milroy and Gordon 2003: 26), due to its inefficiency in terms of weighing costs of achieving it against the limited additional benefits it might provide (Chambers 2009: 44; Milroy and Gordon 2003: 25). Instead, sociolinguistic investigators accept “this rather weaker kind of representativeness attained in most linguistic surveys” and prefer some form of judgment sampling (Milroy and Gordon 2003: 26; also cf. Di Paolo and Yaeger-Dror 2011a: 13; Hoffman 2014: 31).

I stratify my judgment sample according to sex, age and socioeconomic background of the speakers. The status of the other two social variables ethnicity and religion as included by Clarke (1985a,b, 1991) in her St. John’s survey requires some elaboration. The interviewees in the pilot study were all of mixed ethnic heritage, which seems to be the default for inhabitants of St. John’s at least since World War II (Clarke 1985a: 68). Historically, religion is a good determinant for ethnic origin on the island of Newfoundland, which in turn has socioeconomic overtones. The reason for the direct relationship between religion and ethnicity was the power of the denominational school system, in which the Irish predominantly went to Roman Catholic schools and the English to Protestant schools (Clarke 1985a: 68). This system has lost all putative remnants of its former
power with its official abolition in 1998 with an amendment to the province’s constitution after a referendum in 1997 (Bergman, Stoakes Sullivan and Fisher 1997). Ethnic and religious segregation in Newfoundland has thus become insubstantial in the minds of Newfoundlanders in general and inhabitants of St. John’s in particular within the past generation(s). A similar picture emerged from the pilot study: The ethnic mixture of St. John’s inhabitants is accompanied by non-practice of religion, by a change in religion or by being of no denomination whatsoever. In addition, my focus is on supra-regional features, instead of traditional dialect features associated with a particular ethnic origin (e.g. Irish ‘after’ perfect or stopping of interdental fricatives; cf. Clarke 1985a: 68). Hence, I exclude ethnicity and religion as variables in my sample due to the inseparability of the factors within each variable and their rather minor importance in present-day St. John’s.

As outlined in Section 4.2, after one week of my fieldwork between mid-August and late September 2011, I was denied assistance in recruiting respondents so that I arrived at the judgment sample via concentrating predominantly on a friend-of-a-friend approach to sampling (cf. D’Arcy 2005: 331; Milroy and Gordon 2003: 32; Schilling-Estes 2007: 179; and also Hoffman 2014: 31). I particularly focused on selecting respondents from various social networks in order to balance the sample as much as possible (cf. Eckert 2000: 69-84; Schilling-Estes 2007: 180). That is, I attempted to primarily reach parents and/or grandparents (as well as their friends) of already-interviewed students or other young participants via the “snowball” technique (Milroy and Gordon 2003: 32), but not so much groups of (young) friends. This procedure was also necessary because most of the respondents’ social networks consisted of friends non-native to St. John’s, which I generally excluded from the sample. Not more than three people who know each other are included in the final sample outlined below.

I was provided access to a multitude of circles of young friends through colleagues at MUN who offered me to do “recruitment speeches” in their classes in order to attract some of their students to participate in my study (cf. Appendix A.5.1). Furthermore, I set up a website, disseminated flyers and talked to students/staff, informing them about the social and demographic profile potential participants would have to meet (being native to St. John’s, being middle-class, etc.). Other respondents who had heard or read about my research and emailed/called me were asked whether they consider themselves to be a typical middle-class St. John’s person (cf. Subsection 4.2.2.1) and to fill out the Participant Background Information form (cf. Appendix A.5.2). In the meantime, I commenced interviewing Bachelor, Master and Ph.D. students of the Sociolinguistics Lab, the Department of Linguistics and the Department of Folklore at MUN. After the interviews with the students, I successfully inquired whether their parents, grandparents, spouses and friends (of each of the former) would want to participate as well. Additionally, I accompanied some of my first-order informants to local social gatherings (e.g. a knitting club) in order to recruit more respondents.
4.2. Study Design

Since my demographic and social profile would hinder access (or at least make it very difficult) to respondents furthest away from me in this respect (cf. Schilling-Estes 2007: 182), I had planned the social matrix of my respondents to solely comprise middle-class respondents. Nevertheless, I attempted to include working-class members in the beginning of the recruitment process. As expected, the distribution of interviewees in the respective cells turned out to be a difficult endeavor. Female members of the working class and old females of the middle class were most difficult for me to include in my sample. For the latter group, I received invaluable support from Sandra Clarke, but in case of the former, matters were much more complicated. I was only able to convince four and mostly male members of the working class to be recorded. Despite the great help of the friends I had made during the fieldwork, even including those four required a lot of engagement and effort on their and my part. As I learned in hindsight, they were particularly afraid of reading a word list and a text while being recorded, despite my continuous assurances that the recording device would be turned off whenever they wanted and that it was not a reading test. This was particularly regrettable given the great time we had spent together on various Friday and Saturday nights at their houses or at clubs when they would tell me so many different personal stories and I could not record anything.

By the end of my fieldwork, I had interviewed 41 respondents, from which seven are excluded in the analyses of this thesis for various reasons (e.g. maintaining balance of the sample or social status of interviewees). Each participant had to sign a form providing their informed consent to the study. The consent is part of the guidelines of the Interdisciplinary Committee on Ethics in Human Research at Memorial University of Newfoundland which approved this study (cf. Appendix A.5.3). I recorded each interviewee individually in order to avoid the problems with interviewing groups of speakers simultaneously. According to Labov, it is not possible to determine all of the existing groups in a given speech community and which proportion of the total number of groups has been recorded. If it were possible to interview groups from neighborhoods in a study, the factors that enabled their recording would be accidental, so that such a study would not be replicable. In addition, not every member of a group recorded together in an interview displays the same stylistic range, nor do members engage in similar amounts of speech, if they speak at all as part of a group (1984: 49).

Each participant was digitally recorded on CompactFlash solid state memory cards, using a Fostex FR-2LE at a setting of 48 kHz and 24-bit for uncompressed PCM audio files in the Broadcast Wave Format. I used this device because of the high-quality microphone pre-amplifiers and phantom power supply, which, according to Plichta (2014: n.p.), make the Fostex FR-2LE an excellent stand-alone recording device for the needs of a field linguist, unlike devices with built-in microphones. The Fostex’s high gain and low noise pre-amplifiers cancel the need for additional, external pre-amplifiers to adequately drive XLR-connected condenser microphones. I used an Audio-technica omnidirectional,
Cardioid condenser, lavalier microphone AT831b. According to Plichta (2014: n.p.), this microphone is of good quality and particularly useful in noisy environments (e.g. living rooms). I chose a lavalier instead of a head-set or table microphone in order to reduce its obtrusiveness that may remind interviewees continually of the recording situation (cf. Labov 1984: 33). Additionally, head-set microphones may be inconvenient or cause feelings of resistance or intrusion into the private sphere of the interviewee. I asked each respondent to place the microphone roughly at a distance of 15 cm away from their mouths.

Table 4.2 shows the distribution of the middle-class participants that I recorded according to age and sex. The speakers are distributed within the age groups according to the reasoning I outlined in Section 4.2: The cut-off point for the young cohort is 25 years of age, corresponding roughly to the completion of university/college; the old cohort reflects participants who are about to retire from the linguistic marketplace; the middle cohort includes the remaining participants, who typically pursue occupations.

The young and middle-aged cohorts are slightly skewed towards female respondents. The ratio of male to female participant of 1:1.4 in the young age cohort is, however, unproblematic when compared to 1:1.95 in the atlas (Labov, Ash and Boberg 2006b: 28) for their similar age cohort (20- to 29-year-olds), i.e. 41 males versus 80 females. Likewise, 65 males and 119 females form the data set in the cohort of 30- to 39-year-old respondents (2006b: 28). My gender ratio for the young speakers is also unproblematic when compared to 1:1.75 in Boberg’s youngest age group from his Montreal study (2005: 138). Likewise, my 34-participant sample is skewed slightly towards female respondents, with a ratio of 1:1.25. Boberg’s 35-respondent Montreal sample had a ratio of male to female respondents of 1:1.5 (2005: 138). In his PCE data, Boberg included 35 male and 51 female McGill University students, yielding a ratio of 1:1.46 (2008b: 133). In terms of old-aged speakers, I have slightly fewer respondents compared to the other two age groups. In contrast, 24 out of 39 Canadian respondents in the atlas were between the ages of 20 and 40 in their data, which ranged from either 10 to 70 (2006b: 28, Table 4.5) or from 10 to 80+ (2006b: 28, Table 4.6).

In addition to age and gender, Labov, Ash and Boberg (2006b: 220) included city size and education as social variables in the analysis of the 33 Telsur (telephone survey)
participants selected for Canada as a whole (west coast to Atlantic Provinces). Both of
these variables are controlled for in my sample: I selected informants from one city – St.
John’s (106,000 inhabitants in 2011; Statistics Canada 2012: n.p.) –, and education is
indirectly accounted for by the variable occupation outlined in Subsection 4.2.2.1. Instead
of these two variables, I include the local-ness index for each of the speakers in my sample,
which I will explain in detail in Subsection 4.2.2.2 below.

4.2.2.1 Socioeconomic Class Membership

Traditionally, membership in a social class was determined by nationally applicable so-
cioeconomic index (SEI) scores (e.g. derived from census data and the data of the National
Opinion Research Council, NORC, as done by Labov, Ash and Boberg 2006b: 30), in-
cluding various social variables such as education, occupation and family income (e.g. cf.
Cedergren 1973; Labov 1994: 58, 2001a: 61). In small-scale studies such as the present
one, the few respondents included make such SEIs as derived from NORC not necessarily
applicable. In such cases, the same information is usually elicited from the respondents
themselves, and their answers are combined in order to calculate an index score based on
the same social variables (Chambers 2009: 48; Labov 1966; Trudgill 1974). Furthermore,
I doubt that the ratings for the prestige of jobs in the U.S. and/or Canada such as those
in NORC are applicable to St. John’s.

While Labov maintains that combined indices are more desirable in determining so-
cial class membership than a single index, “[i]t is generally agreed that among objective
indicators, occupation is the most highly correlated with other conceptions of social class”
(Labov 2001a: 60). Among the reasons for multiple indices of social class, he lists: aid in
explaining more aspects of class-based behavior, tapping different dimensions of socioe-
conomic status and that the (in)consistency of various indicators will provide additional
information about socioeconomic status patterns (Labov 2001a: 60).

As emphasized by Chambers, combining means for such objective criteria as education
and income with occupation in order to determine social class is an abstraction (2009: 46), i.e. managerial professions can either be entered by someone who has the necessary
education or who has worked their way up. In terms of income, the same professional label
such as architect does not entail the same annual income for each individual architect.
One may have decided to stay close to their family and thus to accept a job in a small
local office; another may have sacrificed social ties for a successful career in a global-
player company. An SEI score (including means for education and income) “characterizes

However, using several class indicators increases the fuzziness of the individuals’ SEI
scores; the fuzzier they are, the vaguer is the correlation of social class and the dependent
linguistic variable (Chambers 2009: 52). In addition, the use of such elaborate, multi-
dimensional class indices has declined since the early studies, as they have proven to be
unnecessary in sociolinguistics (Chambers 2009: 50), with empirical support provided, for instance, by Macaulay (1977: 178).

Furthermore, these objective criteria cannot account for the fact that individuals who have a high annual salary, extraordinary education and a prestigious profession consider themselves to be members of the lower middle class and consequently use linguistic variants associated with the latter instead of the objectively determined putative upper middle class (cf. Schilling-Estes 2007: 170). In fact, the objective nature of such economic indices is seriously called into question, as the self-perception of the respondents’ position in the social hierarchy may bear a more direct relationship to their social status than the former (Milroy and Gordon 2003: 40-47). Usually, the self-perception of individuals correlates more closely with their linguistic behavior than alleged objective measures of economic wealth (Schilling-Estes 2007: 170).

For these reasons, I do not calculate an SEI score based on census/NORC data for my respondents. Instead, I rely on the “touchstone of social-class membership” (Chambers 2009: 46) – the respondents’ occupations – and additionally list their education, as Boberg (2005: 139) did in his Montreal study and Hollett (2007: 33) in her St. John’s study. This procedure does not hinder comparability with the atlas (Labov, Ash and Boberg 2006b), as the social status of their two respondents from St. John’s, Newfoundland, is unclear. One 35-year-old male respondent works in the military, for which NORC does not return a score so that no SEI could be extrapolated (cf. Labov et al. 2006: n.p.), and the other 24-year-old female respondent was a student of civil engineering with an SEI score of 14 for education and of 86 for her ‘breadwinner’s’ occupation – most likely middle class.

Table 4.3 lists the occupation of each respondent, including their highest educational level. For students, I list the occupation of their fathers (cf. Labov, Ash and Boberg 2006b: 30) and for retired respondents the last occupation they have had prior to retirement. I would like to stress that these occupations are only snapshots of the biographies of these speakers, since most work in different occupations or differently ranked occupations throughout their lives. For instance, speaker 17LEMM (17) has started as an unskilled worker in a call center and eventually became ‘customer care manager’. Recall that in the recruitment process, I placed emphasis on the self-perception of my respondents concerning middle-class membership.

The vast majority of the respondents listed in the table live in the west end of or in Downtown St. John’s, having moved back and forth between the two parts of town in their lives. Few of them are from the north end of St. John’s. According to the City of St. John’s website, the residence values in these neighborhoods are typically from entry-level to executive (2014: n.p.). Downtown St. John’s also offers condominium living. The average housing value was approximately $268,600 in 2011 (City of St. John’s 2014: n.p.).
### 4.2. Study Design

Table 4.3: Birth year, sex, education and occupation of the individual respondents (n = 34). The education listed is the highest degree for those respondents who still attend educational institutions or the highest attended institution for those who have not completed their degree/diploma. “Some” refers to education that has not been completed via a degree or diploma. For students, the occupation of their fathers is listed.

<table>
<thead>
<tr>
<th>Speaker #</th>
<th>Birth Year</th>
<th>Sex</th>
<th>Education</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>1951</td>
<td>f</td>
<td>some community college</td>
<td>clerical worker</td>
</tr>
<tr>
<td>7</td>
<td>1952</td>
<td>m</td>
<td>doctorate</td>
<td>university lecturer</td>
</tr>
<tr>
<td>9</td>
<td>1952</td>
<td>m</td>
<td>bachelor</td>
<td>business manager (self-employed)</td>
</tr>
<tr>
<td>20</td>
<td>1952</td>
<td>f</td>
<td>bachelor</td>
<td>civil servant (retired)</td>
</tr>
<tr>
<td>32</td>
<td>1952</td>
<td>m</td>
<td>doctorate</td>
<td>university lecturer</td>
</tr>
<tr>
<td>8</td>
<td>1953</td>
<td>f</td>
<td>completed high school</td>
<td>clerical worker (retired)</td>
</tr>
<tr>
<td>12</td>
<td>1953</td>
<td>m</td>
<td>completed high school</td>
<td>salvage business manager (self-employed)</td>
</tr>
<tr>
<td>21</td>
<td>1953</td>
<td>m</td>
<td>bachelor</td>
<td>research assistant (university)</td>
</tr>
<tr>
<td>40</td>
<td>1953</td>
<td>f</td>
<td>bachelor</td>
<td>nurse</td>
</tr>
<tr>
<td>22</td>
<td>1963</td>
<td>m</td>
<td>some college</td>
<td>programmer (unemployed)</td>
</tr>
<tr>
<td>24</td>
<td>1964</td>
<td>m</td>
<td>some university</td>
<td>cook</td>
</tr>
<tr>
<td>17</td>
<td>1965</td>
<td>m</td>
<td>some university</td>
<td>customer care management</td>
</tr>
<tr>
<td>23</td>
<td>1969</td>
<td>f</td>
<td>bachelor</td>
<td>nurse</td>
</tr>
<tr>
<td>31</td>
<td>1969</td>
<td>f</td>
<td>master</td>
<td>university lecturer (nursing)</td>
</tr>
<tr>
<td>33</td>
<td>1971</td>
<td>m</td>
<td>bachelor</td>
<td>pastry chef (self-employed)</td>
</tr>
<tr>
<td>35</td>
<td>1973</td>
<td>m</td>
<td>completed high school</td>
<td>musician</td>
</tr>
<tr>
<td>2</td>
<td>1976</td>
<td>m</td>
<td>bachelor</td>
<td>programmer/analyst</td>
</tr>
<tr>
<td>34</td>
<td>1976</td>
<td>f</td>
<td>master</td>
<td>international development consultant</td>
</tr>
<tr>
<td>27</td>
<td>1977</td>
<td>f</td>
<td>master</td>
<td>elementary school teacher</td>
</tr>
<tr>
<td>3</td>
<td>1981</td>
<td>f</td>
<td>bachelor</td>
<td>student; business manager</td>
</tr>
<tr>
<td>28</td>
<td>1982</td>
<td>f</td>
<td>bachelor</td>
<td>business analyst</td>
</tr>
<tr>
<td>11</td>
<td>1985</td>
<td>f</td>
<td>bachelor</td>
<td>executive director in private education</td>
</tr>
<tr>
<td>14</td>
<td>1986</td>
<td>f</td>
<td>bachelor</td>
<td>student; civil servant</td>
</tr>
<tr>
<td>15</td>
<td>1986</td>
<td>m</td>
<td>bachelor</td>
<td>student; teacher</td>
</tr>
<tr>
<td>16</td>
<td>1986</td>
<td>m</td>
<td>bachelor</td>
<td>network administrator</td>
</tr>
<tr>
<td>25</td>
<td>1986</td>
<td>f</td>
<td>bachelor</td>
<td>student; police officer</td>
</tr>
<tr>
<td>5</td>
<td>1989</td>
<td>m</td>
<td>completed high school</td>
<td>student; clerical worker</td>
</tr>
<tr>
<td>36</td>
<td>1989</td>
<td>f</td>
<td>completed high school</td>
<td>student; senior officer Can.Rev.Agncy</td>
</tr>
<tr>
<td>38</td>
<td>1989</td>
<td>f</td>
<td>completed high school</td>
<td>student; engineer</td>
</tr>
<tr>
<td>39</td>
<td>1989</td>
<td>m</td>
<td>completed high school</td>
<td>student; financial advisor</td>
</tr>
<tr>
<td>13</td>
<td>1990</td>
<td>f</td>
<td>college diploma</td>
<td>receptionist</td>
</tr>
<tr>
<td>26</td>
<td>1990</td>
<td>f</td>
<td>completed high school</td>
<td>student; university lecturer</td>
</tr>
<tr>
<td>30</td>
<td>1990</td>
<td>f</td>
<td>completed high school</td>
<td>student; manager taxation center</td>
</tr>
<tr>
<td>37</td>
<td>1991</td>
<td>m</td>
<td>completed high school</td>
<td>student; factory worker</td>
</tr>
</tbody>
</table>
4.2.2.2 Local-ness to St. John’s

As outlined above, the most important characteristic for selecting respondents in this thesis was their local-ness to St. John’s (cf. Subsection 4.2.2). I ensured that this requirement was fulfilled either via personal communication with participants prior to the interviews or the Participant Background Information form (cf. Appendix A.5.2). Since the vernaculars have been found to be unstable before the age of 20 (cf. Bailey 2002: 319; Chambers 2009: 207; Cukor-Avila and Bailey 2013: 246; and Section 4.2), or at least 18 (Labov 1972b: 138), and since the vernaculars or speech patterns in pre-adolescent and adolescent years are predominantly influenced by peers (members of the same social network and/or community of practice; cf. Eckert 2000; Milroy 2002) rather than parents, I placed particular emphasis on the fact that my participants are raised in St. John’s, but not necessarily born (90% of the speakers are also born in St. John’s). That is, particularly in the early formative years of the respondents’ lives (one to six years), I ensured that by that time they had been long-term residents of St. John’s and were visiting local schools.

In order to find a means to account for the effects of local parentage and non-local parentage in St. John’s that D’Arcy (2005) found to be influencing the innovative retraction of TRAP among pre-adolescents and adolescents (cf. Section 4.1), I calculated an index score which is supposed to stress the rather small differences concerning local-ness in my relatively homogeneous sample. In other words, virtually everyone in my sample is truly representative of St. John’s English, so that a local-ness index score of small value does not mean the respondent is not local to St. John’s, but less local than others. The scores thus reflect the heterogeneity of the middle class in this respect, rather than a dichotomy.

The local-ness index is comprised of several subordinate indices: First, I calculated an index for generation born in St. John’s (LIgen), second, one for having attended local schools (LIschool), and third, one for having been abroad for a period of at least one year (LIabroad). These scores are then combined to the total local-ness index (LItotal) for each speaker.

Only LIgen is a more or less directly equivalent to the local/non-local parentage variable D’Arcy (2005) included in her analysis. Respondents can score a maximum of three points in this index: one for each parent born in St. John’s and one for themselves being born in St. John’s. I added only half a point if one of the three was born outside of St. John’s, but not outside of Newfoundland. If one of the three was born in mainland Canada, they were given zero points; if a parent was born outside Canada, I subtracted half a point, resulting in a score of -0.5 for that parent. As mentioned above, this categorization replicates D’Arcy’s in terms of local/non-local parentage, with the additional detail of differentiating between parents from outside Canada, the Canadian mainland or the island.
4.2. Study Design

Unlike D’Arcy’s, my youngest respondents are around the age of 20/21, which made it necessary to include the influence of their peers in school. For LIschool, respondents can also score a maximum of three points: one for grade school attended in St. John’s and two for high schools attended in St. John’s. I emphasize the latter in points because the influence of the peers should be stronger for a 17-year-old than for a 7-year-old. In addition, one school year not spent at a school in St. John’s resulted in one point subtracted for the whole category, in order to emphasize the differences between the participants.

My 22 older participants do not have social networks that are relatively stable over longer periods of time such as those in schools. After high school graduation, their vernaculars should be already stable, which is why I excluded tertiary education institutions from the indices. Moreover, not every participant had a chance to attend such an institution. For these older interviewees, I deemed it necessary to include possible years spent outside St. John’s for various reasons (e.g. occupation or relationships; LIabroad). Since such years may have more effect on one’s variety at younger years of age, I chose 30 as a rough cut-off point, i.e. for respondents who have spent more than one year outside St. John’s before the age of 30 I subtracted one point, and for those who have spent a year abroad (again) after the age of 30 I subtracted one (additional) point.

For LIabroad, a maximum of two points was thus subtracted from the indices LIgen and LItotal. The reasoning behind this cut-off point in the category is that alumni who left St. John’s with a Master’s degree and no work experience would have had at least five to six years to accommodate to a more standard variety possibly necessary in the speech communities they have their occupations in. Due to the lack in experience, they may feel more pressured by the constraints of the linguistic marketplace to use a variety not revealing regional origins than speakers older than 30 who can increase their value on the marketplace via expertise. The former speakers may employ this standardized register/style in more formal situations of the interview (reading word lists, text passages, providing demographic information, speaking about language). However, it should not affect their stabilized vernaculars, so that I only subtract one point for such résumés. In order to account for longer periods of time spent outside St. John’s, I subtract one additional point.

Table 4.4 lists the local-ness index scores for each participant. The maximum number of points for LItotal is six: three in LIgen, three in LIschool and zero in LIabroad. Two-thirds of the participants score five points or more; the remaining third scores fewer points mostly because they have spent consecutive time of at least one year outside St. John’s. Speakers 16, 17, 23 and 33 were not born in St. John’s, but moved there before the age of three. Six speakers did not spend every school year at an institution in St. John’s, either in grade school, high school or both. Three detailed examples of how the local-ness index was calculated can be found in Subsection 5.6.3. LItotal is entered as a social predictor variable in the multiple regression analyses.
Table 4.4: Ligen, LIschool, LIabroad and LItotal of the individual respondents (n = 34).

<table>
<thead>
<tr>
<th>Speaker #</th>
<th>Birth Year</th>
<th>LIgen</th>
<th>LIschool</th>
<th>LIabroad</th>
<th>LItotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>1951</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>1952</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>1952</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>20</td>
<td>1952</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>1952</td>
<td>1</td>
<td>3</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1953</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>12</td>
<td>1953</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>21</td>
<td>1953</td>
<td>3</td>
<td>3</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>40</td>
<td>1953</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>22</td>
<td>1963</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>1964</td>
<td>2</td>
<td>3</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>1965</td>
<td>1</td>
<td>2.5</td>
<td>-1</td>
<td>2.5</td>
</tr>
<tr>
<td>23</td>
<td>1969</td>
<td>2.5</td>
<td>3</td>
<td>-2</td>
<td>3.5</td>
</tr>
<tr>
<td>31</td>
<td>1969</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>33</td>
<td>1971</td>
<td>3</td>
<td>3</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>35</td>
<td>1973</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1976</td>
<td>0.5</td>
<td>2</td>
<td>0</td>
<td>2.5</td>
</tr>
<tr>
<td>34</td>
<td>1976</td>
<td>3</td>
<td>3</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>27</td>
<td>1977</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1981</td>
<td>2</td>
<td>2</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>28</td>
<td>1982</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>1985</td>
<td>3</td>
<td>3</td>
<td>-1</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>1986</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>15</td>
<td>1986</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>16</td>
<td>1986</td>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>1986</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1989</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>36</td>
<td>1989</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>38</td>
<td>1989</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>39</td>
<td>1989</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>1990</td>
<td>2.5</td>
<td>1</td>
<td>0</td>
<td>3.5</td>
</tr>
<tr>
<td>26</td>
<td>1990</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
<td>3.5</td>
</tr>
<tr>
<td>30</td>
<td>1990</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>37</td>
<td>1991</td>
<td>2.5</td>
<td>3</td>
<td>0</td>
<td>5.5</td>
</tr>
</tbody>
</table>
4.3 Acoustic Theory

In this section, I will briefly summarize the basics of the acoustic theory of speech production relevant for this thesis. I outline them as they facilitate comprehension of Subsection 4.3.3 and the methodological decisions made for this study in the sections to follow. More details on acoustic theory can be found in Clark, Yallop and Fletcher (2007: 204-295), Johnson (2003: 79-112), Ladefoged (1996: 92-135), Mayer (2010) and Thomas (2011: 31-37). This sections will be concerned with the Source-Filter theory (Subsection 4.3.1), tube models (Subsection 4.3.2) and Linear Predictive Coding (Subsection 4.3.3).

The Source-Filter theory as one way of thinking of human speech production (Ladefoged 1996: 103) states that the vocal folds in the larynx produce sounds and the vocal tract modifies them, before they leave the mouth of a speaker (Johnson 2003: 79). The vocal folds are thus the source and the oral, pharyngeal and for some sounds nasal cavities (vocal tract) constitute the acoustic filter. The source and filter are independent of one another, as a change in the fundamental frequency (F0; related to the perception of the pitch of the voice) of a speaker does not result in a change of the resonant frequencies, given the shape of the filter remains unchanged (Ladefoged 1996: 98).

Tube models may serve to schematize the vocal tract. It can be thought of as a series of tubes that have infinite resonances, depending on their shape, length and diameter and on whether they are closed on both ends or open on one end. The lowest three resonances correspond to the frequency bands that determine the quality of voiced sounds in general and vowels in particular. Perturbation theory is an alternative model to tube models that predicts the resonances of the vocal tract based on pressure and velocity of air particles (cf. Johnson 2003: 107; Ladefoged 1996: 115; Thomas 2011: 32).

Linear Predictive Coding (LPC) was developed in order to automatically locate and determine the frequency bands or formants of vowels. It can be regarded as the inverse of the process of speech production, as the LPC filter is the inverse of the power spectrum of the speech signal that serves as input. The output produced by the LPC filter is as close to zero as possible (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 93; Johnson 2003: 40; Ladefoged 1996: 182; Weenink 2013: 217).

4.3.1 The Source-Filter Theory

More specifically than stated above, the source of a speech sound can be considered more than simply the vibrating vocal folds, and the consideration depends on the sound to be produced. After air is blown out of the lungs, the airstream has to pass through the vocal cords, which can act as a narrow constriction, if not in action for voicing. For (voiceless) fricatives, the source is thus the turbulence caused by the constriction. For stops, the source is the contact of the suddenly released air with inert air (Ladefoged 1996: 111-112; Thomas 2011: 30).
For voiced sounds, the vocal folds vibrate at a certain frequency, referred to as the fundamental frequency. The vibration is caused by the airstream from the lungs which builds up pressure that forces the vocal folds to open. The pressure drops as soon as they are open, so that they close again. They continue this process until the airstream ends. In the opening phase of the vocal folds, the airstream bursts upwards and sets the downstream body of air in the pharynx above them vibrating (cf. Johnson 2003: 79; Ladefoged 1996: 93; Thomas 2011: 30). The vocal folds thus produce a complex periodic, non-sinusoidal waveform – variations in air pressure in the shape of pulses (Johnson 2003: 79).

The vibration of the vocal folds can be modified extensively by muscular tension, so that it is usually categorized as modal versus non-modal phonation. Modal phonation refers to the mode of a fundamental frequency distribution of an individual or the range of fundamental frequencies usually used for speaking (Gerrat and Kreiman 2001: 365). In phonetics, it is considered the typical phonation type and is associated with periodic vocal fold vibration, a well-defined glottal closure and a rich glottal spectrum (cf. Berry 2001: 431; Gerrat and Kreiman 2001: 366). Non-modal phonation is a generic term which is used to refer to any deviation from modal phonation (Berry 2001: 431), including breathy voice, falsetto, vocal fry (creak or pulse) or diplophonia among other phonatory registers (Gerrat and Kreiman 2001: 365-366; also cf. Esling and Edmondson 2011: 137-139; Esling 2013: 123-125). In a simplifying manner, phonation can be conceptualized as a continuum of breathy, modal and creaky voicing (Thomas 2011: 227). Breathy phonation is realized by having the vocal folds vibrate without much or no contact (Johnson 2003: 136), and creaky phonation by having them close sharply and stay closed for a relatively long period, so that the glottal pulses occur at irregular intervals (Ladefoged 2003: 176). Both of these and other non-modal phonation types require special attention, as they add some more considerations to an acoustic phonetic analysis. Since I disregarded vowel tokens uttered in non-modal phonation, especially in breathy and/or creaky phonation, in this analysis, I will focus on the specifics of the source and filter in modal phonation.

The fundamental frequency depends not only on the variation of muscular tension in the glottis, but also on the physiology of the speaker. If speakers stretch the vocal folds tightly, they move faster or vibrate at a higher frequency and create more pulses per second. If the cords are held loosely apart, they vibrate at a slower rate and create fewer pulses per second (cf. Ladefoged 1996: 99). Adult male speakers usually have a larger glottis than adult female speakers and children, so that adult males can create fewer pulses per second than children – perceived as low-pitched voices versus high-pitched voices. The fundamental frequency is measured by the number of times the complex waveform is repeated per second. If one complete cycle or period is 5.88 milliseconds long (the distance between the peaks of pressure in a waveform; the wavelength of a sound; Ladefoged 1996: 115), the fundamental frequency is 1 second $\div 0.0058$ seconds = 172.41 Hertz, but more accurately 170 Hz due to rounding of the milliseconds – an F0
expected of female speakers (cf. Johnson 2003: 79). Male speakers’ (modal) fundamental frequency usually varies from 90 Hz to 140 Hz, that of female speakers from 170 Hz to 220 Hz, and that of children is higher than 200 Hz. As outlined above, all speakers can vary their range of F0 well beyond the modal ranges (Thomas 2011: 31).

When the complex waveform from the glottis is decomposed into its frequency components and their amplitudes via a Discrete or Fast Fourier Transform (DFT or FFT) analysis, the fundamental frequency is represented as the first (or lowest-frequency) peak in this power spectrum. All of the peaks – frequency components or a series of simple waves extracted from the complex waveform – are called harmonics (Johnson 2003: 80). The subsequent peaks have to be integral multiples of the fundamental frequency or the first harmonic (H1), because its wavelength corresponds to the time between vocal fold vibrations. Any waveform with an amplitude of zero at the same places as F0 will ‘fit’ between the glottal pulses, and all multiples of F0 meet that constraint (Thomas 2011: 21): For an adult male speaker, the fundamental frequency (F0 = H1) may be 90 Hz, so that H2 = 180 Hz, H3 = 270 Hz… and H10 = 900 Hz. The height of the fundamental frequency is thus represented by the distance between the individual harmonics: For an adult female speaker, the fundamental frequency may be 200 Hz, so that H2 = 400 Hz, H3 = 600 Hz, etc.

The schema of the power or FFT spectrum of the fundamental frequency outlined in Figure 4.1a represents a voice that would not be filtered by the vocal tract. The frequency in such a spectrum is represented on the x-axis, the amplitude or amount of acoustic energy is represented on the y-axis. Time, as another important information in this regard, is missing (cf. Section 4.4.4). The harmonics are represented schematically as vertical lines or poles. The higher the frequency components, the larger the gradual decrease in amplitude (Thomas 2011: 31). In modal phonation, it usually decreases by approximately 12 decibels for each doubling of the frequency. In other words, the spectrum of the glottal pulse has a slope of -12 dB per octave, as an increase in one octave is a doubling of the frequency (Ladefoged 1996: 103). The relative amplitudes of the fundamental frequency and the second component frequency are related to certain non-modal phonation types: In breathy phonation, the amplitude in F0 is considerably higher than in the second harmonic, and in creaky phonation, the second harmonic is clearly higher in amplitude than the fundamental frequency (cf. Johnson 2003: 137).

After pulses are produced by the vocal cords, the pulses move through the vocal tract, which is considered to be an acoustic filter. There are several types of filters such as low-pass, high-pass and band-pass filters. As their names imply, low-pass filters pass low frequencies and block the higher ones, high-pass filters pass the high frequencies and band-pass filters pass frequencies inside the specified range or band. The vocal tract acts as a band-pass filter that passes some frequencies better than others. Disregarding the movement of non-stationary articulators such as tongue, lower jaw and lips, the filter’s length, diameter and the fact that it is closed at the end of the glottis largely determine
4. Data and Methodology

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
</tr>
<tr>
<td>600</td>
<td></td>
</tr>
<tr>
<td>800</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>1400</td>
<td></td>
</tr>
<tr>
<td>1600</td>
<td></td>
</tr>
<tr>
<td>1800</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
</tr>
<tr>
<td>2200</td>
<td></td>
</tr>
</tbody>
</table>

(a) A schematized power spectrum of unfiltered glottal pulses at 200 Hz.

(b) A schematized power spectrum of schwa after appropriate filtering.

Figure 4.1: Schematized power spectra of the source pulses with no filter in (a) and with the filter function superimposed in (b). The harmonics are represented as vertical lines. The fundamental frequency is 200 Hz in both spectra, so the harmonics come at multiples of 200. The superimposed filter in (b) is appropriate for schwa of an adult female speaker. The peaks in the spectrum are the resonances of the filter. Figures adapted from Thomas (2011: 31-32).

The vibration of the air in the vocal tract causes the air particles at the open end to move backward and forward. This movement eventually causes the air outside the lips to move. As the sound leaves the vocal tract, the movement changes the air pressure that is audible to others only at some frequencies (Ladefoged 1996: 104): the higher the frequency, the greater the response of the surrounding air outside the lips to the air vibrating in the vocal tract. This effect – radiation impedance – boosts the higher frequencies by 6 dB per octave. In other words, radiation impedance reduces the decrease in amplitude at higher frequencies.
of -12 dB per octave in the spectral slope of the pulses to a general slope of -6 dB per octave in a power spectrum of a filtered sound that has left the vocal tract. The peaks in such an output spectrum are superimposed on the general slope (Ladefoged 1996: 105).

As outlined above, the vibration of the vocal folds (source) is independent of the resonances of the vocal tract (filter). The frequencies produced by the vocal folds are determined by the size of the glottis and the muscular tension. Similarly, the resonant frequencies are determined by the shape of the vocal tract and the position of the articulators such as tongue, lower jaw and lips. The voiceless glottal fricative /h/, for instance, is generated without regular pulses from the vocal folds, and consequently has no fundamental frequency. Adjacent voiced sounds such as vowels, however, determine the position of the articulators which are similar in position of the glottal fricative. In *house*, for instance the articulators will be in the position to realize the nucleus of the diphthong *mouth*. The frequency components of the glottal fricative thus have relative amplitudes similar to, but smaller than, those in the adjacent vowels (Ladefoged 1996: 112).

The shape and configuration of the filter specify the qualities of voiced sounds in general and vowels in particular. Among others, the relative frequency of the source pulses over time are a strong indicator of whether a vowel is stressed or unstressed.

### 4.3.2 Tube Models

Tube models attempt to explain how the particular resonances of the vocal tract, the peaks in the spectral slope, arise. In the simplest form, as outlined in Figure 4.2, the Source-Filter Theory assumes the vocal tract is one tube of a uniform diameter of 1 centimeter and of a length of 17.5 centimeters (cf. Johnson 2003: 102). The length is an average value that helps to keep the arithmetics simple, as in reality most vocal tracts of speakers are shorter (Ladefoged 1996: 116). Although the vocal folds vibrate, the distance between them when they are open is comparatively small, so that the uniform tube is considered to be closed at the end of the glottis and to be open at the end of the lips (cf. Harrington 2010: 83). A tube like this occurs with unrounded vowels such as schwa /ə/ (Thomas 2011: 32), which have no significant constrictions in the vocal tract. In order to calculate the resonant frequencies of schwa /ə/, it is important to look at the way air vibrates in the vocal tract in response to the input of the voicing source.

When sound waves are produced, the air particles are perturbed or, in other words, each particle of air moves backward and forward (Ladefoged 1996: 114; Thomas 2011: 18). It thus passes its vibration on to the particles next to it, which takes time. Sound travels in air at about 35,000 cm per second, depending on altitude, temperature and humidity (Ladefoged 1996: 115; Thomas 2011: 34). If the vocal folds produce pulses at a fundamental frequency of 200 Hz, as outlined in the previous subsection, there will be 200 peaks of pressure spaced over 35,000 cm after one second. Put differently, every
4. Data and Methodology

peak of pressure of a sound source will be \(35,000 \div 200 = 175\) cm away from the next, or the wavelength of the sound is 175 cm. If the fundamental frequency is at 90 Hz, its wavelength is 388.9 cm. A digitally recorded (or synthesized) sound wave is typically presented in an oscillogram, as shown in Figure 4.3, for phonetic analysis, so that the wavelength is usually known. In order to determine the frequency, \(f\), the speed of sound in air, \(c\), in cm per second is divided by the wavelength, \(\lambda\) (lambda), in cm (cf. Ladefoged 1996: 115):

\[
f = \frac{c}{\lambda} = \frac{35,000}{\lambda}.
\]

The component frequencies of the complex waveform of the glottal pulses extend up to at least 4000 Hz with notable amplitude and with some of them being similar to the natural frequencies of the air inside the vocal tract. The amplitudes of these frequency components of the glottal pulses will be enhanced and the air is able to build
4.3. Acoustic Theory

up considerable vibrations (Ladefoged 1996: 117). The amplitudes of the frequencies not near the natural frequencies of the air inside the tube will be damped (Johnson 2003: 93). The air at the closed end of the tube cannot move back and forth, but the particles of air at the open end can have their maximum movement. The air in the tube can be compared to the behavior of a spring in that the air is cyclically maximally compressed (compression), under medium pressure (equilibrium) and maximally expanded (rarefaction; cf. Thomas 2011: 19, 33). In order to produce this cycle, the particles in the air may move quickly or slowly. The maximum flow of air particles is at the open end, because there is equilibrium, and zero flow is at the closed end of the tube, because there is either compression or rarefaction (cf. Figure 4.2). As becomes apparent, the air pressure varies within the tube inversely to the movement of air particles: The difference in pressure is zero (a node) at the open end, which is open to the medium atmospheric pressure of the surrounding air, and it is maximal (an antinode) at the closed end, where the particles are either compressed or rarefied. Consequently, the only constraint for tubes open at one end is that the resonances have to have a node at the open end and an antinode at the closed end (cf. Johnson 2003: 89; Ladefoged 1996: 118; Thomas 2011: 33). If the variations in movement and pressure are viewed as part of a sound wave, the wavelength can be estimated: Since the movement of air is zero at the closed end and at its positive maximum at the open end of the tube, it would change to zero again when the tube length was doubled. Consequently, the movement would reach its negative maximum when the tube length was tripled and reach zero again when it was quadrupled (cf. Ladefoged 1996: 119). Put differently, the waveform needs four tube lengths in order to complete one cycle: It starts with compression at the closed end, travels to the open end and reflects off the immobile mass of air outside the tube as rarefaction for the trip back to the closed end. It reflects at this point again, travels to the open end and switches back to compression when it reflects off the open end again, and travels back to the closed end as compression. Compression and rarefaction cancel each other out, so that the pressure equals the atmospheric pressure outside the tube. The waveform thus has to switch from one to the other at the open end of the tube (cf. Thomas 2011: 33-34). Using the equation outlined above, the frequency, $f$, in Hertz of the wavelength, $\lambda$, (which is four times the tube length, $L$, of 17.5 cm) can be determined as follows:

$$f = \frac{35,000}{4 \times 17.5} = \frac{35,000}{70} = 500 \text{ Hz}.$$}

Therefore, a sound wave with a frequency of 500 Hz corresponds to one of the modes of vibration of the air in the vocal tract when it is configured to produce the neutral vowel [ɔ] (cf. Ladefoged 1996: 119); in fact, it is the lowest resonance for that vowel. Another mode of vibration does not only have a maximum movement of air at the open end of the tube, but also at certain areas inside the tube. Likewise, there is not only minimal movement of air at the closed end but also elsewhere. The waveform fits the constraint of having a
Figure 4.4: A graph of a sound wave with $\frac{3}{4}$ of its wavelength within the tube. Figure adapted from Ladefoged (1996: 121).

node at the open end and an antinode at the closed end as well (Ladefoged 1996: 120), but it has switched from compression to rarefaction earlier than the lowest resonance (cf. Thomas 2011: 34). The period of such a waveform is one-third the duration of the lowest resonance calculated above. Put differently, the second resonant frequency is three times the first resonant frequency. In general, resonant frequencies of tubes closed at one end and open at the other are odd multiples of the first resonance (Johnson 2003: 94). Figure 4.4 outlines this relationship for the second resonant frequency. If one cycle of the sound wave is taken as the whole wave, then three-fourths of it are within and one-fourth is outside the tube. The dotted lines show that the tube is divided into thirds. With this division, an extra one-third is needed in order to complete one cycle of this wave. In other words, the wavelength, $\lambda$, is four-thirds of the tube length, $L$, or the wavelength is four times the tube length divided by three: $\lambda = 4 \times \frac{L}{3}$, so that the second lowest resonance has a frequency of 1500 Hz:

$$f = \frac{c}{\lambda} = \frac{35,000}{4 \times \frac{L}{3}} = \frac{35,000 \times 3}{4 \times L} = \frac{35,000 \times 3}{70} = 500 \times 3 = 1500 \text{ Hz}.$$ 

The same general rule of thumb for tubes closed at one end and open at the other end is valid for the third resonant frequency. In this case, the tube can be divided into fifths ($L/5$), because the wavelength of this resonance is shorter, which makes apparent that only four-fifths are needed for a complete cycle. The wavelength is thus four times the tube length divided by five: $\lambda = 4 \times \frac{L}{5}$. The third resonant frequency is therefore five times the first resonant frequency in Hertz (cf. Johnson 2003: 94):

$$f = \frac{c}{\lambda} = \frac{35,000}{4 \times \frac{L}{5}} = \frac{35,000 \times 5}{4 \times 17.5} = \frac{35,000 \times 5}{70} = 500 \times 5 = 2500 \text{ Hz}.$$ 

The neutral vocal tract has one resonant frequency at approximately 500 Hz, another one at approximately 1500 Hz and a third one at approximately 2500 Hz. These resonant
frequencies and their neighboring harmonics (cf. Section 4.3.3 below) show up as peaks in a power spectrum of schwa /ə/ and are commonly referred to as formants (cf. Berry 2001: 433; Harrington 2010: 83; Johnson 2003: 96; Ladefoged 1996: 121; Thomas 2011: 33). The first formant corresponds to the lowest resonance in which $\frac{1}{4}$ of the sound wave is within the tube: $F_1 = \frac{c}{4L}$; the second formant is the next-lowest resonance in which three-quarters of a wave is within the tube: $F_2 = \frac{3c}{4L}$; and the third formant is the next resonance in which five-fourths of the wave are within the tube: $F_3 = \frac{5c}{4L}$. The general formula of the formants (air vibrations), which are caused by the complex waveform of the glottal pulses, in a neutrally configured vocal tract – a tube closed at one end and open at the other – is:

$$F_n = \frac{(2n - 1)c}{4L},$$

where $n$ is any integer, the formant number, $c$ is the speed of sound in air, and $L$ is the length of the vocal tract tube (cf. Berry 2001: 433; Johnson 2003: 96; Ladefoged 1996: 122; Thomas 2011: 34). These considerations do not stop at the third formant, but continue indefinitely. As outlined in the previous section, the spectral slope of a sound past the lips will decrease by six decibels per octave. The amplitudes of the formants above 4500 Hz for schwa uttered by a speaker with a tube length of 17.5 cm will thus be too small to consider. What is important is the pattern of the formants for a neutral vocal tract: There is one formant to be expected every 1000 Hz, starting from 500 Hz and going up to infinity (Ladefoged 1996: 123).

This one-tube model is, however, not capable of calculating the resonant frequencies of higher (closer) or lower (more open) vowels, as it assumes a neutral vocal tract shape. For low back vowels such as merged PALM-LOT, both /ə/ in most American varieties of English, the tongue is low and its root pulled backwards, constricting the pharynx, so that a narrow tube is formed which is closed at the glottis and open to the wider cavity in front of the tongue. More accurately, but nevertheless simplifying, the vocal tract shape of such vowels has to be thought of as two tubes in order to account for the difference in cross-sectional area within the glottis and the oral cavity: a vertical one (or pharynx tube) of a small diameter and a horizontal one (or mouth tube) of a diameter larger than that (cf. Ladefoged 1996: 123; Thomas 2011: 33). In addition, other kinds of tubes – closed at both ends or a Helmholtz resonator (see below) – account for the constriction caused between the tongue and the hard palate with mid and high (close) vowels such as DRESS /ɛ/ or FLEECE /i/ (Thomas 2011: 36).

The pharynx and mouth tubes are both considered to be comparatively closed at one end (the glottis and the tongue root) and open at the other (the tongue root and the lips) and to be of equal length, i.e. each is half the vocal tract length of approximately 17.5 cm. For tubes with a diameter less than maximally one-fourth of the length of the tube, the diameter can be disregarded as a variable in the equation to calculate the resonant frequencies: In the production of vowels, the diameter may vary from a few
millimeters to a maximum of approximately 3.5 cm (Ladefoged 1996: 124). With the tube lengths being half that of the vocal tract tube as a whole, the resonant frequencies should be twice those of the whole vocal tract: \( F_1 = \frac{c}{4L} = \frac{35,000}{4 \times 8.75} = 1000 \) Hz. In fact, they differ from those predicted values, as both tubes are not completely closed at one end as the movement of air particles in the mouth tube interacts with that of the pharynx tube. Consequently, the lowest resonance in the mouth tube is at approximately 1100 Hz and at approximately 900 Hz in the pharynx tube, which is close to the formant frequencies of the merged PALM-LOT vowel (Ladefoged 1996: 124). The movement of the tongue back and fourth in the mouth can modify the length of each of the tubes at the cost of the length of the other. Since the longer tube will always produce the lowest resonant frequency, which corresponds to the first formant, a two-tube model can account for the first formant being always below 900 - 1000 Hz, but only to a value as low as 500 Hz, given the total vocal tract length is 17.5 cm (Ladefoged 1996: 125).

The vocal tract shape of rounded vowels such as GOOSE /u/ can be thought of as one or two tubes closed at both ends (Thomas 2011: 35). Waveforms traveling in such tubes do therefore not switch from compression to rarefaction or vice versa at either end of the tube. They start at the end of the glottis as compression, travel across the tube, reflect off the end of the rounded lips and travel back to the end of the glottis. While they are reaching the end of the tube at the lips, rarefaction sets in at the end of the glottis and travels to the end of the lips after the compression waveform reflected off of that end. Waveforms thus have to fit the constraint of one node in the middle of the tube, as compression and rarefaction cancel each other out at this point, and two antinodes at either end of the tube (cf. Johnson 2003: 89). The general formula of the formants (air vibrations), which are caused by the complex waveform of the glottal pulses, in such a vocal tract – a tube closed at both ends – is (Johnson 2003: 91; Thomas 2011: 35):

\[
F_n = \frac{nc}{2L},
\]

where \( n \) is any integer, the number of formant, \( c \) is the speed of sound in air and \( L \) is the length of the tube.

High (close) vowels usually have frequencies below 500 Hz; for FLEECE /i/ in particular, the lowest resonance is at approximately 250 Hz. The vocal tract shape of such vowels has a narrow constriction added to the open end of a tube where the tongue is moved closely toward the hard palate, which can be simplified as a Helmholtz resonator (Ladefoged 1996: 126; Thomas 2011: 36). Much like a bottle, the vocal tract has a large body of air between the glottis and the constriction at the raised front of the tongue, and a small body of air in the channel between tongue front and hard palate. The area between the constriction and the lips can be thought of as another tube (cf. Figure 4.5). The vibrating air in the large body acts like a spring, and in the small body it acts like a weight, vibrating back and forth on the spring. The rate of vibration (frequency) depends
4.3. Acoustic Theory

Figure 4.5: A schematized graph of a Helmholtz resonator for the vocal tract shape in the vowel FLEECE /i/. The tube closed at both ends and the constriction together form a Helmholtz resonator. Figure adapted from Thomas (2011: 35).

on the volume in the wide tube and on the mass of the air in the narrow tube (Ladefoged 1996: 126). In order to calculate the frequency of a Helmholtz resonance, the volume, $V$, of the wide tube as well as the cross-sectional area, $A$, and length, $L$, of the narrow tube have to be known (cf. Thomas 2011: 36):

$$f = \frac{c}{2\pi} \sqrt{\frac{A}{VL}},$$

where $c$ is the speed of sound in air. If the volume of the larger tube is 60 cm$^3$ from the glottis to the constriction, the cross-sectional area of constriction is 15 mm$^2$ and the length of the constriction is one cm, then the frequency of the Helmholtz resonance would be 280 Hz (Ladefoged 1996: 127):

$$f = \frac{35,000}{2\pi} \sqrt{\frac{0.15}{60 \times 1}} \approx 280 \text{Hz}$$

If the tongue is moved back into the velar region, appropriate for high (close) back vowels such as GOOSE /u/, the length of the larger tube would be reduced by one to two cm, so that the volume of this tube would be reduced likewise to about 50 cm$^3$. The Helmholtz resonance would then have a frequency of 300 Hz, which is an appropriate first formant value for a (non-fronted) GOOSE (Ladefoged 1996: 128). In a similar manner, the cross-sectional area becomes wider for lower (more open) vowels such as KIT /i/ and DRESS /ɛ/, so that the equation for the Helmholtz resonance returns values appropriate for the first formants in these vowels as well. However, the vocal tract configuration does not meet the assumptions of the Helmholtz resonators for other vowels than high (close) ones (Ladefoged 1996: 129).

Another way of calculating the formant frequencies produced by a vocal tract is that of modeling the vocal tract as a series of closely coupled tubes. As noted above, one
or two tubes open at one end occur with unrounded vowels; tubes closed at both ends occur with rounded vowels, as the rounding of the lips narrows the open end of tubes considerably; and Helmholtz resonator-like tubes (a kind of tube closed at both ends) occur with high (close) vowels (cf. Thomas 2011: 32, 34, 36). When the vocal tract is modeled as a series of tubes, the change in formant values varies, depending on the region in the mouth in which the constriction is caused by the tongue. The place of constriction varies roughly according to the place of the contextual consonants of each vowel. The point farthest from the glottis is between the tongue and alveolar region and the point nearest to the glottis is between the tongue and pharyngeal region.

Assuming the vocal tract has a very open lip position, the frequencies of the first formant are very low (250 Hz) at the farthest point and increase gradually as the constriction moves toward the nearest point (500 Hz). The second formant frequencies increase slowly from approximately 1800 Hz to 2500 Hz in the velar region and decrease drastically to approximately 1200 Hz in the pharyngeal region. The third formant frequencies behave inversely to the second formant frequencies, as they decrease until the constriction reaches the velar region from approximately 3200 Hz to 2700 Hz and become higher in the pharyngeal region (3000 Hz; cf. Ladefoged 1996: 132-133). The recalculation of the formant frequencies when lip rounding is added to each vocal tract shape shows that the first formant is largely unaffected, the second formant is lowered when the constriction occurs in the palatal, velar and uvular regions, and the third formant is always lowered. The reason for these changes in resonant frequencies is that lip rounding affects the cavity of the vocal tract in front of the constriction caused by the tongue. When this front cavity is short, its resonant frequency corresponds to the third formant, and the resonant frequency of the back cavity corresponds to the second formant. If the front cavity is lengthened via moving the constriction backward, the resonant frequency that corresponds to the third formant switches to correspond to the second formant after the tongue reaches a certain point (here the velar region), depending on the degree of lip rounding. The third formant then becomes a higher resonance of the remainder vocal tract (cf. Ladefoged 1996: 133).

Perturbation theory provides an alternative to tubes in order to model vocal tract constrictions. The relationship between air pressure and velocity plays an important role in this approach (Johnson 2003: 107). The movement of air particles is manipulated by constriction and flaring in a tube open at one end. Constrictions limit the room air particles have to move around in, so that they reach maximum pressure faster. If the constriction is at an antinode (points of minimal velocity and maximum pressure), the waveform will gain speed, which increases its frequency. If the constriction is at a node (points of maximum velocity and atmospheric pressure), the waveform will loose speed, which decreases its frequency (cf. Johnson 2003: 110; Thomas 2011: 35). Flaring at one end of the tube does exactly the opposite of constriction: air particles have more room to move around and are also sped up (Thomas 2011: 36). The formant values calculated with the assumptions of vocal tract configurations made by the tube models
4.3. Acoustic Theory

Figure 4.6: Two digitally stored waveforms, showing in detail how each point in the smoothed wave at the bottom is the mean of five points in the noisy wave at the top. Figure taken from Ladefoged (1996: 184).

are consistent with the perturbation theory analysis of the same vocal tract configuration. Perturbation theory is, however, more accurate when formant frequencies of sounds with more than one vocal tract constriction as in /æ/ are to be calculated. On a more general note, the assumptions of tube models are more closely met in articulations that have narrow constrictions and that of perturbation theory by articulations in which the vocal tract is mainly unobstructed (Johnson 2003: 111).

4.3.3 Linear Predictive Coding (LPC)

In the 1970s, speech engineers developed several algorithms to locate and determine formants automatically, by calculating the Linear Predictive Coefficients for a set of sample points of a (complex) sound wave (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 93; Johnson 2003: 40). At present, this is the preferred way of determining the quality of vowels in phonetic analysis (Ladefoged 1996: 181, 2003: 120; also cf. Johnson 2003: 99; Weenink 2013: 217). The most extensive description of how LPC analyses work in terms of linguistic applications and uses is that by Ladefoged (1996), so that this subsection will largely be based on his explanations.

The LPC analysis is in some respects the reverse of the process of speech production. Its basic notion is that a speech signal serves as an input to the LPC filter, which is the inverse of the spectrum of the speech signal. The filter will produce an output as close to zero as possible. Despite the same shape, the LPC filter differs from the vocal tract filter in that the spectral characteristics of the glottis and the lip radiation are incorporated into the same filter as the one representing the characteristics of the vocal tract. The LPC analysis assumes that the source and the filter are independent of one another (Ladefoged 1996: 182; Weenink 2013: 217).

In order to outline the function of a digital filter such as the LPC filter, it is necessary to look at the digital representation of a sound wave. A digitized waveform is stored on a computer hard drive via a certain amount of samples representing the amplitude of the analog waveform at discrete moments in time (cf. Figure 4.6). A digital filter such as the
LPC filter modifies these sample points in one way or another, depending on the desired outcome of the filtering process. Ladefoged (1996: 183) suggests the moving average (MA) filter as an example, in order to understand the modification of the sample points. An MA filter takes, for instance, five sample points of a digitized speech wave that may have some added noise components, e.g. after a truck drove by the recording locality. The MA filter calculates the mean of those five points in the noisy wave in order to replace the sample point in the middle of the five points by this mean value in the smoothed output wave (cf. arrow in Figure 4.6). After the filter has replaced every point in the noisy waveform in that way, the output of the MA filter is a smoothed waveform that is close in shape to, but not exactly, the one the speech wave would have had, without a truck driving by (cf. Ladefoged 1996: 184). The number of points considered is of course not limited to five: the more points the MA filter takes into calculation, the smoother the output wave. The two drawbacks of this configuration of the MA filter are that it takes each of the five points into account equally when smoothing the waveform and, more importantly, that it requires two points after the one being considered. This means that the output point at any moment depends on two input points at future moments, which complicates the filter as its output needs to be delayed in order to do the calculations (cf. Ladefoged 1996: 186). An easier filter only takes the points at past and present moments into account. It can be generally specified in that the output amplitude of the \( n^{th} \) sample (\( y_n \)) of the smoothed wave is equal to an arbitrary weight (\( b_0 \)) given to the input amplitude of the \( n^{th} \) sample (\( x_n \)) of the noisy wave, as illustrated in the following formula: 

\[
y_n = b_0 x_n.
\]

To this one sample point \( x_n \) of the noisy wave, previous sample points such as \( x_{n-1}, x_{n-2}, x_{n-3} \), etc. can be added similarly (cf. Ladefoged 1996: 187):

\[
y_n = b_0 x_n + b_1 x_{n-1} + b_2 x_{n-2} + b_3 x_{n-3}, \ldots
\]

The weights (\( b \)) are referred to as the coefficients of the filter. In order to fit the right value to those weights, they are numbered by subscripts according to the sample point in time they belong to. The first coefficient is \( b_0 \), because it refers to time \( n = 0 \), the second coefficient is \( b_1 \), the third coefficient is \( b_2 \), etc. In the example of the MA filter above, each of the coefficients (\( b \)) was \( \frac{1}{5} \) times the input amplitude of the \( n^{th} \) sample (\( x_n \)), because all of the five points considered in the MA filter were weighted equally in the calculation. The difference between the MA filter and this example of a filter is that in the MA filter the additional sample points were not exclusively previous ones (e.g. \( x_{n-1} \)), but two of them were future sample points (\( x_{n+1} \) and \( x_{n+2} \)).

If the input signal (\( x \)) into the filter outlined in the formula above consisted of a single pulse at a time (\( n \)) zero with an amplitude of one, then \( x_n = 1 \) when \( n = 0 \), but \( x_n = 0 \) when \( n \) is not zero (Ladefoged 1996: 187; cf. schematized input on the left in Figure 4.7). Put differently, when \( n = 0 \), then \( x_0 = 1 \), \( x_{(0-1)} = 0 \), \( x_{(0-2)} = 0 \) and \( x_{(0-3)} = 0 \). In order for the filter to have an effect, the coefficients (\( b \)) will be assigned the following
4.3. Acoustic Theory

arbitrary values: $b_0 = 0.9$, $b_1 = -0.5$, $b_2 = 0.6$ and $b_3 = -0.2$. This particular filter is then a fourth-order filter, since there are four coefficients (cf. Ladefoged 1996: 187). When all of these values are substituted for $b$ and $x$ in the equation, the output amplitude ($y_0$) is 0.9:

$$y_0 = 0.9(1) - 0.5(0) + 0.6(0) - 0.2(0) = 0.9.$$ 

When $n = 1$, $x_1 = 0$, $x_0 = 1$, $x_{-1} = 0$, $x_{-2} = 0$,

$$y_1 = 0.9(0) - 0.5(1) + 0.6(0) - 0.2(0) = -0.5.$$ 

When $n = 2$, $y_2 = 0.6$ and when $n = 3$, $y_3 = -0.2$. The output of this filter is thus dependent on the input at the present moment and the inputs at the past moments. The output values correspond to the coefficients: $y_0 = b_0 = 0.9$, $y_1 = b_1 = -0.5$, $y_2 = b_2 = 0.6$ and $y_3 = b_3 = -0.2$. Once the filter coefficients are known, the frequency components of the output wave can be determined (Ladefoged 1996: 190). Put differently, the spectrum of the output wave can be calculated when the shape of the output wave is known (for a graphical representation see Figure 4.7).

A Linear Predictive Coding (LPC) analysis tries to predict a filter that has a resonance curve which is the inverse of the spectrum of an input wave (Ladefoged 1996: 190). In other words, it calculates estimates of the vocal tract resonances by taking a small duration of an acoustic waveform (Johnson 2003: 97). It does so by taking a set of coefficients to predict one point in a sampled waveform from known values of previous points, and multiplying each of them by a coefficient (Ladefoged 1996: 190). The underlying assumptions are that the input wave is nonrandom, that it has regularities that can be predicted, i.e. the waveform characteristics do not change too much, and that the vocal folds are modeled as a pulse train or white noise (Ladefoged 1996: 190; Weenink 2013: 217). Hence, the basic principle is that any point can simply be regarded as the sum of a number of previous points, each of which has been multiplied by a suitable positive or
negative number – the Linear Predictor Coefficients. In particular for damped sinusoidal waves, as shown in Figure 4.8, any two points in a sampled waveform are sufficient to fully determine the value of all succeeding points (Ladefoged 1996: 191). Formants are damped sinusoidal waves (cf. Page 162). Many speech sounds are the sum of a number of formants repeated at intervals corresponding to pulses from the vocal cords (Ladefoged 1996: 192). Since the goal of an LPC analysis is to determine the formants of a speech sound such as a vowel or sometimes a semi-vowel, at least two Linear Predictive Coefficients are needed per formant. As mentioned in Subsection 4.3.2, for a male voice, there will be a formant in each 1000 (to 1100) Hz band of a digital speech wave. For a female speaker, there will be a formant in each 1100 (to 1200) Hz band, due to shorter vocal tracts. In this analysis, the speech waves were digitally recorded at 48,000 Hz and 24 bit. For the LPC analysis, each of the sound signals is re-sampled to 22,000 Hz, which provides a calculation range – the Nyquist frequency\(^{45}\) – of 11,000 Hz. Within this Nyquist frequency, approximately 11 formants can be expected on average, so that the coefficients of the LPC analysis should be set at least to 22 as the default for the LPC calculation: \(11 \times 2 = 22\) (cf. Ladefoged 1996: 193, 2003: 123). In order to account for higher formants that may influence the spectrum or a pole due to the glottal pulse shape, two more coefficients should be added to this rule of thumb: \(11 \times 2 + 2 = 24\). If the vocal tract of a speaker is shorter, few formants can be expected within the same Nyquist frequency, so that the coefficients have to be reduced as well: \(10 \times 2 + 2 = 22\) (Ladefoged 1996: 212, 2003: 125). As can be seen in these examples, choosing the right number of coefficients has to be tested for each speaker individually. Too many coefficients will result in an LPC analysis that produces poles or peaks corresponding to spurious formants (see below). If too few coefficients are specified, two formants in close proximity may be miscalculated as a single peak (Ladefoged 1996: 212). A careful testing of how many coefficients are necessary for an LPC calculation is shown in Subsection 4.4.4.2.

\(^{45}\) The Nyquist frequency describes the bandwidth of a signal that can be used (e.g. for phonetic analysis). In order to digitize an analog sound wave, the sampling rate has to be twice the frequency of the analog signal, so that the peak and the valley (periodicity) of one full cycle or one period can be fully represented in the digital signal. Thus, if the analog wave has a frequency of 100 Hz, the sampling rate has to be 200 Hz. Put differently, the Nyquist frequency is half the sampling rate (cf. Mayer 2010: 87).
In a simplified example of how the LPC analysis predicts succeeding points of a wave, only four Linear Predictive Coefficients \((a_1, a_2, a_3\) and \(a_4\)) are considered. The wave of arbitrary amplitude to be analyzed consists of twelve sample points \((s_1 \text{ through } s_{12})\). In order to estimate the value of a random sample point \((\hat{s}_n); \text{ the circumflex on } s_n \text{ denotes that it is an estimated value})\), the four previous sample points \((s_{n-1}, s_{n-2}, s_{n-3} \text{ and } s_{n-4})\) are multiplied with the yet undetermined LPC values \((a_1, a_2, a_3 \text{ and } a_4); \text{Ladefoged 1996: 193; also cf. Johnson 2003: 97})\):

\[
\hat{s}_n = a_1 s_{n-1} + a_2 s_{n-2} + a_3 s_{n-3} + a_4 s_{n-4}.
\]

In order to determine values for the unknown coefficients, the sample points need to have an arbitrary value attributed to them, so that each equation has only four unknowns. The values of the sample points is arbitrary, because the wave under analysis has an arbitrary amplitude. Since there are thus four unknown coefficients, four equations are necessary to determine them. Assuming that each of the estimated sample points is correctly predicted, they should be identical to the corresponding sample points. Put differently, if the values of the estimated sample points are subtracted from the values of the sample points, the result should be zero: \(0 = s_n - \hat{s}_n\) (Ladefoged 1996: 195). In this case, the equations for estimated sample point six \((\hat{s}_6)\) to estimated sample point nine \((\hat{s}_9)\) with the values from Table 4.5 are (Ladefoged 1996: 194):

\[
0 = s_6 - \hat{s}_6
\]
\[
0 = s_6 - (a_1 s_{(6-1=5)} + a_2 s_4 + a_3 s_3 + a_4 s_2)
\]
\[
0 = -40 - (-74a_1 - 42a_2 + 17a_3 + 50a_4).
\]

The other three equations are reduced and summarized, so that only the last line is presented. The values for the sample points can be found in Table 4.5.

\[
0 = -4 - (-40a_1 - 74a_2 - 42a_3 + 17a_4)
\]
\[
0 = 22 - (-4a_1 - 40a_2 - 74a_3 - 42a_4)
\]
\[
0 = 49 - (22a_1 - 4a_2 - 40a_3 - 74a_4).
\]

Thus, the LPC coefficients are \(a_1 = 0.5, a_2 = -0.6, a_3 = 0.4\) and \(a_4 = -0.7\) (Ladefoged 1996: 196). These values are appropriate for the four estimated sample points, but would

<table>
<thead>
<tr>
<th>n</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude</td>
<td>90</td>
<td>50</td>
<td>17</td>
<td>-42</td>
<td>-74</td>
<td>-40</td>
<td>-4</td>
<td>-4</td>
<td>22</td>
<td>49</td>
<td>38</td>
<td>1</td>
</tr>
</tbody>
</table>
result in an error if used to estimate a fifth point from a different set of four points. The error for any sample point \( s_n \) is indicated by \( e_n \), where \( e_n = (\hat{s}_n - s_n)^2 \). Depending on the exact values of estimated sample point (\( \hat{s}_n \)) and sample point (\( s_n \)), subtracting them from each other may result in a negative value. The error is the square of the result of the subtraction, to make the resulting number a positive one (cf. Ladefoged 1996: 197).

Since the values for the LPC filter should be appropriate for any set of four points, the errors for each set of four points are included and summed to a total error in the LPC algorithm. The algorithm solves the sets of simultaneous equations for the twelve sample points in the window while trying to minimize the sum of the errors. Since four out of the twelve points are needed to determine the succeeding ones, there are \( \binom{12}{4} = 8 \) sets of four points that can be used to determine a fifth point, each with an error \( e_5 \) to \( e_{12} \). The problem is to find values of the coefficients such that each contributes as little as possible to the total error in the window (Ladefoged 1996: 198). After multiplying each term in the parenthesis by all the others of the eight errors and rearranging them, eight terms of the form \( a_1^2 s_n^2 \) (from \( a_1^2 s_5^2 \) to \( a_1^2 s_{12}^2 \)) are derived (plus another eight for each coefficient \( a_2, a_3 \) and \( a_4 \)) and for each term containing \( a_1 \) out of the eight terms there are two of each of the other terms (e.g. \( +a_1s_4a_2s_3 \) and \( +a_2s_3a_1s_4 \)) in the LPC matrix (Ladefoged 1996: 204). The error for the first coefficient \( a_1 \) depends directly on those terms. If the sample points (\( s_n \)) are given all the values from Table 4.5, they can be summed and as a result be expressed in a quadratic equation (Ladefoged 1996: 205):

\[
\text{error } a_1 = 11711a_1^2 - 2a_1 + 18202a_1a_2 + 18202a_1a_3 + 18202a_1a_4,
\]

which is similar to the general quadratic equation

\[
y = mx^2 + nx.
\]

All quadratic equations can be visualized as parabolas with an axis of symmetry parallel to the \( y \)-axis. This means that at a certain value of \( a_1 \), \text{error } a_1 \) stops decreasing and starts increasing. This is the value of the error that minimally adds to the total error of the twelve sample point estimates, as a slight change in \( a_1 \) is enough to reverse the behavior of \text{error } a_1. The minimum value for the error curve or parabola is zero. Differentiating the equation of \text{error } a_1 and setting it to zero changes the equation as follows (Ladefoged 1996: 207):

\[
0 = 23422a_1 - 2 + 18202a_2 + 18202a_3 + 18202a_4.
\]

Similar equations can be formed for the minimum errors of the coefficients \( a_2, a_3 \) and \( a_4 \), so that the four equations with the four unknowns can be solved by the LPC algorithm.

The example has outlined how an LPC algorithm attempts to determine the minimum value for \( a \) that applies to this set of twelve points. The same procedure works for any number of sample points and coefficients (Ladefoged 1996: 207). The coefficients of the
4.3. Acoustic Theory

LPC analysis specify a filter or a response to an impulse in the form of a speech signal that served as an input. An FFT analysis (see below) is then made of the output wave, after it decayed to zero and after more zeros may or may not have been added to the wave. The resulting spectrum has a large number of component frequencies, so that the response of the LPC filter is represented by a smooth curve (Ladefoged 1996: 211; cf. Figure 4.9).

Using LPC Spectra is considered to be the best way to determine formant measurements of vowels (Ladefoged 2003: 120, also cf. Johnson 2003: 99). According to Di Paolo, Yaeger-Dror and Wassink (2011: 93, 97), however, LPC analyses are slightly inaccurate when compared to an FFT spectral analysis. They point out that the inaccuracy of LPCs has been shown in speech synthesis with long steady-state vowels synthesized at a fundamental frequency (or the first harmonic in an FFT spectrum) of 100 Hz – an average frequency for adult male speakers – or less. At this F0, the automatic formant measurements are usually most easily calculated, because the chances that common LPC errors occur are small (see below). In addition, in synthesized speech, the formants have to be entered into the formula that creates speech sounds, so that they are known before the LPC and FFT analyses are made (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 93).

Whether an LPC analysis is really less accurate than an FFT analysis remains to be verified, since both spectral analyses make simplified assumptions about the shape of the resonance tubes of natural speakers (cf. Johnson 2003: 157). Using an FFT spectrum instead of an LPC spectrum for a large-scale acoustic analysis of vowels is, however, not very helpful for the following reasons: An FFT spectrum makes no presumptions about formants. As outlined in the previous section, it simply determines the amount of energy (peaks in the spectrum) at each different frequency (Ladefoged 2003: 120). The peaks reflect the frequency of vibration of the vocal folds (source) and the configuration of the moving and stationary articulators such as the lips, tongue and lower jaw above the glottis (filter; cf. Di Paolo, Yaeger-Dror and Wassink 2011: 94). The peaks and valleys in the FFT spectrum (cf. Figure 4.9) thus reflect the resonance properties of the shape and length of the vocal tract for the vowel sound produced. Since the source and filter are independent of one another, the fundamental frequency produced by the vibration of the vocal folds can become higher (i.e. wider spacing between the harmonics) or lower (i.e. smaller spacing between the harmonics) without a change in resonant frequencies (the peak frequencies) as long as the configuration of the lips, tongue and lower jaw is not changed (Di Paolo, Yaeger-Dror and Wassink 2011: 95). Similarly, F0 may stay unchanged, but lip rounding may cause the peak frequencies corresponding to the second formant to be amplified at around 1500 Hz or more (e.g. a sound to be heard as a fronted GOOSE vowel), and spreading of the lips may cause the same vocal tract configuration to amplify peak frequencies at around 2500 Hz (e.g. a sound heard as a FLEECE vowel).

As outlined in the previous section, the peaks produced by the vocal folds are amplified if they are close in frequency to the resonances of the vocal tract. However, especially
with high-pitched voices ($F_0 \geq 200$ Hz), amplified peaks may not be visible in an FFT spectrum (Di Paolo, Yaeger-Dror and Wassink 2011: 95).

Since the frequency bands – or formants – consist of several neighboring overtones, i.e. a bundle of single harmonics as identified by a narrowband FFT spectrum, the resonant frequencies created in the vocal tract are not exact single frequencies to be measured (Mayer 2010: 83). The mucosae on the vocal tract walls do not reflect the sound energy completely, because they are soft. Instead they absorb some of the sound energy produced by the vibrating glottis, as does the inertia of air in and out of the vocal tract. Consequently the mucosae and the inertia of air damp the resonant frequencies of the vocal tract. This damping process is reflected in the bandwidth of the formants (Johnson 2003: 151). An undamped sine wave has a spectral peak which is infinitely narrow, i.e. the peak has an infinitely small or narrow bandwidth and thus it is sharply defined. Spectra of a damped sine wave have energy spread over other frequencies near the frequency of the peak (Johnson 2003: 150). The further the resonant frequencies are damped, the higher is the bandwidth value in Hertz of the formants. In other words, “the spectral result of damping a sine wave is to broaden the peak around the sine wave’s frequency” (Johnson 2003: 150-151). As a consequence, a formant is characterized by its frequency, which is the position of the peak in the LPC spectrum (cf. Ladefoged 2003: 120), and its bandwidth, which is the visible width of the peak in the LPC spectrum (Mayer 2010: 83). Put differently, a formant is simply a damped sinusoidal wave (Ladefoged 1996: 192). As an FFT spectrum only determines the harmonics, but not which of the neighboring overtones are combined to a resonant frequency, it is not clear which of the highest peaks in an FFT spectrum is the accurate position of where the formant should be measured as a single Hertz value (cf. F2 in Figure 4.9). It may also well be that three or four very high harmonics in a narrowband FFT spectrum are actually two formant peaks that cannot be distinguished, neither by a wideband (e.g. with a bandwidth of 344 Hz), nor by a narrowband FFT spectrum (e.g. with a bandwidth of 15.6 Hz; cf. Ladefoged 2003: 122).

As Figure 4.9 highlights, such a measurement can be more accurately made if the position of the formants in question is determined by an LPC spectral analysis with narrow bandwidths and consequently sharply defined peaks of the frequency band. Figure 4.9 shows an LPC spectrum superimposed on a narrowband FFT (see below) spectrum of a FOOT /u/ vowel, uttered by 28HLMF – a middle-aged urban female – in her word-list section of the interview. The $x$-axis reflects the frequency in Hertz and the $y$-axis the amplitude in decibels. I made this example of the narrowband FFT spectrum by extracting a speech sound of 64 milliseconds in duration at a sample rate of 16,000 Hz around the point of measurement at 33% of the duration of FOOT (cf. Subsection 4.4.1 for the point of measurement). The 64 milliseconds of the analysis window duration did not exactly match the signal period, i.e. the amplitudes of the waveform at the beginning and end of the selection did not cross zero. I thus used a Hanning window in order to
4.3. Acoustic Theory

Figure 4.9: LPC spectrum superimposed on a narrowband FFT spectrum of the vowel /u/. The LPC spectrum (on top of the single harmonics shown by the FFT) highlights the spectral peaks of the vowel (the formant peaks). The width of these peaks is the bandwidth of the frequency bands (formants). The first formant is characterized by the highest peak in the FFT and the first peak in the superimposed LPC, both roughly at 633 Hz. The second formant is characterized by the following highest peak in the FFT spectrum at 1455 Hz and by the second peak in the LPC spectrum at 1437 Hz ($\Delta F_2 = 18$ Hz). Without the superimposed LPC spectrum, it is not clear which of the two peaks in the FFT spectrum should be measured as F2.

set the amplitude of the waveform to zero near the edges of the window (cf. upper panel in Figure 4.10), as many implementations of the Fourier analysis assume such a setting (Johnson 2003: 35). The change in amplitude near the edges is derived from multiplying each point of the waveform with a factor of 0.0 at the edges and gradually changing this factor to 1.0 in the middle of the waveform, which leaves the amplitude values unchanged (Ladefoged 1996: 147). If the amplitude ended or began abruptly instead, the analysis would falsely assume that there is a considerable amount of high-frequency components present in the sampled waveform (Ladefoged 1996: 146). As a result, the spectral analysis of the original complex waveform would have a minimally different frequency, so that the spectrum would show a smaller difference between the amplitude of the highest peak and the lower peaks (Johnson 2003: 35). For the Fourier Transform to be ‘fast’, the window length in points has to be a power of 2 (e.g. 64, 128, 256, 512, 1024; cf. Johnson 2003: 33; Ladefoged 1996: 179). If the window length is set correctly, the FFT implements
the algorithm of a Fourier analysis of discrete signals (Discrete Fourier Transform; DFT) more efficiently than the DFT (cf. Mayer 2010: 91). The sample frequency in kHz times the duration of the window length in milliseconds equals the window length in points: 16 kHz \times 64 \text{ ms} = 1024 \text{ pts} (cf. Ladefoged 1996: 178). The FFT spectrum is a narrowband one, because the bandwidth is 15.6 Hz, as $\frac{16,000}{1024} = 15.6$ Hz (cf. Ladefoged 2003: 118; Mayer 2010: 93). The 1024-point window provides an accurate frequency resolution, which is important for the correct measurements of the component frequencies in vowels, but most likely fails to capture rapidly changing phenomena as in stop bursts (Ladefoged 1996: 177-178). In addition, Figure 4.9 highlights that the comparatively long selection from the signal shows a change in pitch, visible through the differences in spacing of the individual harmonics after the second formant measurement point (which does not substantially affect the filter frequencies).

The basis for the LPC spectrum is the same sound extraction as for the FFT spectrum. The difference is that for the LPC analysis I extracted the sound bit using a rectangular instead of a Hanning window (Johnson 2003: 33), which means that the amplitudes near the edges of the window are not changed to zero (cf. Figure 4.10). The LPC spectral slice was made at 33% of the vowel’s duration (38.38 seconds). The sample rate of the LPC spectrum is at 16,000 Hz and the coefficients (poles) were set to 16, as this proved to be the most reliable value for this speaker in general. I expected seven formants (14 coefficients) within the Nyquist frequency of 8000 Hz for this speaker and added two coefficients for higher formants (cf. Ladefoged 1996: 212, 2003: 125). As two coefficients are needed per formant (Ladefoged 2003: 125), one for the formant and one for the bandwidth (Ladefoged 1996: 211), the rule of thumb suggests 16 coefficients: $2 \times 7 + 2 = 16$.

According to the help section in Praat (Boersma 2011d), automatic formant tracks are produced by making an LPC analysis every 35 milliseconds, using the maximum formant frequency instead of the Nyquist frequency, and number of formants instead of the coefficients. An example of how settings like these were found to be reliable is provided in Subsection 4.4.4.2. The example will also outline the implications of incorrect settings such as the coefficients or poles in an LPC spectral analysis on the change in formant measurements.

In essence: In comparison to an LPC analysis, an FFT spectrum is not necessarily the better choice of measuring formants, because its accuracy is lost when the resonant frequencies to be measured cannot be determined from it. This may either be the case when there are no amplified resonant frequencies visible (sometimes with F0 $\geq$ 200 Hz) or when amplified peaks cannot be distinguished from one another (cf. F2 in Figure 4.9).

Di Paolo, Yaeger-Dror and Wassink (2011: 96-97) further provide two sets of measurements of formants one and two in the vowel FOOT in order to outline the discrepancy between them. One set was measured via an FFT spectral slice, the other one via an LPC spectral slice. Without any further specification of the settings for the calculation of the
4.3. Acoustic Theory

Figure 4.10: Extracted waveform of FOOT using a Hanning window (top) and a rectangular window (bottom).

<table>
<thead>
<tr>
<th></th>
<th>Di Paolo et al.’s</th>
<th>My values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FFT</td>
<td>LPC</td>
</tr>
<tr>
<td>F1</td>
<td>467</td>
<td>F1</td>
</tr>
<tr>
<td>F2</td>
<td>1494</td>
<td>F2</td>
</tr>
<tr>
<td>ΔF1D</td>
<td>17</td>
<td>ΔF1H</td>
</tr>
<tr>
<td>ΔF2D</td>
<td>146</td>
<td>ΔF2H</td>
</tr>
</tbody>
</table>

Table 4.6: Comparison of differences in formant measurements (in Hertz) of FOOT via LPC and FFT spectra. Note that the raw formant values cannot be compared to one another. Delta F1D and Delta F2D refer to the differences in values taken from Di Paolo, Yaeger-Dror and Wassink (2011: 96-97). Delta F1H and Delta F2H refer to the differences in values I measured as shown in Figure 4.9.
spectral slices, they illustrate that the LPC measurement is inaccurate. The differences in formant values are shown on the left in Table 4.6. Even if the measurements made in an FFT spectral slice are more accurate, the differences in the measurements I made in an FFT versus LPC slice for this particular vowel are not as high as indicated by Di Paolo, Yaeger-Dror and Wassink (2011: 96-97).

The accuracy of both spectra has to be understood in light of the following caveats: The calculation of a spectral curve depends on the different assumptions each analysis makes about the spectrum in that they take the fundamental frequency and the influence of spurious formants into account in different ways (Ladefoged 2003: 121). A spurious formant can appear as a formant very close to another (accurately measured) one, indicated by additional peaks in the spectral slices of FFT and LPC spectra. Spurious formants usually have very high bandwidths (> 400 Hz) and return formant values that do not necessarily make sense for the vowel under analysis (cf. Subsection 4.4.4.1 for an example). Usually, the values calculated for formants three and four can safely substitute the values of formant two and three if the spurious formant is located between formants one and two. The frequencies the spurious formants have can be understood as an indicator of the speaker’s voice quality (Ladefoged 2003: 125), instead of a resonant frequency due to the vocal tract shape. The intensity of the fundamental frequency depends on the phonation type of the glottis and the nearness to formant one. It can be measured in relation to the second harmonic in an FFT spectral slice or in relation to formant one. Very generally, the intensity will be much higher in breathy phonation than in modal phonation – the ‘normal’ type that provides the most accurate measurements – and much higher in modal phonation than in creaky voice (Ladefoged 2003: 181).

The spectral contributions of the fundamental frequency and spurious formants should be reduced as much as possible if the purpose of formant analysis is to determine information about tongue and lip movements. Additionally, for nasalized vowels and nasal consonants, the oral cavity is closed or nearly closed and acts as side branch of a larger resonant tube, which consists of the nasal cavity and the frontal sinus cavities (Johnson 2003: 151). The frequency components near the resonant frequencies of this side cavity are canceled in the side branch, and become anti-resonances, or anti-formants, which show up as spectral valleys in the spectrum (formants show up as peaks). The difficulty is then to differentiate passive lack of resonance from active anti-resonance in a spectral valley (Johnson 2003: 154). In addition, if the frequencies of formant and anti-formant are the about the same, the anti-formant does not show up as a valley in the spectrum anymore, but the formant peaks are weakened (Johnson 2003: 157). In summary, nasalization will affect the spectrum in that it introduces new resonances (Ladefoged 2003: 122). Important for this analysis is the vocal tract shape, reflected in the actions of the moving and stationary articulators. It is not clear which type of spectral analysis is able to measure formant frequencies associated with these actions more accurately (Ladefoged 2003: 122).
4.4 Measurements

The LPC analysis is, in addition, infamous for miscalculating formant peaks, especially when formants and harmonics are close to one another (Di Paolo, Yaeger-Dror and Wassink 2011: 94). For instance, if the fundamental frequency is higher than 200 Hz, the LPC algorithm may confuse formant one with the fundamental frequency. This affects only high or close vowels such as FLEECE and GOOSE, and upglides in diphthongs such as FACE, GOAT and PRICE. In addition to a high fundamental frequency, formants one and two are in close proximity for GOOSE and GOAT, so that the LPC analysis may consider formant one to have a broad bandwidth (taking formant one and two to be a single formant). With regard to FLEECE, formants two and three are in close proximity, so that the LPC algorithm may confuse those as well, and return values for formant two that suggest an unusual front position in the vowel space (Di Paolo, Yaeger-Dror and Wassink 2011: 94). Many of these errors can be found and to some extent accounted for when the spectral slices calculated with different sets of coefficients are carefully examined and checked against perception. However, if there are formant readings derived from an LPC analysis which appear to be incorrect despite testing different coefficients, an FFT analysis might be helpful to check measurements again before they are disregarded (Di Paolo, Yaeger-Dror and Wassink 2011: 94).

4.4 Measurements

This section lays the foundation for the extraction of the individual vowel tokens in the analyses of this thesis, based on the fundamentals of acoustic theory outlined in the previous section. The most important methodological decision for the analysis of the vowel tokens per speaker is the number of measurements per vowel, discussed in Subsection 4.4.1. In the standard sociolinguistic procedure, vowels are measured at a single point (Evanini 2009: 57), usually based on a visual analysis of the spectrogram in order to determine a rather stable portion of a vowel (steady state) and to exclude transitional effects of the linguistic environments (Di Paolo, Yaeger-Dror and Wassink 2011: 90; Ladefoged 2003: 104-105; Schützler 2011: 29). Recent research in acoustic phonetics experiments with multiple points of measurements determined either by a fixed amount of milliseconds (e.g. every ten milliseconds – default distance approach; Di Paolo, Yaeger-Dror and Wassink 2011: 91) or by steps in percentages (e.g. every ten per cent – proportional distance approach; Di Paolo, Yaeger-Dror and Wassink 2011: 92) of the total vowel duration, from vowel onset to offset (e.g. Gardner 2010; McDougall and Nolan 2007). In this fashion, the entire formant trajectories per vowel should be compared across speakers (cf. e.g. Fox and Jacewicz 2009; Fabricius 2007a,b). Some research even attempts to include more than two dimensions in such comparisons. While multiple points of measurements and inclusion of dimensions in addition to formant one and two certainly improve the amount of information obtained from measuring a vowel, they
hinder comparability to earlier studies (Evanini 2009: 59). A related problem is the quantification of the point of measurement which I will outline in Subsection 4.4.2.

Very generally, the determination of the stress level of a vowel (unstressed, primary, secondary, etc.) is usually based on estimating the fundamental frequency (F0) of a speaker’s voice most commonly via autocorrelation, which is displayed as pitch tracks in a spectrogram (Ladefoged 2003: 75). Despite its efficiency, the method is prone to errors, especially with speech signals including background noise such as those recorded in living rooms (cf. Thomas 2011: 37). The practicability of using autocorrelation with my speech samples will be discussed in Subsection 4.4.3. The spectrogram and formant tracking settings are discussed in the subsequent subsection (4.4.4), followed by the influence of linguistic context on the vowels under analysis (4.4.5). The former is based on the theoretical fundamentals outlined in the previous section, and the latter has been well established in research conducted from the 1960s to 1980s (e.g. Labov, Yaeger and Steiner 1972; Lehiste 1964, 1970; Lehiste and Peterson 1961; Summers 1987, 1988). The final subsection (4.4.6) summarizes the methodological decisions made with regard to measuring vowels for the present analyses.

### 4.4.1 Single-point versus Multiple-point Measurements

As mentioned earlier, the most common procedure in vowel analysis is to look for a steady state in formants one and two, usually in the middle of a stable portion (midpoint) of a vowel (e.g. Chen et al. 2009; Pierrehumbert et al. 2004; Steinlen 2002). Another general approach to arrive at a single formant reading for one vowel is to inspect the formant frequency contours and select a measurement point based on formant values that are most characteristic of a given vowel (point of inflection; e.g. Evanini 2009: 60; Labov 1972b: 15; Labov, Ash and Boberg 2006b: 36; Peterson and Barney 1952: 177). Di Paolo, Yaeger-Dror and Wassink (2011: 90) claim that such single-point measurements are an overly simplistic projection for monophthongal vowels, since perception studies have confirmed that listeners use more information than provided by one measurement point in order to identify the quality of a vowel (also cf. Labov, Ash and Boberg 2006b: 37). While the methodology employed in perception studies might arguably lead to an interpretation that vowels should be measured at multiple points, variable multiple-measurement approaches have been used in acoustic phonetic and more recently in sociophonetic studies.

Kendall and Thomas (2012: n.p.), for instance, suggest to use at least three points of measurement for a given vowel at regular intervals of, for instance, 35 milliseconds (ms): at the midpoint, –35 ms towards the onset and +35 ms towards the offset of each vowel (default distance approach; Di Paolo, Yaeger-Dror and Wassink 2011: 91; also cf. Hillenbrand et al. 1995: 3100; Hillenbrand, Clark and Nearay 2001: 750). The disadvantage of fixed distances is that they require the vowel to be at least of a certain

---

46 Diphthongal vowels are commonly measured at one single point in the nucleus and one single point in the glide (cf. Ladefoged 2003: 105).
duration, in this case 70 ms, which may be problematic with some lax vowels such as KIT in connected speech. As Hall-Lew (2009: 132-133) and Evanini and Huang (2013: 5) point out, they included all vowel tokens of at least 60 ms and 50 ms, respectively, as shorter vowels are known to have a schwa-like quality (/ə/; cf. Lindblom 1963: 1779; Subsection 4.4.3). While it is of course possible to adjust the default distance to, for instance, 30 ms and 25 ms, respectively, the idea of this multiple-points measurement approach is flawed when analyzing vowels of 150 ms to 200 ms in the same data set, as only the middle portion of 60 ms or 50 ms is considered (cf. Fox and Jacewicz 2009: 2605).

Among others, Adank, van Hout and Smits (2004: 1731), Fox and Jacewicz (2009: 2605), Gardner (2010: 64) and McDougall and Nolan (2007: 1826) use a proportional distance approach to measure vowels along their trajectories. Fox and Jacewicz consider five equidistant temporal locations at the “20%-35%-50%-65%-80% point of the vowel” (2009: 2605) based on its duration as providing enough information about (dialect-specific) formant trajectory changes that may remain “unnoticed while sampling the formants at only two or three points” (2009: 2605). Unlike Fox and Jacewicz, Adank, van Hout and Smits (2004: 1731) use nine equidistant temporal locations based on the vowel’s duration, which they most likely consider as sufficient information about the formant contours of the vowels in Dutch; Gardner (2010: 64) and McDougall and Nolan (2007: 1826) rely on ten such temporal locations at every 10% of the vowels’ durations.

This brief comparison suggests that the actual number of measurements at different temporal locations is not only disagreed upon, but also arbitrarily chosen. While three points of measurement may fail to capture changes in formant contours that five can, the latter may fail to capture the ones that nine could, which in turn may fail to capture a change that a tenth point of measurement could. Measuring vowels every 5% may provide even more formant contour changes that may or may not distinguish, for instance, dialect regions from one another. Arbitrarily selected multiple points may, in fact, capture changes in formant trajectories that are measurable, but not socially meaningful. They may, for instance, be so slight that listeners do not recognize them. Likewise, it is unclear whether a change in formant frequency of 30 Hz in formant one or 70 Hz in formant two at two different temporal locations is to be considered as a difference in dialect region or as intra-speaker variability. In fact, Adank, van Hout and Smits point out that for the majority of the vowels they analyzed for 160 speakers of different Dutch varieties, the single-(mid)point measurement at 50% of the vowel duration sufficed to separate the vowels from one another “based on their steady-state characteristics for their first two formants frequencies alone” (2004: 1737; also cf. 2007: 1132).

In addition, according to Di Paolo, Yaeger-Dror and Wassink (2011: 93), the disadvantage of the proportional distance approach is that the beginning and endpoints are taken arbitrarily: Adank, van Hout and Smits (2004: 1731) take the first of their nine formant readings “at the start of the vocalic portion of the token” and the last at the
end of the vocalic portion. In the analysis of the tokens, they only include the formant measurements between the third point (25%) and the seventh point (75%) in order to exclude the transitional influence of the consonantal context preceding and following the vowel (2004: 1731). That is, the transitional phases are arbitrarily restricted to the 0%-25% and the 75%-100% portion of the vowel, regardless of the consonant and vowel under investigation.\footnote{Differences in transitional phases caused by the different consonant-vowel combinations are well documented (cf. Ladefoged 2005; Ladefoged and Disner 2012).}

In the interval approach, measurements are taken at regular intervals across the vowel (e.g. every 20 ms). Since vowels vary in duration, a quantitative study of vowel formants will result in a different number of measurements for each vowel (Di Paolo, Yaeger-Dror and Wassink 2011: 93). Thus, comparison of formant trajectories is limited to the smallest number of measurements, taken from the shortest vowel in the data set. Assuming that the shortest vowel is 70 ms in duration, the earliest measurement point may be at 20 ms after the transitional phase of the formants from the preceding consonant, and the latest measurement may be 20 ms before the transitional phase of the formants to the following consonant. This would leave a measurable duration for that vowel of 30 ms, resulting in, for instance, three points of measurements at every 10 ms or less. The number of intervals is, however, much larger for vowels 200 ms in duration, and yet every point of measurement after the first three is unusable in a comparison of formant trajectories. In addition, the reasoning for choosing the length of the interval is once more unmotivated.

While this problem could be circumvented by using percentage steps instead of milliseconds, it remains for different reasons: Automatic token extraction from speech recordings with a large number of vowel tokens (e.g. more than 4000; Hillenbrand, Clark and Nearey 2001: 750) will result in measurement errors. While single-point measurements can relatively easily be repeated manually for each faulty measurement, ten-point measurements cannot. In cases where even manual re-measurements do not provide accurate formant readings (LPC analysis with large bandwidths of more than 400 Hz; Ladefoged 2003: 117), tokens have to be excluded from the analysis (cf. Subsection 4.5.1). Data cleaning may thus result in different numbers of measurement points per vowel, even if the point selection is based on percentages.

Another major problem of multiple-point measurements is connected speech. Most of the studies outlined above base their multiple-measurement points on vowels in citation form in word lists. In my data, word-list tokens are roughly 30% longer in duration than those in connected speech. Such reduced vowel duration is mostly caused by reducing the duration of the steady state of a vowel, but not the transitional phases to the neighboring segments (cf. Thomas 2011: 149). In connected speech, the vowel is thus not necessarily characterized by changes in formant trajectories, but by its target or steady state for which a single-point measurement may suffice (cf. Adank, van Hout and Smits 2004: 1737, 2007: 1132).
4.4. Measurements

Labov, Ash and Boberg (2006b: 37-38) employ such a single-point measurement (the maximal displacement approach) when selecting an F1/F2-paired measurement for each of their 134,000 vowels acoustically analyzed from 439 speakers (2006b: 119). Despite the crucial importance of other acoustic events such as F0, F3, nasal formants, tone and laryngeal tension for vowel quality (Di Paolo, Yaeger-Dror and Wassink 2011: 93; Labov, Ash and Boberg 2006b: 37), a host of research has confirmed the importance of the first two vowel formants, F1 and F2, in vowel production (e.g. Peterson and Barney 1952; Labov, Yaeger and Steiner 1972; Hillenbrand et al. 1995; Hagiwara 1997; Ladefoged 2003; Labov, Ash and Boberg 2006b; Ladefoged and Johnson 2011) so that I will rely solely on F1/F2 pairs of measurement for determining vowel quality (cf. Boberg 2005, 2008b, 2010; Labov, Ash and Boberg 2006b). According to Labov, Ash and Boberg, a plot of F1 against F2 sufficiently “illustrates salient social differences in the pronunciation of the vowels of North American English, including both vowel shifts and differences in phonemic inventory” (2006b: 37).

In large quantitative studies such as Adank, van Hout and Smits’ (2004, 2007), Boberg’s (2008b, 2010) and Labov, Ash and Boberg’s (2006b) a series of paired measurements “at every pitch period” (Labov, Ash and Boberg 2006b: 37) introduces additional practical issues to the discussion. An array of sequential measurements is easy to plot and read for a single vowel, but becomes illegible when plotted for 300 vowels. Moreover, plotting 300 formant trajectories for each of the 439 speakers included in the atlas obscures any pattern and precludes the goal of describing the vowel systems of North American Englishes (Labov, Ash and Boberg 2006b: 37). Additionally, the central concern of dialectological and sociolinguistic research are inter-speaker comparisons which are simply not feasible with trajectories, where the great difficulty lies in establishing precise points of comparison which makes quantitative analysis problematic (Labov, Ash and Boberg 2006b: 38).

For these reasons, Labov, Ash and Boberg (2006b: 38) measured vowels at a single point, namely the strongest point of tongue inflection, since for many vowels such a point can be clearly determined in one or both formants at a specific location in the nucleus. In this manner, pre- and post-nuclear transitional formant values are excluded from the measurements (also cf. Harrington 2010: 85). A point of inflection marks the moment when a speaker’s tongue ceases movement away from a pre-nuclear transition into the vocalic nucleus and commences movement away from the nucleus toward the position necessary for the following segment (or into a glide in the case of a diphthong; Labov, Ash and Boberg 2006b: 38). Among the advantage of this single-point measurement method, Labov, Ash and Boberg list that it is the best representation of the vowel’s overall quality, it provides an accurate account of the extent to which a speaker engages in a vowel shift, and listeners appear to be sensitive to such points of inflection (2006b: 38).
4.4.2 Quantifying the Maximal Displacement Approach

The identification of points of inflection has been described in detail by Labov, Ash and Boberg (2006b: 38), on which the following paragraphs are based. It crucially depends on an analysis of the central tendency or target of a vowel, i.e. the main trajectory of the tongue during the vowel’s articulation. For most short and many long upgliding vowels, the trajectory is a downward movement of the tongue into the nucleus of the vowel, followed by a rise of the tongue out of the nucleus. Acoustically, this tongue movement is expressed in a rise-and-fall pattern in first formant frequencies so that the maximum value in the first formant indicates the lowest point the tongue reaches. In fact, all consonants except pharyngeals lower F1, so that maximum F1 in a vowel is the point farthest from the consonants (Thomas 2011: 49). Consequently, vowels with a rise-and-fall tendency in the first formant should be measured at the maximum first formant value, including the corresponding second formant at the same temporal location. If the second formant was selected independently at a similar maximum value, the resulting F1/F2-pair would not describe the vowel quality perceived by a listener, but a non-existent one (cf. Labov, Ash and Boberg 2006b: 38), i.e. formant one is never independent of formant two and vice versa. While this observation seems trivial, it becomes crucial in the statistical modeling (collinearity; cf. Subsection 4.6.2.5): if formant one or two is entered as the dependent variable in the model, the corresponding formant can never be entered as an independent variable in the same model (but cf. Hoffmann 2011 for the violation of this rule).

Ingliding vowels provide a major exception to the principle of measuring the maximum F1 value, because the central tendency is a tongue movement toward and subsequently away from the front or periphery of the vowel space (Labov, Ash and Boberg 2006b: 38): Acoustically, the tongue movement toward and away from the front periphery is marked by a rise-and-fall in formant two, so that such vowels should be measured at their second formant maximum; tongue movement toward and away from the rear periphery is indicated by a fall-and-rise in formant two, so that such vowels should be measured at their second formant minimum.

Typically, ingliding vowels arise in two situations: first, in historically long and ingliding vowels, e.g. /æh/ (/æ:/) and /oh/ (THOUGHT), and originally short vowels that

---

48 Labov, Ash and Boberg (2006b: 12) characterize long or tense vowels in American English as possibly diphthongal, reflected in their binary notation. This notation reduces the redundancy of a unary notation by assigning one symbol to short vowels (denoting their nuclear quality) and two symbols to long vowels: The first one denotes the nuclear quality, which is identified as ‘the same’ as that of a corresponding short vowel (/i/ = /i/) and the second symbol denotes the quality of their glide. The three basic types of glides include: front upglides, represented as /y/; including all the varying end positions [j, i, e, c]; back upglides (/w/ for [w, u, o, y]); and inglides or long monophthongs (/h/ for lengthening and centralizing toward /o/). Front long upgliding vowels are thus: /iy/ (FLEECE), /ey/ (FACE), /oy/ (CHOICE), /ay/ (PRICE); back upgliding vowels are: /iw/ (suit), /iw/ (GOOSE), /ow/ (GOAT), /aw/ (MOUTH; Labov, Ash and Boberg 2006b: 12-13).

49 Ingiding vowels are: /ah/ (PALM), /oh/ (THOUGHT), /æh/ (/æ:/ plus inglides) and the vowel in the lexical set NURSE. Only for these vowels rounding is contrastive (Labov, Ash and Boberg 2006b: 12-13).
4.4. Measurements

have been raised and tensed along the peripheral track as part of a chain shift such as /æ/ in the Northern Cities Shift and /e/ and /i/ in the Southern Shift; and second, in high upgliding vowels followed by liquids (e.g. *fear*, *pool*), because the latter are articulated in mid-central position and consequently share acoustic characteristics with central inglides. Labov, Ash and Boberg (2006b: 38) also use the trajectory of formant two in cases where a more precise point of measurement within a steady-state in formant one is needed. This is especially necessary when a point of inflection in formant two seemingly marks the maximal distance from preceding and following consonantal transitions at the same time (midpoint measurements). Thus, the points of measurements between steady-state and non-steady-state vowels are not as precisely comparable between speakers as this single-point measurement method allows for when compared to multiple-point measurement methods (see previous subsection).

The main problem with this single measurement that Labov, Ash and Boberg (2006b: 38) identified is its failure to indicate the presence and quality of off-glides for monophthongal and diphthongal vowels. Whenever they have a contrastive function on a phonemic or geolinguistic level, their role becomes essential in the study of regional differences and sound changes so that their quality should be acoustically determined. Despite the importance of the glides, Labov, Ash and Boberg (2006b: 38) found “that an actual measurement of the glide target was not necessary” and that the absence/presence of glides and their quality could be effectively determined with a code in the annotators’ comments to the measurements (some glide measurements are, however, included; cf. Labov, Ash and Boberg 2006b: Chapter 13).

Since this thesis is primarily concerned with the lax vowels of St. John’s English (and thought), I adhere to the same policy of measuring the nuclear quality only, regardless of the vowel under investigation. The maximal displacement method is, however, difficult, if not impossible, to automatize. In an “overly simplistic projection” (Di Paolo, Yaeger-Dror and Wassink 2011: 90), Di Paolo, Yaeger-Dror and Wassink insinuate that when the midpoint method is used, the measurement point of a vowel coincides with the one derived from employing the maximal displacement method in most cases (2011: 92, Figure 8.2). This would suggest that in a quantitative analysis, dividing the vowel duration by two and have a script automatically take formant readings via LPC analysis at this point would result in findings no different from those yielded by manual measurements at the strongest point of tongue inflection.

In an elaborate analysis and discussion of the measurement methods used by Labov, Ash and Boberg (2006b), Evanini (2009) found a way to quantify their maximal displacement method. He analyzed the atlas and extracted a subset of 406 speakers, including 110,399 pairs of non-normalized manual F1 and F2 measurements. Based on this subset, Evanini (2009: 65) obtained formant one and two values automatically, according to five procedures: he read formants at 25%, 33% and 50% of the vowel’s duration, used a simplified version to differentiate ingliding from upgliding vowels and used Lennig’s Coefficient
4. Data and Methodology

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Algorithm for determining measurement point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid(point)</td>
<td>measure at one-half of vowel duration (t/2)</td>
</tr>
<tr>
<td>Third</td>
<td>measure at one-third of vowel duration (t/3)</td>
</tr>
<tr>
<td>Fourth</td>
<td>measure at one-fourth of vowel duration (t/4)</td>
</tr>
<tr>
<td>Lennig (1978)</td>
<td>measure at point where ( c_i ) is smallest</td>
</tr>
<tr>
<td>ANAE simplified</td>
<td>if vowel == /æ/, measure at F2 max, else if vowel == /ʊ/, measure at F2 min, else measure at F1 max</td>
</tr>
</tbody>
</table>

Table 4.7: Summary of automatic vowel analysis procedures under comparison. Table adapted from Evanini (2009: 64).

of Change (1978). The first three procedures, based on the vowels’ temporal domain, are relatively straightforward; the latter two, based on the vowels’ formant measurements, require some elaboration: In the fourth procedure (ANAE simplified), Evanini (2009: 64) treats all of the TRAP/BATH and THOUGHT tokens generically as ingliding vowels, so that the former were always measured at their F2 maximum and the latter at their F2 minimum. For short and upgliding vowels, maximum F1 is taken as an indicator of the vowel target. In the fifth procedure (Lennig 1978), a Coefficient of Change, \( c \), is calculated at each formant measurement, \( i \), which is based on the absolute value of the difference between that measurement and the preceding as well as the following one. The formula for calculating the coefficients is taken from Evanini (2009: 62):

\[
c_i = \frac{|F1_i - F1_{i-1}| + |F1_i - F1_{i+1}|}{F1_i} + \frac{|F2_i - F2_{i-1}| + |F2_i - F2_{i+1}|}{F2_i}.
\]

According to Evanini (2009: 62), this approach was developed in order to automatically detect a vowel’s steady state. The measurements are taken at the location in the formant trajectory where \( c_i \) is the smallest. This approach is simplistic insofar as it only accounts for the formant values in the immediate vicinity instead of a theoretically larger window, and as it treats the different ranges of variation between F1 and F2 similarly (Evanini 2009: 63).

Table 4.7 provides a summary of the procedures used by Evanini (2009: 64) in order to determine which of them reads formant frequencies most closely in value to the manual points of measurements chosen by the analysts of the atlas (Labov, Ash and Boberg 2006b). “The overall results compare the mean absolute differences between the manual measurements and the automatic measurements for each method” (Evanini 2009: 65). Additionally, the range of raw Hertz measurements is different for formants one and two, so that Evanini (2009: 65) also calculated the ratio of the measurement differences to the manual measurements made by the annotators of the atlas.
4.4. Measurements

<table>
<thead>
<tr>
<th>Procedure</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1 (Hz)</td>
<td>F2 (Hz)</td>
<td>F1 (%)</td>
</tr>
<tr>
<td>Third</td>
<td>64.7</td>
<td>216.1</td>
<td>10.4</td>
</tr>
<tr>
<td>Fourth</td>
<td>67.1</td>
<td>216.4</td>
<td>10.7</td>
</tr>
<tr>
<td>Mid(point)</td>
<td>70.0</td>
<td>232.6</td>
<td>11.3</td>
</tr>
<tr>
<td>ANAE simplified</td>
<td>103.8</td>
<td>247.8</td>
<td>18.5</td>
</tr>
<tr>
<td>Lennig (1978)</td>
<td>110.8</td>
<td>304.4</td>
<td>17.4</td>
</tr>
</tbody>
</table>

Table 4.8: Mean differences between manual and automatic formant measurements for five different measurement points (n = 110,399). The procedures are ordered according to the difference between the manual and automatic measurements. Table adapted from Evanini (2009: 65).

The results in Table 4.8 show that best automatic measurements were made at 25% and 33% of the vowel duration (Evanini 2009: 65) and that the measurements made automatically at the midpoint do not reflect optimal values for the vowels analyzed in the atlas (Evanini 2009: 70), unlike implied by Di Paolo, Yaeger-Dror and Wassink (2011: 92, Figure 8.2). In addition, the two automatic procedures that are based on the vowels’ formant trajectories perform substantially worse than the three that are based on the vowels’ temporal domains. However, the overall mean differences in Table 4.8 provide a somewhat misleading picture of the performance of each procedure as the distributions are skewed (Evanini 2009: 66). In case of the Third procedure, for instance, the majority of the differences for the measurements are less than the mean (for F1 67.6%, for F2 73%). As Figure 4.11 shows, the distribution is thus skewed to the left with a long tail of large differences which are, according to Evanini (2009: 66), most likely caused by formant tracking errors. Thus, the results do not suggest that automatic formant readings based on a vowel’s temporal domain perform better. For both formant measurements, F1 and F2, the largest number of automatic measurements are within 5% of the manual measurements. That is, if the automatic measurements yield an F1 for kit of 500 Hz and an F2 for the same vowel of 2000 Hz, the manual measurements will differ within the range of 487 Hz to 513 Hz for F1 and of 1950 Hz to 2050 Hz for F2 for the vast majority of measurements.

In addition to the comparison of automatic and manual selections of the points of measurements, Evanini (2009: 66) quantified the manual point of measurement selected in the atlas (Labov, Ash and Boberg 2006b), in order to support the relative success of measuring vowels automatically at one-third of their duration. The measurement point values he derived represent the ratio of the distance from the vowels’ onsets to the vowels’ durations, that is, the difference between the measurement point and the vowels’ onset to the time difference between the vowels’ offset and onset. A small ratio thus indicates a measurement point close to a vowel’s onset and a large ratio one close to a vowel’s
(a) The differences between automatic and manual F1 measurements.

(b) The differences between automatic and manual F2 measurements.

Figure 4.11: Histograms of the differences between automatic and manual formant one (a) and two (b) measurements using the Third method (n = 110,399). Figures taken from Evanini (2009: 67-68).

offset. Again, he finds that on average the manually selected points of measurements for vowels in the atlas are around one-third of the vowels’ durations (ranging from 0.25 to 0.39 per vowel; Evanini 2009: 69). None of the vowels have mean measurement points later than 0.40 (40%), providing further support for automatic midpoint-measurements being non-optimal in comparison to the manual point-of-inflection method.

In line with Evanini’s (2009) results, my script automatically selects formant readings based on the temporal domain of the vowels under analysis in percentages in order to maximize comparability to the results of Boberg (2005, 2008b, 2010) and Labov, Ash and Boberg (2006b). Due to the different results of the two single-point measurements (midpoint versus maximal displacement), the results of my study are not directly comparable to those of Hollett (2006, 2007), who also chose the midpoint method in her study of St. John’s lax vowels (cf. Subsection 4.1). I measure each vowel at 15 temporal locations: the first (zero) corresponds to the vowel’s onset, followed by measurement points at 10%-20%-25%-30%-33%-40%-50%-60%-66%-70%-75%-80%-90%-100%, with the last temporal location corresponding to the vowel’s offset. For the analysis of the vowels, only the 33% measurement point is included. The other temporal locations serve primarily as a basis of comparison for the formant readings for each vowel. The temporal locations close to the 33% point, 20%, 25%, 30% and 40%, serve as indicators for the accuracy of the formant and its bandwidth readings (cf. Subsection 4.5.1). The corresponding 60% - 80% measurement points determine the presence/absence and quality of off-glides, which are not included in the analyses (cf. Labov, Ash and Boberg 2006b: 38).
4.4. Measurements

4.4.3 Phrasal Accent

Vowel analysis should be based on vowels that are uttered in stressed, focused positions (Lehiste 1970: Chapter 4), as unstressed vowels are reduced in quality toward a central position in the vowel space, i.e. their quality corresponds much more to a schwa (/ə/) than that of their stressed counterparts, also referred to as undershoot (cf. de Jong 1995: 499; Harrington 2010: 91; Lindblom 1963: 1779; Summers 1987: 854). However, stress could neither be automatically identified in the 1970s (Lehiste 1970: 110) nor thirty years later as there is still “no known algorithm” to automatically measure stress acoustically (Ladefoged 2003: 94). What seems to be an acoustic commonplace is that in instrumental terms, stress is a combination of pitch, duration and loudness (Ladefoged 2003: 90; Lehiste 1970: 2), the first two playing the most crucial role.

The acoustic correlate of loudness is a sound’s acoustic energy or intensity in decibels (dB), which depends on the extent of the variation in air pressure (Ladefoged 2003: 90). One decibel corresponds to the smallest change in loudness that can be heard; 120 dB is the loudest sound humans can hear. Intensity is least indicative of stressed words or syllables: As Ladefoged demonstrates, the intensity of vowels produced with open lips (e.g. in three) is always higher than that of vowels produced with less open lips (e.g. in four). Even among monosyllabic lexical items with the same vowel differentiated only by contrastive stress (e.g. in I saw three bees and not four versus I saw three bees but no wasps), the intensity remained the same, regardless of whether three or bees was stressed in different sentences. The marker of contrast between those two lexical items was pitch, not intensity (2003: 92).

Like Labov, Ash and Boberg (2006b: 39), I determined sentence stress impressionistically. This decision is based on the lack of procedures to determine stress automatically and on the comparability to Labov, Ash and Boberg’s atlas results. In addition, I used pitch values and vowel duration as indicators of phrasal-accented lexical items, which I will outline in the subsections to follow.

4.4.3.1 Pitch

While pitch is an auditory feature, its acoustic correlate is a sound’s fundamental frequency (F0) – the rate at which vocal fold pulses recur – in Hertz (Ladefoged 2003: 75). Although F0 is one of the important markers of stress, the assumption that it is the most important one is wrong, since it is possible to emphasize words without using an increase in pitch (Ladefoged 2003: 93).

Pitch is most commonly determined via autocorrelation, which compares one cycle of a waveform of the individual pulses of the vocal folds with another one within a predetermined window. There are different window types available, e.g. Hamming, Hanning, Gaussian, etc., which do not differ substantially from one another in terms of efficiency. What is important about such windows is that they de-emphasize the waveform near
4. Data and Methodology

Figure 4.12: An expanded view of part of the waveform of the lexical item *something* uttered by male speaker 07HPOM. The two rectangular shapes show Hamming windows enclosing parts of the wave, the dashed rectangle shows a square window. Figure based on Ladefoged (2003: 78).

The edges in comparison to the part in the middle of the window (cf. Figure 4.10 in Subsection 4.3.3), unlike a square window which does not account for an abrupt change in minimum or maximum amplitude to zero. The window length is the duration of the piece of the waveform used in calculating the frequency, which must be long enough to include at least two cycles of the waveform (Ladefoged 2003: 77).

Figure 4.12 shows a waveform with six cycles in one Hamming window, plotted from speaker 07HPOM. In order for the software, which is Praat 5.3 (Boersma and Weenink 2011) in the present thesis, to calculate the length of one cycle, it first compares one part of the wave within the window with another part within the same window, and then determines that they have similar shapes. The software has to find a complete cycle before it can determine how long that cycle is (Ladefoged 2003: 78). In the example outlined in Figure 4.12, one cycle is six milliseconds (ms) in length, which means that the window length of 40 ms suffices. In case of very low pitches such as 45 Hz, the glottal pulses would be 22 ms apart (see below), so that a 40 ms Hamming window is not long enough to include at least two cycles, and the automatic pitch tracking algorithm would consequently yield false (or no) values.

The step size indicates the amount of time that the window is moved to the right before the pitch is calculated again. A value of ten ms is usually sufficient according to most descriptions of tone and intonation (Ladefoged 2003: 78). In Praat, the window
4.4. Measurements

length and the steps between two windows are set via the pitch floor (Boersma 2011b: n.p.). The pitch floor constitutes the lower end of the calculation range for the pitch, and the pitch ceiling constitutes the upper end. By default, the calculation range of the pitch for the autocorrelation algorithm and the display range of the pitch in the spectrogram are identical. The former requires a setting of 75 to 300 Hz to be appropriate for many male voices, and a setting of 100 to 500 Hz for many female voices (Boersma 2011b: n.p.; Ladefoged 2003: 81). Fundamental frequencies beyond the calculation rate are ignored (Boersma 2011b: n.p.) so that the resulting pitch readings are inaccurate.

At a pitch floor setting of 75 Hz, which I set for most of my male participants, Praat 5.3 will use a window length of $\frac{3}{75} = 0.04$ seconds (40 ms). The numerator, three, refers to the minimal number of cycles to be included in the analysis window (Boersma 2011b: n.p.). Likewise, a pitch floor setting of 100 Hz causes the analysis window to be 30 ms in length. The step size is calculated via the formula $\frac{0.75}{\text{pitch floor}}$, which yields a step size of 10 ms at 75 Hz and of approximately 8 ms at 100 Hz (Boersma 2011b: n.p.).

An autocorrelation algorithm calculates correlations between windowed waveform chunks set apart by the step size over a range of possible period lengths and reports the length that produced the highest correlation, i.e. estimates of F0 as a function of time (Johnson 2003: 30). The lag duration resulting in the highest correlation of the second, lagged signal and the first, unlagged signal is taken as the duration of one period (cf. Johnson 2003: 30; Thomas 2011: 37). If $T$ is the duration of one period in seconds, then the fundamental frequency in Hertz is equal to $\frac{1}{T}$ (Johnson 2003: 30) or $\frac{1000}{T_{ms}}$ (Ladefoged 2003: 81; Thomas 2011: 37).

If the autocorrelation algorithm identifies a pitch of, e.g., 100 Hz in one window and of 150 Hz in the next window, it falsly interpreted what constitutes a vocal fold pulse (Ladefoged 2003: 79). Two types of errors are common with autocorrelation: pitch-halving and pitch-doubling. The former occurs when the algorithm identifies two pitch periods and mistakes them for a single period; the latter occurs when autocorrelation mistakes half a pitch period for a whole one (cf. Johnson 2003: 31; Thomas 2011: 37). The reasons for these mistakes are quite obvious: first, the calculation range is set incorrectly, second, the recordings contain background noises, and third, the phonation type is non-modal (breathy, vocal fry or diplophonia; also whispering), which introduces aperiodic and/or irregular pitch periods (cf. Johnson 2003: 31; Ladefoged 2003: 80; Thomas 2011: 37).

In order to verify the automatically calculated pitch values, autocorrelation can be manually imitated by visually inspecting the waveform in the oscillogram, zooming in on one pitch period, marking it to extract the exact length and dividing 1000 ms by the duration of the pitch period (Ladefoged 2003: 81). Alternatively, if sudden jumps and falls occur in the pitch track, it is likely that a speaker breaks into or out of creaky voicing. If these are not accompanied by a change in spacing between glottal pulses, the pitch track is inaccurate (Thomas 2011: 38). In the example shown in Figure 4.12, the
interval between Hamming windows one and two – the “lag” – is not exactly the duration of one cycle. As mentioned earlier, the duration is six ms and the step size is set to ten ms. The automatically determined F0 at a pitch floor setting of 75 Hz is 166.6 Hz during articulation of the stressed STRUT vowel in the lexical item *something*. That is, the autocorrelation algorithm identified the lag duration between the two Hamming windows to be \( \frac{1000}{6} = 6 \) ms. Manual determination of one pitch period cycle within that STRUT vowel yields 6.1 ms, i.e. a pitch of \( \frac{1000}{6.1} = 163.9 \) Hz. The pitch in the lexical item *something* uttered by speaker 07HPOM, determined automatically via autocorrelation, is correct at a pitch-floor setting of 75 Hz and a pitch-ceiling setting of 300 Hz. If individual vowels are realized in creaky voice or if this phonation type is a general characteristic of a speaker, the pitch floor should be decreased in value, e.g. to 50 Hz, in order to derive accurate pitch readings (Ladefoged 2003: 81). However, at this setting, most of the automatic pitch readings in Praat were inaccurate.

According to Ladefoged (2003: 81), most pitch-tracking systems make errors when the calculation range is unnecessarily large. Although Praat uses one of the most accurate – unbiased – autocorrelation algorithms (based on Boersma 1993; cf. Boersma 2013: 383), the automatic pitch tracker often yielded erroneous or no pitch values in the reading passage and connected speech, regardless of the calculation-range settings. If so, I manually calculated the pitch for each speaker in connected speech at portions where they used a rather low and a rather high pitch in order to find the accurate per-speaker calculation range values. I measured the pitch at 33% of the vowels’ durations, in order to arrive at precise points of comparison. I then compared the pitch value with the preceding and following lexical item that was impressionistically unstressed, in order to determine possible differences. In the word list, where all words should be stressed, I extracted the pitch automatically for each speaker at 33% of the vowels’ durations, excluding those lexical items in which speakers realized the vowels of interest in creaky voice. I calculated an average pitch for the remaining items per speaker in order to get a reference value. I used this value to estimate whether a stressed lexical item was sentence-stressed or emphatically stressed. The latter has to be treated separately from the word or sentence stress (Lehiste 1970: 38), especially when the stressed item is not a function word. In the example shown in Figure 4.12, the pitch of approximately 165 Hz indicated emphatic stress of the stressed vowel in *something*, as the average pitch of this old, male speaker was 103 Hz, and the average pitch of impressionistically sentence-stressed vowels was between 120 Hz and 140 Hz (cf. Lehiste 1970: 82).

Additionally, pitch measurements at 33% of the vowels’ durations disregard pitch differences due to the segments of speech. For instance, the vocal folds initially vibrate at a higher rate when a voiceless aspirated stop is realized as opposed to a voiced unaspirated stop. In the former, a higher pitch is due to the high rate of airflow, but not due to the sentence-stress level of the lexical item that contains the vowel. Likewise, when realizing a voiced labiodental fricative in intervocalic position, the consonant’s pitch is
4.4. Measurements

much lower than that of the preceding stressed vowel due to a drop in airflow (micro-
prosody; Ladefoged 2003: 87).

4.4.3.2 Vowel Duration

As mentioned above, duration of a vowel or a series of sounds is one of the phonetic
manifestations of stress. In English, stressed vowels are on average 50% longer than
the average unstressed vowel (Lehiste 1970: 36). Vowels in unstressed position tend to
be undershot; their quality usually centralizes and resembles that of /\a/ (cf. de Jong
Especially in studies on vowel shifts such as the present one, it is important to rely on a
sound methodological background in order to be able to attribute a downward shift of a
lax vowel on the non-peripheral track to the phonological system and not to undershoot.
At the same time, vowels must not be too long, especially in function words, as that is
indicative of emphatically stressed vowel tokens.

Duration also interacts with coarticulation in that the extent of the latter is pre-
dictable from the durational decrease (Harrington 2010: 93-95). Phonetically long vowels
are only affected by coarticulation near the edges; short vowels show coarticularatory ef-
fects throughout the vowel (Thomas 2011: 50). That is, when vowels are shortened due
to several different factors, the steady state of the vowels is what suffers most. Such
factors can, for instance, be an increase in the rate of speech, weaker stress or a following
voiceless segment. Vowels so affected may loose their steady state altogether, so that
the transitional phases from preceding to following segment may be all that is left in
connected speech (Thomas 2011: 149).

In addition, an increased speech rate generally shortens vowel duration (Harrington
and Cassidy 1999: 71; Thomas 2011: 144; also cf. Watt 2013). However, the effect of
speech rate is dominant for the steady state of a vowel, as outlined above, and shortens
unstressed syllables rather than stressed ones (Lehiste 1970: 40). In order to examine
contrastive or contextual vowel length, Wassink suggests to normalize segmental duration
via z-scores: She first calculates the mean duration for each vowel class, e.g. KIT, DRESS,
TRAP, LOT, THOUGHT and STRUT. Second, she determines a grand mean of duration,
$\bar{D}_{o,k}$, for speaker $k$ across all vowel-class durations via the formula:

$$
\bar{D}_{o,k} = \frac{\sum_{j=1}^{n} (\sum_{i=1}^{n} \frac{D_{ijk}}{n_j})}{n},
$$

where $D_{ijk}$ is the observed vowel duration “for token $i$ of vowel $j$ for speaker $k$” (2006:
2345). Third, she normalizes the vowel durations via the normalized value of duration,
$\delta_{ijk}$, which is equal to the observed vowel duration minus the vowel class’ grand mean
(Wassink 2006: 2345):

$$
\delta_{ijk} = D_{ijk} - \bar{D}_{o,k}.
$$
For this normalization of per-speaker vowel durations, the individual observations need to be distributed normally and outliers need to be absent, since the mean is taken to be the representative of the duration of a vowel class (cf. Subsection 4.6.2.1).

This approach is, however, unnecessary in my analyses for the following reasons: 1) I do not include vowel duration as a factor in the analysis in line with most, if not all, other studies on the Canadian Shift (e.g. Boberg 2005, 2008b, 2010; Clarke, Elms and Youssef 1995; D’Arcy 2005; Hollett 2006, 2007; Labov, Ash and Boberg 2006b; but see below). 2) An increase in speech rate is achieved largely at the expense of unstressed syllables (Lehiste 1970: 40), which are generally excluded from my analyses. 3) Instead of normalizing durations with non-normally distributed data, I calculated per-speaker rates of speech per portion of connected speech in syllables per second (cf. Watt 2013) in order to determine a reference value during the segmentation process. Due to effects such as undershoot and coarticulation, Evanini and Huang (2013: 5) and Hall-Lew (2009: 132-133) suggest to exclude all vowel tokens shorter than 50 ms and 60 ms, respectively, from the analysis. Visual inspection of spectrograms based on my data suggested that this vowel length may be too short to exclude non-sentence-stressed vowel tokens. I thus set the minimal threshold to 70 ms, which yielded negligibly fewer tokens for speakers with a higher rate of speech. At the same time, I discarded tokens of vowel lengths longer than 250 ms in order to exclude emphatically stressed tokens. It may be due to these thresholds that the remaining vowel durations are not distributed normally per speaker.

Although vowel length or phone duration is not included in the regression analyses of this thesis, it is indirectly accounted for by including number of following syllables as a predictor variable (cf. Labov, Ash and Boberg 2006b: 39). According to Lehiste (1970: 41), listeners interpret the duration of a particular sound by relating it to the duration of the lexical item it occurs in as a whole, because vowel duration remains contrastive (phonemically long versus phonemically short vowels). In multisyllabic words, the longer the lexical item that contains long vowels is, the shorter these vowels are realized (Lehiste 1970: 40).

### 4.4.4 Tracking Formants Automatically

A spectrogram consists of several Fast Fourier Transform (FFT) spectral slices of a certain window length at various time steps, which are coded, turned and added together.

As shown in Figure 4.13, the values of the amplitudes are translated into different levels of gray (1). The darker the gray spots in the spectrogram, the higher the amplitude of the transformed signal. The higher amplitudes of formants are amplified, so that equally dark areas do not mean the formants have the same amplitude (Ladefoged 2003: 109). Subsequently, the spectrum is turned by 90° counterclockwise (2), so that the y-axis constitutes the frequency of the spectrum, and thus the x-axis is unused (3).
4.4. Measurements

Abbildung 3.19: Vom Spektrum zum Spektrogramm I: Amplitudenwerte im Amplitudenspektrum werden in Graustufen kodiert (1); das Spektrum wird um 90 Grad gedreht (2); die Frequenz wird auf der y–Achse abgetragen, die x–Achse ist ungenutzt (3).

Figure 4.13: FFT to spectrogram I. Figure taken from Mayer (2010: 98).

Abbildung 3.20: Vom Spektrum zum Spektrogramm II. Oben: 4 aufeinanderfolgende Spektren am Anfang, nach dem ersten Drittel, nach dem zweiten Drittel und am Ende des Diphtongs erstellt; in Graustufen kodiert und um 90 Grad gedreht. Unten: Aneinanderreihung der 4 schematischen Graustufenspektren auf der Zeitachse (links) und der entsprechende Abschnitt markiert in einem Breitbandspektrogramm (Gesamtäußerung: Hast du einen Moment Zeit?).

Figure 4.14: FFT to spectrogram II. Figure taken from Mayer (2010: 99).
In the upper panel of Figure 4.14 are four spectra taken of the diphthong PRICE at the beginning, after one third, after two thirds and at the end of the diphthong. The turned spectra are consequently added, so that the formerly unused $x$-axis can now represent time in seconds. The diphthong is part of a longer utterance shown in the wideband spectrogram at the bottom right. The levels of gray are added together to form a spectrogram, like slices form a loaf of bread. The spectrogram settings and their implications are outlined in the following subsection (4.4.4.1). A correctly specified spectrogram aids in determining the automatic formant tracking settings, outlined in Subsection 4.4.4.2.

### 4.4.4.1 Spectrogram Settings

Spectrograms visualize the sounds of a language to a certain degree in that they (color-)code the frequencies of the acoustic energy of those sounds and in that they indicate the most important acoustic properties of vowels – the formants of vowels (Ladefoged 2003: 104). The first formant inversely indicates the height and the second formant indicates the advancement of vowels (cf. e.g. Boberg 2010: 143; Hillenbrand et al. 1995; Labov, Ash and Boberg 2006b: 37; Peterson and Barney 1952; also cf. Subsection 2.3.2). In order to use the visualization provided by spectrograms properly, they need to be set correctly.

The most important differentiation is made between wideband and narrowband spectrograms. The former usually has bandwidths of 260 Hz or more and shows the individual formants better than a narrowband spectrogram. The latter usually has bandwidths of 100 Hz or less and reveals the individual harmonics within each formant (Ladefoged 2003: 106), which provide very good information on the pitch (Ladefoged 2003: 107). Figure 4.15 shows a wideband and a narrowband spectrogram in comparison. The dark bands between the vertical crosshairs between 57.53 s and 57.63 s in Figure 4.15a are the formants of the DRESS vowel in the lexical item *veterinary*. The wideband spectrogram shows the precise time of the occurrence of each vocal fold vibration (vertical lines; Ladefoged 2003: 111). The more regular these vocal fold vibrations appear, the more likely is an [omphonic type](Ladefoged 2003: 113). In Figure 4.15b, the spectral analysis is more precise in the frequencies it shows, separating out the individual harmonics (cf. Subsection 4.4.4.2; also cf. Ladefoged 2003: 109). All of the individual harmonics rise between the two vertical crosshairs for the speaker. Although the fundamental frequency cannot be determined exactly, the tenth harmonic rises from roughly 1800 Hz to approximately 2000 Hz. F0 must thus have risen by a tenth of this, from 180 Hz to 200 Hz (cf. Ladefoged 2003: 108). In fact, the automatically determined pitch at the onset of the vowel is 185 Hz, and that at the offset is 210 Hz. The token is thus likely to be emphatically stressed and not included in the analyses in this thesis.

Generally, the best spectrograms for looking at formants have a bandwidth just high enough not to show the individual harmonics. Ladefoged (2003: 108) suggests a bandwidth of 200 Hz for male voices and 300 Hz for female voices. Although Labov, Ash
4.4. Measurements

(a) A wideband spectrogram with a bandwidth of 300 Hz.

(b) A narrowband spectrogram with a bandwidth of 33 Hz.

Figure 4.15: Spectrograms at two bandwidths of the lexical item veterinary as uttered by male speaker 07HPOM in his reading passage. The vertical crosshairs mark the stressed DRESS vowel, which is 105 ms in length.
and Boberg (2006b: 38) use spectrograms with a bandwidth of 500 Hz, I determined the bandwidth settings individually for each speaker, roughly equaling bandwidths of 300 Hz for male speakers and 400 Hz for female speakers. I based the decision on the optimal time resolution of the spectrograms in order to identify the formants more accurately. In Praat 5.3, the bandwidths of the spectrograms are determined via the window length. The exact calculation of the bandwidth depends on the window type, e.g. Hamming, Hanning, Gaussian, etc. (Boersma 2011c: n.p.), and is by default set to Gaussian. For the Gaussian window, which I used, the calculation of a –3 dB bandwidth is done via the equation $\frac{1.2982804}{\text{window length}}$ (Boersma 2011c: n.p.). A bandwidth of 300 Hz thus requires a window-length setting of $\frac{1.2982804}{300} = 0.004$ s (4 ms).

Other spectrogram settings include the frequency range and the dynamic range. The frequency range is limited by the Nyquist frequency, which in turn is determined by dividing the sampling frequency by 2 (cf. Subsection 4.4.4.2 below). If the frequency range extends over the Nyquist frequency, the spectrogram will show a white background only (Boersma 2011c: n.p.). The dynamic range adjusts the range of contrasts that are shown in the spectrogram. The two examples shown in Figure 4.15 have been made with a dynamic range of 45 dB. The default value, 50 dB, usually yields a very dark background in which the formants are less straightforwardly identifiable (cf. Ladefoged 2003: 109).

The dark bands in wideband spectrograms do not indicate similar amounts of acoustic energy. Spectrograms give boost to the higher frequencies, so that waveforms of equal amplitudes with frequencies of 50 Hz to 800 Hz will be hardly visible in a spectrogram as compared to those with frequencies of, for instance, 1000 Hz to 5000 Hz (Ladefoged 2003: 110-111).

The correct spectrogram settings are important in order to determine where the formants of a vowel are. Very generally, there is a formant at every 1000 Hz, but there are exceptions to this rule. For low back vowels (LOT, THOUGHT), there are two formants below 1000 Hz, which may show up as one wide dark band in the spectrogram (Ladefoged 2003: 114). Since the third formant of such vowels is usually between 2000 Hz and 3000 Hz, the dark band below 1000 Hz most likely indicates presence of two formants. However, such a wide band in a spectrogram may also indicate acoustic energy above or below the first formant that is not a quality of the vowel (i.e. an additional, spurious formant) but a general characteristic of a speaker’s voice, especially when such energy portions around a formant occur irrespective of the vowel (cf. Subsection 4.3.3).

4.4.4.2 Formant Tracker Settings

Figure 4.16 shows the sound editor of Praat 5.3 (Boersma and Weenink 2011). The sound sample presented here is the lexical item veterinary uttered in the reading passage of speaker 11DCMF – a middle-aged urban female – between 24.84 and 25.44 seconds. The upper panel in the figure is the graphic representation – the oscillogram – of the sound wave that was recorded when veterinary was uttered by that speaker. The numbers on
the left of that panel take the amplitude of the sound wave into account and range from -0.3 to 0.24 Pascal (cf. Boersma 2011a: n.p.). Of more importance here are the other two panels: The first one is the visible wideband spectrogram of the sound wave, which ranges from a frequency of 0 Hz at the bottom to 7000 Hz at the top of the panel, as the numbers on the left indicate. The vertical crosshairs help to show the whereabouts of the segmentation of *veterinary* into its individual sounds and are extensions of the segments in the lowest panel – the textgrid. The second vertical crosshair (t) from the left marks the point of measurement of formants one and two at 33% of the vowel’s duration (at 24.98 s). The textgrid consists of seven interval tiers, namely sentence, phone, phoneme, word, word type and part of speech. All of these interval tiers should be self-explanatory. The latter two tiers are of negligible importance for the present thesis.

Interval tier means that each segment must consist of two boundaries: the vowel dress on the interval tiers phone and phoneme is marked by two boundaries. The opposite of interval tiers are point tiers, in which, for instance, one boundary anywhere within the duration of the vowel dress could mark a possible spot for taking formant measurements. The preceding segments on the phone and phoneme tiers indicate that the mid front vowel is preceded by the voiced labiodental fricative /v/. Unlike the following segments on the two tiers, the preceding sound’s phonetic realization is very close to its phonemic description. The following sound is realized as a voiced alveolar tap/flap (/t/), or arguably a voiced alveolar plosive (/d/), although phonemically it should be realized as a voiceless alveolar plosive (/t/). Tapping voiceless alveolar plosives in non-final position is, however, the norm in most North American varieties, including middle-class speech in St. John’s, Newfoundland.

As mentioned earlier, the panel in the middle shows the spectrogram of the sound wave. I set the the range to 7000 Hz in order to see the frequencies above 5000 Hz which may show formants four and five for high-pitched female voices and the spectral behavior of fricatives, which are usually differentiated from one another (e.g. voiced alveolar fricatives from unvoiced ones) at frequencies around 9000 Hz. The maximum frequency up to which Praat can calculate a spectrogram is determined by the sampling frequency. Since the recordings were made at a sampling frequency of 48,000 Hz, Praat could calculate a spectrogram from 0 to 24,000 Hz (Nyquist frequency). However, the recordings were generally re-sampled to 22,000 Hz for the analyses, yielding a Nyquist frequency of 11,000 Hz (cf. Subsection 4.3.3).

The pitch value for the female speaker 11CDMF is indicated on the right of the panel. I set the calculation range in this particular example to 100 Hz (pitch floor) and 400 Hz (pitch ceiling; cf. Subsection 4.4.3.1), so that the automatic pitch track accurately determines the pitch of 223 Hz in this case (on average approximately 200 Hz). This is a very high pitch which is usually expected with children (Di Paolo, Yaeger-Dror and Wassink 2011: 94) and indicates that the vocal tract – the filter – is rather short. For automatically tracking the pitch over a longer sound file, it is important to set the limits of
4. Data and Methodology

the calculation range apart accurately (cf. Subsection 4.4.3.1), so that all the variations in pitch (e.g. non-/emphatic stress, etc.) can be measured (unfortunately not very reliably in connected speech) by the computer. In this example, I measured the pitch values manually as outlined in Subsection 4.4.3.1 and set the limits for the figure to 165 Hz at the bottom and 310 Hz at the top for the purpose of visualization. The thin line in the spectrogram shows the movement of pitch within the word veterinary over time. The steep rising pitch curve towards the end of the vowel dress is an indicator of (sentence) stress. The dark dots are the automatically calculated formant tracks via LPC, based on the formant settings in Praat 5.3.

In Figure 4.16, the these settings are, first, 5500 Hz as the maximum formant frequency (default setting; in the LPC spectral analysis in Praat, this frequency is automatically determined by the Nyquist frequency), i.e. the frequency range in which Praat is supposed to look for formants. This value is derived from the predictions about adult female vocal tract lengths and usually serves as a starting point to check for the correct maximum frequency for each speaker and across all the vowels under analysis. Importantly, once the best settings are found they should be used consistently, so that the formant readings per vowel are comparable, since the physiological configuration of the vocal tract and the articulators do not change for one speaker in the course of an interview.

The predictions provide a rough guideline which suggests to look for formants of adult male speakers in each 1000 Hz - 1100 Hz band and of adult female speakers in each 1100 Hz - 1200 Hz band (cf. Subsection 4.3.3; Ladefoged 2003: 125). This does, however, not mean that the maximum formant frequency can be set to 2200 Hz for a female speaker in this analysis of vowel formant values, since the analysis is interested in a two-dimensional F1xF2 plot (cf. Section 5.1). As will be shown below, the automatic formant determination (LPC analysis) works a lot more reliably and accurately if more formants are asked for than are actually needed, especially when the automatic method misses formant tracks visible in the spectrogram. In the case of speaker 11CDMF, the second setting – number of formants – is set to five, as \( \frac{5500}{5} = 1100 \) Hz (this setting has the same effect as the number of coefficients in an LPC analysis).

Praat 5.3 allows three additional settings for automatic formant tracking, namely window length in seconds, dynamic range in dB and dot size of the formant tracks in the spectrogram. The first one determines how much time of the signal (the sound wave) is used for the formant calculations at each step. Its default value is 0.025 s. The value also affects the time steps at which each dot in the formant track is calculated. I use Praat’s default setting in which the time steps are set to 0.00625 s. They do not affect the quality of analysis, but assign more dots to the formant track if the value is increased. The dynamic range, the second additional setting, is by default at 30 dB, which I do not alter. If it is changed to a slightly lesser value of, for instance, 20 dB a formant track may not be shown (this did not affect detection of formant five as outlined below); if it is set to a slightly higher value, the formant tracks may become erratic, as there are too many
4.4. Measurements

Sarah Perry was a veterinary nurse.

Figure 4.16: Sound editor Praat 5.3 for the word *veterinary* uttered by female speaker 11CDMF at a setting of 5500 Hz and 5 formants. The spectrogram shows that the formant settings hinder accurate automatic formant tracking of the vowel DRESS /ɛ/.
peaks in the spectra for Praat to determine the formant contour. The last additional setting is a matter of taste. I personally prefer 0.5 as a dot size in order to see both the automatically tracked formant and the underlying formant in the spectrogram.

The wideband spectrogram (400 Hz) in Figure 4.16 shows that near the horizontal crosshair at 5250 Hz, there could be an untracked fifth formant very low in amplitude and high in bandwidth, as its gray area in the spectrogram is slightly darker than that above or below the formant. As mentioned earlier, spectrograms boost higher frequencies, so that equally dark areas for formants in the spectrogram do not mean the formants have the same energy (Ladefoged 2003: 109). The suspicion of a fifth formant being present in the signal is also supported by the high-pitched voice. In other words, the maximum formant frequency of 5500 Hz is too low for this speaker. Although only the first two formants are needed and four have been tracked, it is clearly noticeable that the automatic track of the fourth formant deviates from formant four in the spectrogram, roughly in the middle of the vowel. The rough guideline mentioned above suggests to take 6000 Hz to 8000 Hz as the starting point for the setting of the maximum formant frequency, since the pitch of female speaker 11CDMF is as high as that of children.

A setting at 6000 Hz and five formants did not force the automatic formant tracker in Praat to recognize the fifth formant, and the tracked formants two to four deviated considerably from the same three formants visible in the spectrogram. Changing the maximum frequency and number of formants stepwise to a higher setting resulted in the detection of the fifth formant at 6600 Hz and six formants, but also in the detection of an additional formant track between formant one and two, as can be seen in Figure 4.17a. However, since there is no acoustic energy visible in the spectrogram that would justify detection of this extra formant track, since there are no transitions into the preceding and following consonants (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 94), and since its value does not make sense for a dress vowel (see below), it can be considered as an artifact of the settings (a spurious formant; cf. Subsection 4.3.3).

Similarly to the extra formants in this speaker’s dress vowels, formants one to three were erratic in her kit and trap vowels at the setting of 6600 Hz and six formants. In addition to the detection of formant five, the contour of formant four was tracked more accurately than the one visible in Figure 4.16. The continuous stepwise increase of maximum formant frequency and number of formants arrived at the level of 8500 Hz and seven formants until reliable readings for this speaker were returned (cf. Figure 4.17b). At this setting, the extra formants disappeared almost entirely, although tracking for formants four and five in the transitional phase at the vowel’s onset matched the spectrogram’s formants more accurately at the setting of 6000 Hz and six formants (cf. Figure 4.17a).

As this procedure indicates, finding formants automatically in Praat is difficult at a lower setting of maximum formant frequency and number of formants. I followed this procedure for every speaker (cf. Evanini 2009: 80). The issues outlined above were also
4.4. Measurements

Figure 4.17: Sound editor Praat 5.3 for the word *veterinary* uttered by female speaker 11CDMF at a setting of 6600 Hz and 6 formants (a) and 8500 Hz and 7 formants (b). The spectrogram shows that the adjusted formant settings produce better results, but only the latter setting allows for accurate automatic formant tracking of the vowel DRESS.

The formant measurements in this particular example are summarized in Table 4.9. These values are raw formant values that cannot directly be compared to the ANAE (Labov, Ash and Boberg 2006b) normalized values outlined in Chapter 5. The differences in the formant readings are quite large, compared to those of other DRESS tokens of that speaker in her reading passage: $\Delta F1 \approx 10$ Hz, $\Delta F2 \approx 50$ Hz and $\Delta F3 \approx 100$ Hz between the three measurements, when settings (1) to (2) and (2) to (3) are compared. The high bandwidth for formant three in all three settings indicates that these measurements are not straightforwardly accurate. For the present analysis, however, formant three is pertinent to, for instance, male speakers with low pitches (e.g. of 85 Hz with settings at 5300 Hz and 5 formants).
4. Data and Methodology

Table 4.9: Raw formant (F1 to F3) and bandwidth (B1 to B3) values in Hertz for DRESS in veterinary in three different settings (1 to 3) as uttered by female speaker 11CDMF in her reading passage.

<table>
<thead>
<tr>
<th></th>
<th>F1</th>
<th>B1</th>
<th>F1</th>
<th>B1</th>
<th>F1</th>
<th>B1</th>
<th>F1</th>
<th>B1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>716.8</td>
<td>66.7</td>
<td>707.2</td>
<td>68.3</td>
<td>718.8</td>
<td>68.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1894.5</td>
<td>275.0</td>
<td>1954.9</td>
<td>262.7</td>
<td>1901.8</td>
<td>246.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2677.1</td>
<td>870.5</td>
<td>2886.2</td>
<td>543.26</td>
<td>2743.3</td>
<td>697.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Sett. (1): 5500 Hz/5 Formants | Sett. (2): 6600 Hz/6 Formants | Sett. (3): 8500 Hz/7 Formants |

not needed, so that the measurement pairs of formants one and two do not have to be discarded/re-measured.

The values for formants two and three at setting (2) are the manually corrected values. The automatically derived ones return F2 = 1370.1 Hz (B2 = 1534.9 Hz), F3 = 1954.9 Hz (B3 = 262.7 Hz) and F4 = 2886.17 Hz (B4 = 543.26 Hz). The value for formant two (spurious formant; cf. Ladefoged 2003: 122) can be disregarded, because it would mean that the stressed vowel DRESS is audibly realized as something closer to a STRUT vowel with the low back merger in place. Instead, it impressionistically clearly sounds like the mid front vowel /ɛ/, so that this value is regarded as incorrect and consequently discarded. The example also stresses the importance of confirming the automatically derived formant values impressionistically. In addition to the smaller differences in the formant values derived from the three settings in other DRESS tokens in the reading passage, the open question section of speaker 11CDMF revealed further consistency in the results of her lax vowels derived from setting (3).

4.4.5 Linguistic Context

The preceding and following phonological segments play a crucial role in the quality of the vowel. The formants used to determine the quality of vowels are also the most important correlates of consonants’ places of articulation (Thomas 2011: 98). Depending on the place of the segment, the formant transitions from the preceding segment and to the following segment differ between vowels (Hillenbrand, Clark and Nearey 2001: 754; also cf. Harrington 2010: 92). A very general pattern of the effects of places of articulation on vowel formants during the transitional phase is outlined in Table 4.10. This pattern can be summarized by a rule of thumb: labial consonants lower F2 for neighboring vowels, dorsal consonants raise F2 and coronal consonants raise F2 for back vowels but lower F2 for front vowels (Thomas 2011: 49).

However, these transitions are accounted for by the different measurement approaches outlined in Subsection 4.4.1 and/or by including linguistic environments of vowels as variables in regression analyses. Single-point measurements are usually made at the
midpoint of the vowel, which is furthest away from both transitional phases. The maximal distance approach takes measurements at maximum F1 in most of the cases. As the table shows, virtually all consonants cause F1 to be lowered during the transition, so that the maximum value of formant one per vowel indicates the point furthest away from the consonants, except for pharyngeals (/h/, which does not occur after a vowel in English). In multiple-point measurements, the transitions are usually arbitrarily excluded by disregarding the first and last two or three temporal measurement locations (e.g. Adank, van Hout and Smits 2004: 1731). In connected speech, very short vowels may have no steady-state portion (cf. Subsection 4.4.3.2), so that measurements will include formant values from these transitions, regardless of the different measurement approaches. I attempt to account for this issue by excluding vowels shorter than 70 ms.

In terms of including linguistic context in the regression analyses, recall that most studies on the Canadian Shift have included neither preceding nor following segments (e.g. Boberg 2005, 2005b, 2010; Hoffman 2010; Roeder 2012). Clarke, Elms and Youssef (1995) and D’Arcy (2005) included following phonological segment in their impressionistic analyses; Hall (fc.), Hollett (2006, 2007) and Labov, Ash and Boberg (2006b) included both, preceding and following segments. However, Hollett’s findings are somewhat inconclusive (cf. Section 4.1), and Labov, Ash and Boberg (2006b) never discuss any influence of the preceding segment on the vowels of the Canadian Shift. Hall (fc.) emphasizes that preceding alveolars favor lowering of the front lax vowels, whereas preceding nasals and liquids tend to inhibit it. She further stresses the importance of including preceding alveolars in research on the Canadian Shift, due to their frequencies of occurrence in English. Only the two impressionistic studies, Clarke, Elms and Youssef (1995) and D’Arcy (2005), discuss the relative contribution of following linguistic variables and social variables, but the latter were partially controlled for (young urban middle-class females only; role of style in D’Arcy 2005). None of the other studies revealed the relative contribution of social and preceding linguistic factors in the Canadian Shift on the mainland, so that I

<table>
<thead>
<tr>
<th>Place of Articulation</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial</td>
<td>Lowered</td>
<td>Lowered</td>
</tr>
<tr>
<td>Labiodental</td>
<td>Lowered</td>
<td>Lowered</td>
</tr>
<tr>
<td>Dental</td>
<td>Lowered</td>
<td>Raised (back rounded vowels), lowered (front vowels)</td>
</tr>
<tr>
<td>Alveolar</td>
<td>Lowered</td>
<td>Raised (central and back vowels), lowered (mid and high front vowels)</td>
</tr>
<tr>
<td>Palatal</td>
<td>Lowered</td>
<td>Raised</td>
</tr>
<tr>
<td>Velar</td>
<td>Lowered</td>
<td>Raised</td>
</tr>
<tr>
<td>Pharyngeal</td>
<td>Raised</td>
<td>Lowered</td>
</tr>
</tbody>
</table>

Table 4.10: Generalized effects of the common places of articulation on the first two formants during the transitional phases in consonant-vowel-consonant (CVC) contexts (cf. Thomas 2011: 101).
leave preceding context to future investigations and include only following segment as a variable in the regression analyses.

Other coarticulatory effects cannot be as easily controlled for, as their effect (contextual assimilation) is greater on the formant values further into the vowels’ durations or on the vowels as a whole. Consonants that have such effects in English are semivowels (glides) and liquids (cf. Harrington and Cassidy 1999: 72-73), but also nasals and to a certain degree velars, which may cause F2 and F3 to converge (Thomas 2011: 100). Glides preceding and following vowels generally result in an undershoot of vowel quality or formant quality which is perceptually compensated (Harrington 2010: 92; Harrington and Cassidy 1999: 72). Vowels before /r/ are broadly treated as a separate analytical category (cf. e.g. Wells 1982a). Syllable-coda /l/ is often vocalized when co-articulated with preceding vowels, which has created a series of back, rounded vowels from /ol/, /ul/ and /ul/ (Thomas 2001: 56). This series may have motivated the fronting of /o/, /o/ and /u/ in other phonetic contexts (Labov 1994: 332). In addition, coarticulation of /r/ and /l/ with preceding vowels can neutralize phonetic cues used to distinguish vowels, resulting in conditioned mergers (Thomas 2001: 50). The most common pre-/l/ mergers (not affecting the vowel system as a whole) are, for instance, the still/mill (FLEECE-KIT) and the fool/full (GOOSE-FOOT) merger in western Pennsylvania, the southern Midwest and the West of the U.S., but also the fail/fell (FACE-DRESS) merger (Labov, Ash and Boberg 2006b: 120; Thomas 2001: 50).

As before /l/, vowels before nasals may cause conditioned mergers of preceding vowels (Thomas 2001: 52). One common merger is the pin/pen or him/hem (KIT-DRESS) merger in the South of the U.S., southern California and southern parts of the Midwest (Labov, Ash and Boberg 2006b: 120; Thomas 2001: 52). The antiformants that nasals generally introduce may make the first formant more susceptible to being canceled in those areas, because DRESS is generally higher in the South than in the North of the U.S. After canceling of the first oral formant, the first nasal formant is reinterpreted as F1. The antiformants created by nasality make it particularly difficult to obtain first formant readings for preceding (front) vowels (Harrington 2010: 114; Ladefoged 2003: 135; Thomas 2001: 52; also cf. Subsection 4.3.3). The pin/pen merger has been reported to continually spread northward by diffusion as far north as Saint John, New Brunswick, and Halifax, Nova Scotia (Labov, Ash and Boberg 2006b: 220). In some varieties of North American English such as African American Vernacular English, following nasals are deleted and the preceding vowel is nasalized (Bailey and Thomas 1998: 89), a pattern usually found in French (Ladefoged: 136). Increased formant bandwidths are generally an indication of nasality (Ladefoged 2003: 137).

In varieties of North American English, a following nasal causes /æ/ to be raised and fronted to a quality of [ɛ] or [e], to which Labov, Ash and Boberg refer as the nasal system (2006b: 174; also cf. Boberg 2000: 5; Labov 1991: 5, Labov 1994: 197, 503; Thomas 2001: 52). Although this system is generally absent in Canada (Labov, Ash and
Boberg 2006b: 175), speakers from Halifax and Sydney, Nova Scotia, Saint John, New Brunswick and St. John’s, Newfoundland, show differences of 200 - 300 Hz or more in first formant values between /æ/ before nasals and /æ/ before other consonants (Labov, Ash and Boberg 2006b: 176). Especially in the Maritimes, /æ/ is similarly raised and fronted before /g/ (Labov, Ash and Boberg 2006b: 223).

For these reasons, I exclude the same consonantal environments as Labov, Ash and Boberg (2006b: 77) from the analyses in this thesis: following and preceding liquids (/l, r/), preceding glides (/j, w/) and in case of the front lax vowels /i, e, æ/ additionally following nasals (/n, m, ŋ/). Furthermore, I exclude /æ/ before /g/. All of these environments are, however, included for the purpose of normalization.

4.4.6 Tokens for Analysis

Based on the methodological considerations outlined above, I digitized the speech signals at 48,000 Hz and 24 bit. I re-sampled the recordings to 22,000 Hz for the analyses, which yields a Nyquist frequency of 11,000 Hz. Before determining the correct spectrogram and formant tracker settings in Praat 5.3 (Boersma and Weenink 2011), I investigated several different vowels per speaker manually. In a next step, I segmented the re-sampled speech, using impressionistic cues to determine the phrasal-accented lexical items. Emphatically stressed tokens (higher pitch) and vowels shorter than 70 ms were disregarded. Before I extracted the segmented vowel tokens for LPC analysis automatically, I determined the quality of each of the vowels via auditory impression during segmentation.

This procedure was repeated for each of the 34 speakers included in this thesis. I collected five to ten tokens of each vowel allophone and limited the most frequently occurring allophones to approximately ten tokens, in order to prevent skewing of a speaker’s vowel space by an over-representation of one or two vowels (cf. Labov, Ash and Boberg 2006b: 37). For the diphthongs FACE and GOAT I extracted two to three tokens and for vowels before /r/ one to two tokens per lexical set (cf. Appendix A.2). Depending on the amount and length of personal narratives, I collected between 250 and 450 vowels per speaker, which resulted in an average of 350 vowels (cf. Labov, Ash and Boberg 2006b: 37) and yielded a total of 11,803 vowels. After cleaning the data (cf. Subsection 4.5.1), 10,731 vowels remained for normalization. 4558 of these vowels were uttered in linguistic environments excluded in the analyses. The remaining 6173 tokens are comprised of the lax vowels KIT, DRESS, TRAP, LOT, STRUT and FOOT, the tense vowels FLEECE, THOUGHT and GOOSE and the phonological diphthong FACE (cf. Section 5.1).

50 I excluded all PALM tokens due to the small size of the lexical set. The word list contained the lexical item father, the reading passage contained the lexical item palm, but the free speech section contained virtually no tokens from this set.

I excluded all PALM tokens due to the small size of the lexical set. The word list contained the lexical item father, the reading passage contained the lexical item palm, but the free speech section contained virtually no tokens from this set.
4.5 Data Preparation

Prior to any analysis of the vowel tokens, I prepared the data set in two steps. First I cleaned the individual tokens, based on auditory impression, the spectrogram, Fast Fourier Transform (FFT) and the formant as well as the bandwidth values I automatically derived in Praat. Second, I normalized the cleaned data points based on Nearey’s (1977) log-mean normalization and the modifications made by Labov, Ash and Boberg (2006b: 39-40). The two steps are illustrated in detail below.

4.5.1 Data Cleaning

During the segmentation phase of the vowel token collection, I disregarded passages in the sound files with external noise such as traffic outside the buildings in which the recordings were made, short-term noise from within the buildings such as door slams and the like. In addition, very low frequencies caused by heartbeats and similar noise as well as hissing sounds caused by strong exhalation sometimes added to the complex sound wave of the speech signal. Vowels occurring under such circumstances have not been included in the final data set. Figure 4.18 shows the lexical item *vacation* uttered by speaker 09PDOM in the spontaneous speech part of the interview (segmented for visualization purposes). The oscillogram in the top panel shows the sound wave of the lexical item with additional noise in the last third of the vowel *face*. Such immeasurable vowel tokens were particularly unfortunate when they occurred in the word list and reading passage, as the total number of tokens in these two styles is limited to begin with. Overall, such tokens occurred, however, quite infrequently in the sound files.

After the segmentation of the lexical items of interest, a script was used to extract the formant readings and to write them to a spreadsheet. The automatically derived spreadsheet including the vowel tokens per speaker were imported into an MS Access database, which I used to adapt and change the layout of the different spreadsheets needed for plotting in R 3.0.3 (R Development Core Team 2006-13) or with the vowels package (Kendall and Thomas 2010), for normalization via NORM 1.1 (Thomas and Kendall 2007) and for the statistical analyses in R 3.0.3 and SPSS 21 (Statistical Package for the Social Sciences). Data cleaning was done in two subsequent steps: First I checked the formant bandwidths of the vowels per speaker and subsequently the formant values of the vowels per speaker.

Formant readings with bandwidths greater than 400 Hz should be either discarded or remeasured, as the formants have low amplitudes (Ladefoged 2003: 117), so that the LPC spectral analysis yields no sharply defined peaks on which the formant is located. Figure 4.19 shows an LPC spectral slice of the vowel *thought* at around 33% of the vowel’s duration in the lexical item *daughter* uttered by a young female (speaker 26PBYF) in interview style. The second formant is marked by a bandwidth of 2599 Hz and clearly much smaller in amplitude than the first formant. While this LPC spectral slice may yield
4.5. Data Preparation

uh, vacation places internationally we’ve been

| k | *FACE1 | f |
| k | *FACE | f |

*vacation

*Nr

Figure 4.18: The lexical item *vacation uttered by male speaker 09PDOM in the spontaneous speech section of his interview. Noise is added to the complex waveform towards the end of the FACE vowel.

a second formant with such a high bandwidth due to the settings of the LPC analysis, these settings have proven to yield accurate values for most of her vowels. The figure serves primarily for the purpose of illustrating a formant reading with a very high bandwidth. The second formant value with such a bandwidth may lie anywhere between 1370 Hz and 1520 Hz ($\Delta F2 = 150$ Hz). As mentioned in Subsection 4.4.5, a high bandwidth may also be indicative of nasality, which does not apply here.

For back-rounded vowels such as THOUGHT in particular, high bandwidths are indicative of possible automatic misreadings, because formants one and two are close together in frequency and in many varieties of English both below 1000 Hz (cf. e.g. Ladefoged and Johnson 2011: 196), which can be resolved by adjusting the coefficients for the LPC algorithm (cf. Subsection 4.3.3). Likewise, for high front vowels such as FLEECE, formants two and three are close together in frequency and might equally be confused by an automatic tracking algorithm. The latter issue was in fact the case in the majority of the
vowel tokens with high bandwidths for formant two in my data set (approximately 20% of the automatically tracked tokens; see below). For FLEECE and KIT vowels uttered by females in particular, formant two often had bandwidths up to 1000 Hz, while formant three had bandwidths of below 100 Hz, and the difference between the two formant frequencies was at times smaller than 50 Hz. In addition, the fundamental frequency was mistracked as formant one for a small number of females with high-pitched voices (also cf. Watson and Harrington 1999: 461).

In order to clean such faulty measurements, I ordered the spreadsheet according to height of the bandwidth values and investigated those tokens more closely where the bandwidth exceeded a value of 400 Hz at the 33% measurement point (cf. Subsection 4.4.2). If the bandwidths were below the threshold of 400 Hz for the measurements at 20%, 25%, 30% and 40%, I manually remeasured the tokens via different Fast Fourier transform (FFT) settings in order to investigate some of the individual harmonics more closely in relation to the automatically extracted formants (cf. Watson and Harrington 1999: 461). If the bandwidths exceeded 400 Hz at all five temporal locations (20%, 25%, 30%, 33% and 40%), I discarded the vowel tokens.

Out of the 11,803 vowels automatically tracked in Praat 5.3 (Boersma and Weenink 2011), 378 vowel tokens had bandwidths greater than 400 Hz for formant one at the measurement point of 33% of the vowel’s duration, and 1983 vowels had such bandwidths for formant two. From these I had to discard 83 vowel tokens based on formant one

Figure 4.19: LPC analysis of THOUGHT in daughter uttered by female speaker 26PRYF. The bandwidth of the second formant is 2599 Hz.
misreadings and 989 vowels based on formant two misreadings, adding up to 1072 deleted vowel measurements and 1289 manually re-measured vowels.

After I cleaned the data set based on bandwidths, I ordered it according to the formant readings per vowel in order to identify formant one and two readings that are too high or low for the respective vowel. If, for instance, a DRESS vowel had a formant two value of 1500 Hz or less, I listened to the respective lexical item again in order to impressionistically confirm or reject the accuracy of the measurement. In many of these rare cases, the accuracy had to be rejected (e.g. the vowel in said did not sound like a schwa) so that I re-measured such tokens manually as well. Fortunately, none of the respective tokens had to be deleted. In rare occasions, I interpolated the 33% measurement point based on the 20% and 40% measurement points. If the formant values were less apparently inaccurate, i.e. closer to the main dispersion of the respective vowel per speaker, style and lexical item, I compared them to other formant readings taken from the same speaker in order to investigate whether such a reading was a single occurrence for that speaker or whether other tokens of the same vowel had similar automatically tracked formant values. In very few individual cases, I based my decision of whether to re-measure a token or not on the literature. For instance, TRAP tokens in the lexical set catch are often raised and fronted in Newfoundland English, so that I considered young females’ TRAP tokens after voiceless velars and before voiceless affricates (and possibly stops and fricatives) with F2 readings of 1900 Hz or more as accurate, although the majority of their TRAP tokens yielded F2 readings of 1700 Hz and less. Such lexical effects are accounted for by modeling lexical items as random effects in the multivariate statistical analysis (cf. Section 4.6). Their total occurrences in the data set are, however, negligible.

4.5.2 Normalization

The need to normalize acoustic measurements of vowel tokens uttered by different speakers goes back to studies conducted as early as of the 1950s. Peterson and Barney (1952: 182), for instance, have shown with predominantly General American speech data that the spectral peaks or formants have remarkably different frequencies for the very same vowel token in hVt contexts when these tokens are uttered by different speakers. Peterson and Barney (1952: 183) conclude that in general, children have the highest formant frequencies, women intermediate ones and men the lowest. This finding generally correlates negatively with the vocal tract sizes, i.e. small vocal tracts will produce high formant frequencies. They have further shown that different vowels have quite similar formant frequencies when uttered by speakers with different vocal tract sizes (Peterson and Barney 1952: 182; also cf. Disner 1980: 253). In summary, no single vowel token uttered by the same speaker in the same linguistic environment will have the same for-

---

51 A (voiceless) glottal fricative is followed by a vowel, which in turn is followed by a voiceless alveolar stop (e.g. heat, hot).
mant frequencies; in fact intra-speaker variation was statistically significant in Peterson and Barney’s study (1952: 182; also cf. Clopper 2009: 1430).

Over the decades of acoustic phonetic research, consensus has been reached that normalization of formant readings from more than one speaker is necessary (e.g. Hindle 1978; Disner 1980; Hillenbrand et al. 1995; Thomas 2002; Thomas and Kendall 2007; Clopper 2009; Watt, Fabricius and Kendall 2011). Some authors have taken Peterson and Barney’s generalized conclusion at face value, maintaining that normalization is necessary when children, women and men are compared in a data set, although they acknowledge the underlying reason: the difference in vocal tract sizes and not merely in speakers’ ages and/or gender (cf. Baranowski 2013b: 410; Watt, Fabricius and Kendall 2011: 111). In fact, a further goal of normalization, in addition to factoring out differences in formant frequencies due to physiological differences between speakers (1), is to preserve inter-speaker variation due to social differences such as age and gender, due to regional differences and due to sound changes (2). Two further goals have been collectively identified by the authors mentioned above: (3) to preserve vowel class and phonological distinctions and (4) to model the cognitive processes that allow human listeners to identify/normalize vowels uttered by different speakers (e.g. Disner 1980: 253; Fabricius, Watt and Johnson 2009: 415; Flynn 2011: 2; Thomas 2013: 112; Thomas and Kendall 2007: n.p.; Watt, Fabricius and Kendall 2011: 112). For sociophonetic research such as the present study, the latter two goals are the least important, so that normalization techniques specifically aimed at meeting them are not applicable here (cf. Thomas and Kendall 2007).

Various procedures have been proposed to meet the goals outlined above. The procedures can be categorized into groups, based on the amount and type of information that is required (Adank, Smits and van Hout 2004: 3099; Clopper 2009: 1431). They can either apply to individual vowels (vowel-intrinsic) or to sets of vowels (optimally all vowels; vowel-extrinsic; Baranowski 2013b: 411). Vowel-intrinsic normalization techniques use the information available in one vowel token such as F0, F1, F2, F3, sometimes F4, formant bandwidths and/or amplitudes to normalize that token (Disner 1980: 254; Flynn 2011: 3; Thomas 2002: 174; Thomas and Kendall 2007: n.p.). In addition, vowel-intrinsic techniques typically involve a nonlinear transformation of the frequency scale (log, mel and Bark scales; Adank, Smits and van Hout 2004: 3099). Vowel- (and formant-) intrinsic methods such as mel, Bark and ERB (equivalent rectangular bandwidth) are, however, not normalization procedures as such, but psychoperceptual transforms (Clopper 2009: 1432; Fabricius, Watt and Johnson 2009: 417), i.e. they were developed in order to model human vowel perception (Adank, Smits and van Hout 2004: 3099; cf. goal four above). Vowel-extrinsic techniques use the information available in multiple vowel tokens uttered by the same speaker (Fabricius, Watt and Johnson 2009: 416; Flynn 2011: 3; Thomas and Kendall 2007). They were developed to obtain higher percentages of correctly classified vowel tokens for automatic speech recognition purposes (Adank, Smits and van Hout 2004: 3099).
The procedures can also be categorized as to whether they are formant-intrinsic or formant-extrinsic, i.e. a procedure may use only formant one measurements to normalize a formant one value, or it uses F1, F2 and F3 measurements to normalize a formant one value (Fabricius, Watt and Johnson 2009: 416; Flynn 2011: 3). A final classification for normalization techniques is whether they use information from one speaker (speaker-intrinsic) or from a population of speakers (speaker-extrinsic; Fabricius, Watt and Johnson 2009: 416; Flynn 2011: 3; Thomas and Kendall 2007: n.p.). Although the latter technique is rarely used in acoustic phonetics literature due to its complexity (Flynn 2011: 3), it is relatively common in North American sociolinguistic research such as in the *Atlas of North American English* (ANAE; Labov, Ash and Boberg 2006b; cf. Fabricius, Watt and Johnson 2009: 416).

Lack of consensus is maintained with regard to the question of which normalization procedure to use (Flynn 2011: 2; Thomas 2002: 174). None of the proposed normalization algorithms is truly optimal, but some of them are quite effective in minimizing inter-speaker variation (Thomas 2002: 174). Among the rather famous and effective procedures are, for instance, Lobanov’s (1971), Nearey’s (1977) and Watt & Fabricius’ (2002). A number of normalization algorithm comparisons have been conducted in order to provide rough guidelines of which algorithms perform better and worse (e.g. Adank 2003; Adank, Smits and van Hout 2004; Clopper 2009; Disner 1980; Fabricius, Watt and Johnson 2009; Flynn 2011). The studies concur that vowel-intrinsic techniques such as Syrdal and Gopal’s (1986) formant-extrinsic Bark-transformation model, in which differences between Bark-converted F0, F1, F2 and F3 are computed to model vowel height and advancement (Baranowski 2013b: 410; Clopper 2009: 1433), are least effective in removing variation in the raw formant Hertz values due to speaker age and gender (Adank 2003: 98; Adank, Smits and van Hout 2004: 3103; Clopper 2009: 1440; Flynn 2011: 22). The detailed findings of these studies naturally differ, most likely due to the different varieties/languages/regions the data were sampled from and the different number of speakers and evaluation methods used in the studies.

Very generally, the advantages of vowel-intrinsic normalization methods are that they do not require the measurement of the complete vowel system of a speaker (Adank, Smits and van Hout 2004: 3105; Baranowski 2013b: 410; Clopper 2009: 1433). As a consequence, they are better suited to compare languages with differing vowel inventories than vowel-extrinsic normalization techniques (Adank 2003: 98; Baranowski 2013b: 410; Fabricius, Watt and Johnson 2009: 417; also cf. Disner 1980: 257; Thomas 2013: 112). They also perform better than vowel-extrinsic normalization procedures at examining vocal setting, i.e. the habitual shifting of an articulator in some direction (Thomas 2013: 112). Another advantage of vowel-intrinsic normalization procedures is that they are perceptually plausible models of human speech processing. That is, human listeners are perfectly able to identify a vowel uttered to them without having to listen to the speaker’s complete vowel system (Clopper 2009: 1440). However, in addition to their rather poor perfor-
mance, vowel-intrinsic normalization techniques have another serious drawback: they rely heavily on F3 measurements which are most often far from accurate in sociophonetic research (Baranowski 2013b: 410). Particularly since the homes of the respondents provide relatively many sources of noise, but also due to some voice qualities, automatic formant measurement settings work unreliably for F3. In addition, rhoticized vowels lower F3, which consequently affects vowel-intrinsic normalization techniques (Baranowski 2013b: 411).

As mentioned above, most studies evaluating normalization techniques attested great performances to the vowel-extrinsic normalization techniques such as Lobanov’s (1971), Nearey’s (1977) and Watt & Fabricius’ (2002) – when compared to vowel-intrinsic ones. Lobanov’s formant-intrinsic (and speaker-intrinsic) method expresses formant values relative to the hypothetical center of a speaker’s vowel space. According to Adank, Smits and van Hout (2004: 3101), Clopper (2009: 1438), Disner (1980: 255) and Flynn (2011: 7), it simply converts each formant for each speaker (\(F_i\)) into a z-score by subtracting a speaker’s mean formant value across all vowel tokens (\(\mu_i\)) and subsequently dividing by the standard deviation for the formant (\(\sigma_i\)) across all vowels for that speaker:

\[
z = \frac{(F_i - \mu_i)}{\sigma_i}.
\]

Thomas and Kendall (2007: n.p.) stress, however, that in his original formulation of the normalization procedure in his (1971) publication, Lobanov used the rms (root mean square) deviation and not standard deviations as the denominator. This seems to have originated as a typographical error, as in his equation two he writes “\(\sigma_i\)” (the mathematical symbol for standard deviation; Lobanov 1971: 606), but in the following text he maintains that “[...] \(\delta_i\) is the rms deviation of \(F_i\) [...]”. Thomas and Kendall (2007: n.p.) add that in general both deviations yield very similar values. A transform into z-scores (standardizing), as reported by Adank, Smits and van Hout (2004: 3101) referring to Lobanov’s normalization procedure, is often used in statistics to transform normally distributed data sets to a uniform normal distribution (Flynn 2011: 7). It is particularly necessary to compare values coming from different scales (Gries 2009: 121), as is the case for the ranges of formants one and two. The resulting z-scores of this transformation simply indicate how many standard deviations each formant value deviates from the mean of all formant values (cf. Gries 2009: 122). The vowel spaces of all speakers are then anchored at the individual formant means. The range of possible normalized formant values is scaled for each speaker so that the vowel space is modeled within ±2 standard deviations of the mean for each formant (Clopper 2009: 1438), i.e. 95% of all vowel tokens, provided they are normally distributed.

Nearey’s (1977) vowel-extrinsic normalization technique(s) is actually comprised of two algorithms, a formant-intrinsic one and a formant-extrinsic one (both are speaker-intrinsic). The former is sometimes referred to as single log-mean method (Adank, Smits
and van Hout 2004: 3101) or individual formant mean normalization (Clopper 2009: 1436) and the latter as shared log-mean method (Adank, Smits and van Hout 2004: 3101) or grand mean normalization (Clopper 2009: 1436). In both algorithms, the formant values are first log-transformed. In a second step, each speaker’s vowels are scaled by subtracting the mean log formant value across all vowels for the same speaker from each individual formant value (Clopper 2009: 1435). After the natural logarithm of a speaker’s formant value ($F_i$) is taken in the formant-intrinsic algorithm, the mean of the log-transformed formant value across all vowels for the speaker ($\mu_{\ln(F_i)}$) is subtracted from it (Adank, Smits and van Hout 2004: 3101; Clopper 2009: 1435; Disner 1980: 255; Flynn 2011: 7), yielding the normalized formant value ($F_i^N$):

$$F_i^N = \ln(F_i) - \mu_{\ln(F_i)}.$$

In the formant-extrinsic algorithm, the mean of the log-transformed formant value of all formants of all vowels for the speakers ($\mu_{\ln(F_j)}$) is subtracted from the natural logarithm of a speaker’s formant value ($F_i$; Adank, Smits and van Hout 2004: 3101; Clopper 2009: 1435; Disner 1980: 255; Flynn 2011: 7):

$$F_i^N = \ln(F_i) - \mu_{\ln(F_j)}, \forall j = 1, \ldots, n.$$

The effect of Nearey’s normalization is to align all of the speakers’ vowel spaces at the mean formant frequency for each speaker. The alignment point is the intersection in the F1xF2 plane of either the individual means of formants one and two or the grand mean across all formants (Clopper 2009: 1435).

Watt & Fabricius’ (2002) normalization procedure – typologically formant- and speaker-intrinsic (cf. Fabricius, Watt and Johnson 2009: 417) – expresses values relative to the constructed centroid of a speaker’s vowel space, similarly to Lobanov’s (1971) algorithm (cf. Flynn 2011: 7). The vowel space is modeled as a triangle determined via the minimum and maximum F1 and F2 per speaker, as outlined in Figure 4.20. The first point in the F1xF2 plane is identified via a speaker’s mean F1 and F2 values of the fleece lexical set and labeled [i]. These two mean values represent the speaker’s minimum F1 and maximum F2 value. The second (or lowest) point of the triangle is derived via the same speaker’s mean F1 and F2 values of the trap lexical set, and labeled [a] (alternatively of the start lexical set; Watt and Fabricius 2002: 163). The mean F1 value represents the speaker’s maximum F1. The third point of the triangle is constructed, unlike the former two. The minimum F1 observed in the fleece lexical set is used to construct the F1 and F2 coordinates of the third point, labeled [u']). These two constructed values represent the minimum F1 and F2 possible for the speaker. The centroid, $S$, of this triangle is determined as the grand mean of points [i], [a] and [u']) using the following equation (cf.
Figure 4.20: The conceptualization of the vowel space and the construction of the centroid $S$ in Watt & Fabricius’ (2002) normalization technique. Figure taken from Flynn (2011: 7).


$$ S(F_i) = \frac{F_i[i] + F_i[a] + F_i[u']} {3} $$

This original formulation of Watt & Fabricius’ normalization procedure has received some criticism of distorting the bottom of the vowel space (e.g. Bigham 2008: 136), because it relies heavily on minimum and maximum F1 and F2 values. As Thomas and Kendall (2007: n.p.) note, some varieties may show one vowel at the very bottom of the vowel space, whereas others show two low vowels in a slightly higher position, particularly in North American varieties of English (“a butterfly-like pattern”; Baranowski 2013b: 411). That is, in some varieties [a] may have an F2 value significantly higher or lower than the median value in the F2[i] to [u’] line – the center point (Fabricius, Watt and Johnson 2009: 420-421). In response to this criticism, Fabricius, Watt and Johnson (2009: 421) modified their original procedure in that the F2 values are normalized without the F2 value of the TRAP vowel (also cf. Flynn 2011: 7):

$$ S(F_i) = \frac{F_i[i] + F_i[u']} {2}, \ i = 2. $$

According to Thomas and Kendall (2007: n.p.), this modification lessens the distortion introduced by the original version of Watt & Fabricius’ (2002) method. Among others, Bigham further modified the formulation of this normalization procedure in that he modeled the vowel space of his North American speakers as a quadrilateral rather than a triangle, which he “feel[s] better represents most dialects of American English” (Bigham
4.5. Data Preparation

2008: 135). The quadrilateral was conceptualized from the KIT vowel, via the observed (not constructed) GOOSE vowel and the averaged LOT and THOUGHT vowels to the TRAP vowel (Bigham 2008: 135). His feeling can be understood as an indication that the effectiveness of a normalization procedure is dependent on the variety of English to be normalized (see below).

All three of the outlined vowel-extrinsic normalization techniques share further disadvantages (Thomas and Kendall 2007: n.p.): They require the whole vowel system to be included in the normalization algorithms, because including a subset of the vowel inventory of only one speaker skews the normalized values. They may further be impaired when two varieties with a differing vowel inventory are compared – the domain in which vowel-intrinsic normalization procedures perform much better. As Thomas observes, Nearey’s (1977) vowel-extrinsic algorithm poses particular problems for inter-varietal comparisons of, for example, North American Englishes. If speakers from a variety with fronted GOAT, FOOT and GOOSE vowel systems are compared to speakers from a variety with backed GOAT, FOOT and GOOSE vowel systems, the scaling factor based on all vowels would be skewed towards higher F2 values (fronting) in the former variety and towards lower F2 values (backing) in the latter variety (2002: 175).52 This observation was extended by Thomas and Kendall (2007: n.p.) to all vowel-extrinsic normalization algorithms, which would consequently be appropriate only for intra-varietal speaker comparisons, as is the case in the present study.53

As mentioned earlier, studies that evaluated different normalization procedures generally find no substantial difference in the performances of the three vowel-extrinsic procedures outlined above, such as Clopper’s (2009: 1440) impressionistic comparison based on a male and a female speaker from a General American variety. Other studies looked at several goals of normalization procedures. Adank, van Hout and Smits, for instance, looked at the preservation of phonemic (goal three) and sociolinguistic variation (goal two) as well as the reduction of the effects of physiological difference between speakers (goal one), based on 160 Dutch speakers. They excluded Watt & Fabricius’ (2002) normalization technique, but for the remaining two they found that Nearey’s (1977) formant-intrinsic procedure performed second best after Lobanov’s (1971) with regard to goal three. Gender-specific differences (in the physiology; goal one) in the fundamental frequency were dealt with equally well by Nearey’s and Lobanov’s algorithms at a statistically significant level, whereas gender-specific differences in F1, F2 and F3 were dealt with best by Nearey’s formant-intrinsic procedure. In terms of goal two, the latter again performed better than Lobanov’s procedure (cf. Adank, van Hout and Smits 2004: 3102-3104; also cf. Adank 2003: 98).

52 Note that he does not maintain this observation in his article on sociophonetics in the second edition of The Handbook of Language Variation and Change (cf. Thomas 2013).

53 Note that this observation goes most likely back to Disner’s inter-language comparisons with Nearey’s and Lobanov’s normalization procedures, among others (cf. Disner 1980: 257-260).
Adank, van Hout and Smits’s findings are in line with those of Disner (1980: 256), who also found that no specific procedure is the most effective, but in general Nearey’s performed best and only slightly better than Lobanov’s. She investigated the effects of normalization in terms of scatter reduction in the F1xF2 plane (1980: 253), linguistic validity, i.e. the quality of a vowel that is attributable to the language itself, rather than to the speaker (1980: 256), and cross-language comparisons. Her analysis is based on data from male speakers of six languages: English, Norwegian, Swedish, German, Danish and Dutch (1980: 254). Nearey’s and Lobanov’s methods performed both best in scatter reduction, but along with the others Disner tested they fared poorly when evaluated for their linguistic validity (1980: 257). Likewise, they performed poorly when evaluated for their cross-language validity (cf. Disner 1980: 257-260): Nearey’s performed particularly poorly in the Danish-English comparison and Lobanov’s in the German-English comparison (1980: 260).

Disner’s results suggest that the performances of the normalization procedures depend on the language (1980: 260). In the same vein, a certain variety of English may thus be better normalized by a certain procedure than another. This point of view has received support from more recent evaluations such as Fabricius, Watt and Johnson’s (2009) and Flynn’s (2011). The former study found, for instance, that while Nearey’s (1977) formant-intrinsic procedure performed crucially worse than Lobanov’s (1971) and Watt & Fabricius’ (2002) algorithms when applied to data from RP speakers, it performed much better (more similar to the latter two) when applied to data from speakers of Aberdeen English in terms of equalizing vowel space areas (Fabricius, Watt and Johnson 2009: 424). Flynn finds similar results for his speakers from Nottingham, England, in that (among others) Lobanov’s performs best, followed by Fabricius, Watt and Johnson’s (2009) modified (after Bigham 2008) and subsequently Watt & Fabricius’ (2002) original normalization technique. Nearey’s procedures both perform less well when compared to these with regard to equalizing vowel spaces (2011: 16). In terms of aligning vowel spaces, Flynn (2011: 17) finds that Watt & Fabricius’ techniques in general (including Bigham’s 2008 modification) outperform Lobanov’s and Nearey’s when applied to data from speakers of Nottingham English.

It has to emphasized that each of the evaluative studies used a different conceptualization in order to compare the vowel spaces between different speakers: Flynn (2011), for instance, used a vowel quadrilateral, Clopper (2009) added lines between all vowels, except for the FLEECE and GOOSE vowels, yielding no particular shape of the vowel space. As Flynn (2011: 23) notes, differences in this conceptualization may confound the measure of ‘equalizing vowel spaces’ and ‘aligning vowel spaces’ between speakers. Other authors, for instance, hypothesized that a pentagon (Jacewicz, Fox and Salmons 2007) or even a polygon (Fox and Jacewicz 2008) better modeled the complete vowel space area used by speakers and thus provided more accurate percentages of normalized vowel space overlap between speakers.
As mentioned earlier, for General American English, the three vowel-extrinsic normalization algorithms do not differ substantially from one another (cf. Baranowski 2013b; Clopper 2009; Disner 1980; also cf. Adank 2003; Adank, van Hout and Smits 2004 for Dutch), and Nearey’s (1977) formant-extrinsic procedure is the most commonly used one in North American sociophonetics (cf. Baranowski 2013b: 411; Fabricius, Watt and Johnson 2009: 416; Flynn 2011: 3; Watt, Fabricius and Kendall 2011: 115). For the atlas (ANAE), Labov, Ash and Boberg found that “[Nearey’s] log-mean normalization was most effective in eliminating male–female differences due to vocal tract length and preserving the social stratification of stigmatized variables that had been established by auditory impressions” (2006b: 39; based on research for the Philadelphia project reported in Labov 1994). In order to maximize comparability to the atlas (Thomas and Kendall 2007: n.p.),54 I decided to use the same normalization procedure that Labov, Ash and Boberg employed. For this reason, I subjected the raw vowel measurements in Hertz to the NORM suite (Thomas and Kendall 2007), using the Labov ANAE Telsur G option (cf. Bigham 2008: 134; Fabricius, Watt and Johnson 2009: 423).

The ANAE normalization method is a variation of Nearey’s formant-extrinsic procedure with a speaker-extrinsic overlay in the form of a parameter, the group log mean G, or Telsur G (Labov, Ash and Boberg 2006b: 39; Watt, Fabricius and Kendall 2011: 115). The parameter was successively updated in the atlas as the number of North American respondents increased. Once the sample size reached 345 speakers, a significant change in G ceased so that it was kept at that figure, G = 6.896874 (Labov, Ash and Boberg 2006b: 40; Watt, Fabricius and Kendall 2011: 115). This value was derived via the following formula:

\[
G = \frac{\sum_{k=1}^{p} \left( \sum_{j=1}^{m} \left( \sum_{i=1}^{n} \ln(F_{i,j,k}) \right) \right)}{m \times \sum_{i=1}^{p} n_i},
\]

where \(p\) is the number of speakers measured; \(m\) is the number of formants (two in the Telsur data: F1 and F2); and \(n\) is the number of vowel tokens measured for a given speaker (Labov, Ash and Boberg 2006b: 39). The group log mean G is subtracted for the individual log mean S for that speaker, in order to normalize any given speaker (Labov, Ash and Boberg 2006b: 39):

\[
S = \frac{\sum_{j=1}^{m} \left( \sum_{i=1}^{n} \ln(F_{i,j}) \right)}{m \times n}.
\]

The anti-log (exp) of this difference is the uniform scaling factor F for that speaker (Labov, Ash and Boberg 2006b: 40):

\[
F = \exp(G - S).
\]

54 Note that Bigham normalized using ANAE’s (Labov, Ash and Boberg 2006b) method in addition to his modified version of Watt & Fabricius (2002) “for those specific points in the discussion where a direct comparison to the findings of the ANAE is needed” (Bigham 2008: 134).
For male speakers, F has a value greater than one, so that their vowel spaces will be expanded (large vocal tracts yield smaller formant values and vice versa; cf. Peterson and Barney 1952). Likewise, the F value for female speakers has a value smaller than one, so that their vowel spaces will be contracted. In this way, the formant means of different vowels in the F1xF2 plane display the course of the sound change in progress (Labov, Ash and Boberg 2006b: 40).

Flynn (2011: 13) tentatively adds that “it could be argued that [taking the anti-log] might reverse some of the effects of the normalisation, since the exponential function is the inverse function of the natural logarithm”. He does, however, not discuss the consequences of taking the exponential in any further detail. Thomas and Kendall (2007) only caution insofar as they explicitly emphasize that scaling of normalized formant values back into Hertz values may undo much of the normalization procedure, unless the formant values for all speakers are used. Since Labov, Ash and Boberg’s normalization, which returns normalized Hertz values, is one of the few speaker-extrinsic ones, it does include all speakers’ formant values (cf. Labov, Ash and Boberg 2006b: 39; Watt, Fabricius and Kendall 2011: 115). In addition, many, if not most, acoustic phonetic studies on the Canadian Shift in different speech communities report normalized Hertz values as well (e.g. Boberg 2005, 2008b, 2010; De Decker 2002; Hoffman 2010; Hoffman and Walker 2010; Roeder and Jarmasz 2007; Roeder 2012; Roeder and Gardner 2013; Sadlier-Brown and Tamminga 2008).

4.6 Data Analysis and Statistical Modeling

Before I select any data exploration method or statistical test for the analyses of the Canadian Shift vowels, I review and summarize the respective assumptions of these statistical methods. This review forms the core of this section, starting with data exploration methods in Subsection 4.6.1 and continuing with statistical tests in Subsection 4.6.2. In Subsection 4.6.3, I subsequently review and motivate the selection of the statistical tests I employ. Any statistical evaluation of the data was preceded by testing the respective assumptions for each statistical method in order to preclude the possibility of significant results being due to the violation of these assumptions.

The plot of all the vowels included in the analysis in Section 5.1 provides a first overview of the distribution of the data and reveals which subsequent steps have to be taken in order to assess the behavior of the vowels in question statistically. The plot is based on the arithmetic means per vowel of all the speakers per age group, including the respective standard deviations. Since the tokens per vowel and age group are not normally distributed, the means are an inaccurate summary of the data set. In such cases, the median may be the better measure of central tendency. A plot based on medians, however,

---

55 The statistical test for normality I used is explained in detail in Subsection 4.6.2.1. The results of the tests can be found in Appendices B.8 to B.10.
4.6. Data Analysis and Statistical Modeling

provided an identical picture of the vowels under analysis (cf. Appendix C). The only remarkable difference is the behavior of the FLEECE vowel of young speakers: Its median is less retracted than the mean, but the total number of tokens is far too small to make valid assumptions in this fashion. I decided to use the vowel plot based on means, because both plots are essentially the same.

4.6.1 Assumptions of Data Exploration Methods

The two methods I employ to explore my data set in order to identify clusters are Decision Trees and Optimal Binning. Decision Trees are an excellent tool for revealing significant groupings in the data set (Mendoza-Denton, Hay and Jannedy 2003: 131) and are increasingly used in sociolinguistic research – in particular a complex type, random forests (cf. e.g. Eddington 2010; Tagliamonte and Baayen 2012).

As a first step, a Decision Tree algorithm takes all of the data at once and attempts to split it into two significantly different groups or nodes. Ideally, the splits minimize variation within categories and maximize variation across categories, while attempting all possible classifications of the independent variables (Mendoza-Denton, Hay and Jannedy 2003: 129). If, for instance, the first and second formants of one vowel are entered as the dependent variables and phonological contexts are entered as the independent variables, the classification tree attempts to separate the vowel’s mean formant values according to preceding or following phonetic environment. Ideally, the algorithm separates all of the vowels before and/or after approximants from the rest of the data, as approximants are known to lower second formant values (cf. e.g. Thomas 2001: 50). In that way, the variation in vowel means within approximants is reduced, but is increased across groups of manner of articulation. At any given node, the algorithm identifies the best split by using the maximum reduction of deviance over all possible splits. The algorithm stops as soon as the number of cases reaching each leaf is small or the leaf has reached sufficient homogeneity relative to the root node (Mendoza-Denton, Hay and Jannedy 2003: 129).

One of the main problems of Decision Trees is that they are prone to overfitting. They attempt to find an explanation for each case in the data set, so that they become recursive (they continue repeating the fitting process) in the end. As a result, Decision Trees or random forests are not suitable to identify possible predictors for the dependent variable in the data set. Such exploration of data is thus unsuitable for formulating a posteriori hypotheses. Although methods such as pruning help to reduce overfitting, it is safer to only use independent variables that are a priori known to predict the behavior of other variables (cf. e.g. Eddington 2010: 267, 272; Mendoza-Denton, Hay and Jannedy 2003: 129; Tagliamonte and Baayen 2012: 163). Apart from this disadvantage, Decision Trees are an attractive method for linguistic data exploration, as they handle interactions between independent variables automatically, are completely non-parametric and make no

Since Decision Trees can model the relationship of two or more continuous (independent) variables, their computational calculation is resource-exhaustive and frequently inaccurate (Fayyad and Irani 1993: 1022). Optimal binning converts continuous variables to discretized or nominal variables for the purpose of optimal data fitting (Cleophas and Zwinderman 2013: 38), so that it helps to improve the outcomes of Decision Trees (Fayyad and Irani 1993: 1022). However, the categories of following and preceding phonological context are finite, unlike continuous variables, so that I consider the output of the Decision Trees as accurate.

Optimal binning is a so-called non-metric method for describing a continuous predictor variable in the form of best fit categories for making predictions. [...] It uses an exact test called the entropy method, which is based on log-likelihoods. It may, therefore, produce better statistics than traditional tests (Cleophas and Zwinderman 2014: 123).

Optimal binning therefore does not assume that the data is normally distributed. Similar to Decision Trees, it tries to find best splits in continuous data sets, so that infinitely possible values such as formant frequencies can be categorized into greater and smaller than a certain value. I employed Optimal Binning to derive those mean values of formants one and/or two at which the youngest speakers of my data set behave significantly different, so that the dependent variables in the logistic regression can be discrete or in this case binary. Although this method is a form of data mining (cf. Eddington 2010: 266), I can combine the result of optimal binning with a priori hypotheses (cf. Bortz and Döring 2006: 380): higher mean values of the first formant indicate lowering of a vowel, and lower mean values of the second formant indicate retraction of a vowel. In addition to the hypotheses, the results of linear regression are triangulating the results of logistic regression.

4.6.2 Assumptions of Statistical Tests

Before the analytical evaluation of the data, the nature of the statistical tests has to be determined. Since many of the dependent and independent variable combinations are normally distributed and have homogeneous variances, I use the same parametric tests for the analyses as the relevant literature suggests (e.g. Boberg 2005, 2008b, 2010; Clarke, Elms and Youssef 1995; D’Arcy 2005; Di Paolo, Yaeger-Dror and Wassink 2011; Hollett 2006, 2007; Labov, Ash and Boberg 2006b) such as the t-test for comparison of means, analysis of variances (ANOVAs) for a multiple comparison of means and linear (and logistic) regression (cf. Quinn and Keough 2002: 436).

A graphical representation of the data distribution in the form of box-and-whisker-plots (boxplots) preceded any descriptive or analytic statistical test. As boxplots contain
all the necessary information about the distribution of the data, they help to select the appropriate statistical test as its assumptions as well as most of its results can be anticipated (cf. Bortz and Döring 2006: 374). Since boxplots rely on the median as measure of central tendency of a data set, I added the mean in numbers and in the form of a plus-sign to the graphs (cf. Gries 2009: 118), so that the difference between mean and median is visible (in an ideally normally distributed data set with no outliers, the mean and median are identical in values; cf. Section 4.6.2.1), and I added a dotted line to indicate the grand mean of the data subset under analysis (cf. Gries 2009: 276).

The size of the boxes is determined by the interquartile range of the data set as the measure of dispersion, i.e. if all the data points are sorted in ascending or descending order, the interquartile range extends from the data point at 25% to the data point at 75% (the median divides the sorted data set in two equal halves of data points and the median of each half represents the 25% and 75% values; cf. Gries 2009: 107), and outliers are thus excluded. Boxplots help to determine outliers in a data set in a more or less simplistic – but sufficient – way, as they classify each data point more than 1.5 interquartile ranges away from either the 25% boundary or 75% boundary as an outlier (cf. Gries 2009: 119). The position of the box within the whiskers, or rather the unequal length of the whiskers, reveals much about the normality and skewness of the distribution of the data set; the length of the box hints at the kurtosis of the number of data points near the median/mean (cf. Chapter 5 for examples). However well a graphical representation aids in inferring the nature of the distribution of the data points, it has to be tested for normality.

### 4.6.2.1 Normality and Outliers

One such statistical test with a very high power is the Shapiro-Wilk test (Seier 2002: 1). It is very conservative and especially well suited for small data sets (3 < n < 50), but it is very sensitive to outliers and the amount of identical values in different data points (i.e. ties; cf. Gries 2009: 150; Ivezic et al. 2014: 159; Seier 2002: 3, 5; Shapiro and Wilk 1965: 610). A sensitivity to outliers means that with outliers being present, the test is likely to reject the null hypothesis, although it is actually true (Gibbons, Bhaumik and Aryal 2009: 301). The null hypothesis for the Shapiro-Wilk test is that the distributions of two samples or populations are normally distributed and the alternative hypothesis is that the distributions are not normal. In other words, the null hypothesis states that the distribution is due to chance.

In order to reject the null hypothesis in favor of the alternative hypothesis, the probability value of error (p) has to be smaller than the threshold for statistical significance alpha (α). The probability of error refers to the probability one is likely to err if one rejects the null hypothesis in favor of the alternative hypothesis given the observed data (Gries 2009: 32). By convention, α is set to 0.05 (cf. Bortz and Döring 2006: 500), which means that there is only a probability of five per cent that the observed distribution will occur, given the null hypothesis is accurate. Hence, if the p-value is below 0.05, the
distribution is most likely not normal (Gries 2009: 150). Other conventional levels for statistical significance are $\alpha = 0.01$ and $\alpha = 0.001$. If $p$-values are below the threshold of $\alpha = 0.01$, the null hypothesis is very unlikely or the observed variation is statistically very significant; if the $p$-values are below the threshold of $\alpha = 0.001$, the observed variation is highly significant (Gries 2009: 32). Since language is variable (see below), I set the significance level at $\alpha = 0.05$ in order to reject the null hypotheses in general. Despite the two disadvantages of the Shapiro-Wilk test, I chose it due to its above-mentioned power, compared to other tests for normality, such as the Kolmogorov-Smirnov test, and because it is implemented by default in the standard statistical programs SPSS (Statistical Package for the Social Sciences) and R (R Development Core Team 2006-13).

The role of outliers cannot be underestimated: the classification of data points as outliers is mostly arbitrary. Those which deviate strongly from the rest of the data points or those which are untypical are considered outliers. This definition is not very helpful when it comes to language, because it does not reveal anything about the manner to proceed about the outlying data points. Without a theory or good justification, they can be neither excluded nor included (cf. Caspary 2013: 41).

Classic examples to illustrate the problem are body height, body weight or average hours of sleep for human beings. Based on convention, it is more than apparent that a body height of three meters, a body weight of 250 kilograms and an average of 67 hours of sleep are clearly outliers in a data subset about average human health conditions, as they simply cannot occur. Thus, they seem to originate in measurement errors and need to be excluded from further analysis. As the example of body weight already implies, there are some cases in which humans do weigh more than 200 kilograms, but an exclusion of such data points would still be justified, because such humans are not part of the average healthy group of human beings.

Other data sets do not have such an obvious distribution of data points. In stock exchange, for example, it is the norm that within one day some stocks have huge gains and losses and others do not. In such a wide distribution, outliers are an inherent feature and can therefore not simply be excluded. It is thus the nature of stock exchange to have outliers. The same is true for language. Language is variable; it varies within one sentence (Chambers 2009: 13) and it may change in the course of two to four generations (Labov 1994: 44-45), which can ultimately lead to communal change (Labov 1994: 84).

In acoustic phonetic analysis, the dependent variables that should be normally distributed, ideally have homogeneous variances and no outliers are most often formants one and two – ratio-scaled variables. According to the sociophoneticians Labov, Baranowski

---

56 For other tests for normality see, for instance, Seier (2002).
57 This common sense definition of outliers is very simplistic. Despite very thorough attempts (cf. e.g. Caspary 2013: 40-42), there is no clear-cut definition of the term ‘outlier’, and scientifically sound methods for the determination of outliers do not exist (Caspary 2013: 41). Nevertheless, there are numerous statistical tests for the identification of outliers. The choice of test depends crucially on the estimated behavior of the sample and the conventional threshold of alpha (Caspary 2013: 41).
and Dinkin (2010), outliers may be indicators of language change and are thus an inherent part of language (also cf. Labov and Baranowski 2006). Other phoneticians such as Ladefoged (2003: 129) suggest to remove those data points from the set under analysis that behave untypical. This putative contradiction may originate from the theory underlying the conducted analysis. Whereas Ladefoged’s analysis attempts to establish a typical vowel space of a speaker of a certain language (possibly endangered by extinction), Labov, Baranowski and Dinkin’s concern is to model language change or the vowel space of speakers whose ongoing change in vowel positions naturally requires outlying data points in order to complete the change at some point in time. In such cases, after the faulty measurements in formants one and two have been discarded or re-measured (cf. Section 4.5), outliers can only originate from the social reality of the speakers (e.g. young upwardly mobile females) and/or the linguistic reality in which the vowels occur (e.g. dorsals adjacent to vowels increase the values of the second formants; cf. Subsection 4.4.5).

This interpretation of outliers is further supported by the fact that for some speakers in my data set, repeated removal of outliers resulted in new outliers with each repetition (see below). The results of the Shapiro-Wilk test changed accordingly: While it was impossible to arrive at a clearly normal distribution of first and second formant values per vowel for such speakers, inclusion of outliers helped appropriate such a distribution most closely. This behavior of data points in relation to one another seems to indicate that outliers are an inherent feature of such – if not all – speakers in my data set.

If a data set seems to consist mostly of outliers, both arithmetic mean and median may not be appropriate measures of central tendency, i.e. they may fail to provide a good summary of the data set (cf. Gries 2009: 106). Clearly, one extreme outlier in a data set tremendously changes the mean of that data set. Since the median is more robust in such cases (Gries 2009: 108), it may hence misrepresent the changing nature of the observations from a speaker. The mean thus provides the better measure of central tendency in my data, because it takes outliers into account, which I understand to be an inherent feature of my informants’ speech.

4.6.2.2 Homoscedasticity

If the Shapiro-Wilk tests suggested normally distributed data sets, I tested them for homoscedasticity (homogeneity of variances) via an $f$-test for two samples, Bartlett’s test for more than two samples and Peña and Slate’s global test (2003, 2006; cf. Subsection 4.6.2.4). Multiple statistical comparisons of the same samples have to be avoided, because the probability threshold that the observed variation is likely under the null hypothesis increases. If, for instance, the variances of the first formant values per age groups old, middle and young are to be compared, the threshold of the probability has to be defined carefully. In this case, a proper definition is that at least one out of three group variances is (statistically) significantly different.
Multiple separate $f$-tests of the age group variances with $\alpha = 0.05$ result in an increase of the probability of significance from $p = 0.05$ to $p = 0.143$, which is to say that there is a probability of 14%, instead of 5%, that at least one of the three variances will reach statistical significance just by chance (cf. Baayen 2008: 105). The alpha error inflation increases logarithmically with the number of statistical tests, so that there is a probability of 30%, instead of 5%, that one of the values of comparison will reach statistical significance when nine tests are conducted (cf. Bühner and Ziegler 2009: 325; Crawley 2007: 482). Thus, a false positive result can be obtained, as such testing has higher chances of producing significant $p$-values.

False positive results lead the analyst to reject the null hypothesis in favor of the alternative hypothesis in the analysis of the samples, although the null hypothesis is true in the population (real world). This is the most severe mistake that can happen in statistical analyses, usually referred to as Type I error (cf. Bortz and Döring 2006: 498; Johnson 2008: 54-55). For this reason, the influencing factors with more than two levels cannot be evaluated with statistical tests that assume two levels only such as a $t$-test or an $f$-test.

$F$-tests and Bartlett’s test are sensitive to a non-normal distribution of the samples and rely on a confidence interval of 95% by default in order to determine whether the variances of the two or more samples are similar (null hypothesis) or different (alternative hypothesis). The variance is the sum of all squared values of the data points after they have been subtracted from the mean. Since data point values are usually smaller (left of the mean) and larger (right of the mean) than the mean, subtraction will yield positive and negative numbers, so that their sum will be zero. To avoid this result, the subtracted values are squared.

The value of the variance is difficult to interpret in relation to the original non-squared values of the data set, so that the square root is taken. The result is referred to as standard deviation and is thus derived from the mean, stating that 68.2% of the data points fall within its range of -1 to 1 standard deviations (Oakes 1998: 8). If, for instance, the mean of first formant values of $\text{kit}$ is 523 Hz with a standard deviation of 34.5, then 68.2% of the values are between $523 - 34.5 = 488.5$ Hz and $523 + 34.5 = 557.5$ Hz. If the standard deviation was smaller, but the mean value stayed the same, the values of the data set would vary much less around their mean. For this reason, measures of central tendency (e.g. mode, median and mean) should not be reported without their corresponding measure of dispersion of the data set (relative entropy, interquartile range or quantiles and the standard deviation or variance; cf. Gries 2009: 111-112).

The choice of the measure of central tendency depends on the variable’s level of measurement: the mode (which data point values occur most often in a data set) should be used for nominal/categorical variables (e.g. linguistic context: voiced versus unvoiced [nominal] or labial versus dorsal versus coronal, etc. [categorical]; cf. Gries 2009: 15), the median for ordinal variables (e.g. age group: young speakers are ranked as low,
middle-aged ones as middle and old ones as high; cf. Gries 2009: 16) and the mean for interval/ratio variables (e.g. formant values: an F1 of 300 Hz for kit is different from one of 600 Hz [the nominal information], the first value is smaller than the second [the ordinal information] and the second value is twice as high as the first; cf. Gries 2009: 16, 106).

4.6.2.3 Independence of Observations

T-tests compare the means of two samples and analyses of variance (ANOVAs) compare variances of more than two samples. The most important assumption of two-sample t-tests and ANOVAs is independence of the data points in the samples. The data points are independent of one another, if each data point is produced by a different subject (Gravetter and Wallnau 2007: 604). This is due to the fact that the intra-speaker variation in one vowel such as dress may range from first formant values of log-normalized (or ANAE normalized) 500 Hz to 700 Hz, whereas another speaker’s internal variation may range from 400 Hz to 600 Hz for the same vowel. This means that the observations of first formant values “all from one [speaker] are often more highly correlated with each other than they are with observations of other people”, so that “individual observations of one speaker are not independent of each other” (Johnson 2008: 122). If the selected statistical procedure assumes independence of observations in the samples and independence is violated, the result is more likely to be statistically significant than when independence is not violated (Saito 1999: 453-454), i.e. a Type I error is very likely.

A t-test may, however, also be used for dependent samples, if the observations can be paired (Gries 2009: 212). One example for which such a test can be used is the difference in first and second formant values of the low back vowels in minimal pairs (in word lists) such as hod and hawed, and bot and bought, respectively (cf. Bigham, White-Sustaita and Hinrichs 2009). This poses the question of whether it is possible to find similar pairs in free speech, i.e. whether it is valid to assume that, for instance, two LOT vowels can be considered as a pair, if they occur in relative vicinity to each other. Such contemplations are also supported by results from studies of persistence phenomena which suggest that the variant used in the preceding instance may have a positive (variant re-occurs) or negative (variant does not re-occur) effect on the realization in the succeeding instance (cf. e.g. Bernolet, Hardhuiker and Pickering 2009; Gries 2005; Potter and Lombardi 1998; Scherre and Naro 1991). The question then revolves around the fact of when to consider two tokens of the same speaker to be dependent, i.e. is it, for example, the short time that passes between the two tokens, is it the same topic under which the tokens occur or is it rather the same lexical item in which two tokens occur.

59 The same is true for paired ANOVAs. They are usually referred to as repeated measures ANOVAs.
Since such theories have not been established to an extent that would have justified using them in my analyses, I decided to satisfy the assumption of independence by using one observation per speaker. This does not mean that the analysis should be restricted to literally one token per vowel and speaker, but to at least three to five tokens per independent variable (linguistic cotext, style, age, sex, social class, etc.) from which the mean (per variable and speaker) is taken to be the best summary of the quality of each of the vowels under analysis. The mean values of KIT, DRESS, TRAP, LOT and THOUGHT per speaker thus satisfy the independence of observations, so that t-tests for independent samples and ANOVAs can be employed (cf. Gries 2009: 205).

Concerning LOT-THOUGHT in particular, Gorman and Johnson (2013: 230) note “that unpaired t-tests [for independent samples] are a poor tool for quantifying merger, since the tests frequently [assign] a significant result for vowel class even to [those] merged in production”. They concur that this may be due to ignoring to role of phonological context (Gorman and Johnson 2013: 230), because t-tests only compare two means of ratio-scaled variables. According to Labov, Ash and Boberg (2006b: 59), the low-back merger is favored by LOT-THOUGHT tokens before nasals (essentially /n/) and subsequently progresses to other consonants.

If the low-back merger is in the early stages of change in a speech community, it would sound reasonable to first calculate arithmetic means per following and/or preceding phonological context and employ t-tests for LOT and THOUGHT before nasals first, rather than lump the means for all phonological environments together and doubt the suitability of t-tests (for multiple comparison of means a combination of ANOVAs with post-hoc tests such as Tukey’s HSD is necessary). As a consequence, t-tests should be a ‘richer tool’ for speech communities in which the merger has long been established, as is the case in St. John’s English (e.g. Kirwin 1993: 74). In fact, t-tests of speaker means in all age groups yielded insignificant results (cf. Subsection 5.2.1).

### 4.6.2.4 Additional Assumptions of General Linear Models

In addition to independence of observations, ANOVAs assume linearity (specified by a link function; see below) which is given if, for instance, all of the data points in a two-dimensional coordinate system appear in the shape of an ellipsis, in which low values on the x-axis correspond to low values on the y-axis and vice versa. This means that the dependent variable is continuous and a change in values in the independent variable correlates with a change in values in the dependent variable (e.g. the more y, the more x). In such a distribution of data points, a linear function can be superimposed. The vertical distance of each data point in the set to the estimated data point on the regression line (linear function) in the graph is referred to as a ‘residual’ (Gries 2009: 143). In addition to

---

60 The null hypothesis states that there is no difference in the first and second formant means between LOT and THOUGHT.
homogeneous variances in the groups in the samples, these residuals need to be distributed
normally in order for the ANOVA to produce reliable results (cf. Wickens 2004: 3, 8).

Although most of these assumptions can be best investigated in graphs such as scatterplots, histograms, plots of residuals versus fitted values or normal Quantile-Quantile plots (cf. Gries 2009: 282), and although plotting the data is usually highly recommended as a first step (cf. Baayen 2008: 97; Gries 2009: 276), plot interpretation is subjective, the plots can be quite misleading and it is unclear how the combination of violated assumptions could affect the resulting plot, as usually one graph is plotted to visualize and assess one assumption (Peña and Slate 2003: 5). I thus minimize the multitude of possible graphical representations of the data distribution to one – boxplots – and use the appropriate tests in order to assess the status of the assumptions of ANOVAs in particular and linear models as well as statistical tests in general.

The assumptions of linear models are violated (a) if the samples are skewed, (b) if they deviate from the normal distribution kurtosis of the error distribution, (c) if a mis-specified link function is used (i.e. non-linear behavior of the samples, as the residuals are non-normally distributed) or other predictor variables are absent (i.e. the residuals would be normally distributed if the predictor or independent variable accounted for more data points) and (d) if heteroscedastic errors (the variances of the samples are not homogeneous) and/or dependent errors are present (Peña and Slate 2003: 11).

In order to assess the four assumptions of linear models, statistical testing is required. This, however, poses a similar problem as outlined above with regard to the Type I error probability when the results of the four statistical tests are combined. Peña and Slate (2003, 2006) thus propose a global test that controls the Type I error rate, is based on residual vectors and possesses local optimality properties (Peña and Slate 2003: 7). In addition, its components can be used as a directional test in order to determine the assumptions that have been violated (Peña and Slate 2003: 8). According to Gries (2009: 283), ANOVAs are relatively robust, so that they may be able to yield accurate results when it comes to violations of their assumptions. However, this is not true for violating the assumption of independence of observations in the samples (Johnson 2008: 112), and the extent of robustness is hard to gauge, especially with regard to normality of the dependent variable (Quinn and Keough 2002: 359).

Another type of linear model are multivariate analyses of variance (MANOVAs), which make the same assumptions as outlined above, but additional considerations have to be made when the influence of one independent variable (predictor) on two dependent variables is assessed simultaneously (e.g. first and second formants). The correlation between the dependent variables has to be assessed prior to employing the MANOVA. When no outliers are present and the data are normally distributed, Pearson’s correlation coefficient $r$ can be used to determine the correlation. It should be within the range of $r = 0.3$ and $r = 0.9$ for positive correlations (e.g. the higher F1, the higher F2) and not more than $r = -0.4$ for negative correlations (cf. Mayers 2013: 323). When outliers are
present or the assumption of normality is violated, Kendall’s $\tau$ (tau) is a better choice, as it is based only on the ranks of the variable values (Gries 2009: 144).

Higher correlations would obscure the need for a MANOVA as the dependent variables would ultimately measure the same concept. Likewise, a lower correlation would possibly neutralize a multivariate effect (Mayers 2013: 323). Additionally, the univariate between-group variance has to be equal. If this assumption is violated, unequal sample sizes may negatively influence the outcome of MANOVAs. The homogeneity of multivariate variance-covariance matrices can be assessed with Box’s M test in SPSS. If the result of this test is insignificant, the variances between the samples and the correlation between the dependent variables does not differ significantly. Otherwise a MANOVA should not be used, as its results are unreliable (cf. Mayers 2013: 323).

### 4.6.2.5 Assumptions of Generalized Linear Mixed Models

The statistical evaluation of my data is two-part: First, I test the data based on speaker mean values of formants one and two (independence) per age group. Second, I triangulate and extend these results with those of linear and logistic regression. Generalized Linear Models (GLMs) were originally developed by Nelder and Wedderburn (1972) and developed further by McCullagh and Nelder (1989) in order to unify several statistical tests, which relate the dependent variable to a linear combination (sums) of independent variables via a link function (Jackman n.y.: 1; Quinn and Keough 2002: 360). The link function transforms the expectation ($\mu$) of the dependent variable ($Y$), $\mu \equiv E(Y)$, to a linear predictor ($\eta$; Fox 2008: 379) and depends on the distribution of the dependent variable, which was originally assumed to fall within the exponential family of distributions (Hedeker 2005: 729). Examples of members of the exponential family are Gaussian (normal), binomial, or Poisson families of distributions (cf. Fox 2008: 381).

One common link function ($g(\cdot)$) is the identity link, which simply returns its argument unaltered ($\eta = g(\mu) = \mu$; Fox 2008: 380) and models the mean or expected value of $Y$ ($\mu$; cf. Crawley 2007: 514; Quinn and Keough 2002: 360). It is used to synthesize standard linear models for continuous dependent variables (cf. Fox 2008: 379; Jackman n.y.: 1; e.g. first and second formant values in linear regression). Another common link function is the logit link ($g(\mu) = log\frac{\mu}{1-\mu}$) which is used for binary data (e.g. smaller or larger than a certain first/second formant value) and synthesizes logistic regression (Quinn and Keough 2002: 360). The other two components of GLMs are the random or stochastic component, specifying the dependent variable and its probability distribution, and the linear predictor, which is the linear function of independent variables (or predictors) in the model (cf. Crawley 2007: 512-513). Since a probability distribution is specified for the dependent variable (and therefore for the errors/residuals of the model), GLMs originally were parametric (Quinn and Keough 2002: 360). According to Crawley (2007: 511), GLMs excel at dealing with non-constant variance and non-normal distribution of errors.
Extensions of the GLMs after their original formulation allow also for inclusion of unknown distributions of the dependent variable: The quasi-likelihood estimation method allows for much more flexibility, because it estimates the dispersion from the data rather than constraining it to the value implied by a specific distribution (Quinn and Keough 2002: 360). The method uses only the variance-mean relationship of the dependent variable, which is usually sufficient to estimate as efficiently as Maximum Likelihood (ML; Crawley 2007: 517). Maximum Quasi-Likelihood models are also useful when the dependent variable has a binomial distribution (for logistic regression) and a variance greater or less than expected from the mean. Logistic regression does not make assumptions such as normality in terms of distribution of the dependent variable (Hedeker 2005: 729).

Predictors may be continuous (as in linear models) and/or categorical (as in ANOVAs) and may also include transformations of the independent variable, polynomial functions and interactions (Fox 2008: 379). Interactions between predictors have to be more or less anticipated based on logic or a priori hypotheses. If, for instance, age and education are entered as predictors in a model with formant (one or two) as dependent variable, they will most likely interact. The validation behind entering these three variables lies in the fact that young and highly educated speakers produce other vowel qualities than old and uneducated speakers. However, there are no very young speakers that have completed school, there are also no young speakers that have completed university, etc., so that in a cross-tabulation of age and education the cell values of ‘secondary/tertiary education completed’ and ‘15 to 20 years of age’ will be zero.

In my analyses this problem applies especially to linguistic environment. With regard to, for instance, following phonological segment, independent variables (factor groups or predictors) that influence the vowel quality are glottal state, place and manner of articulation. In a cross-tabulation of nasals and voicing, all of the cells for voiceless labial, alveolar and velar nasal will be empty, as these sounds only occur in their voiced variants in English (i.e. collinearity; cf. Tagliamonte 2006: 230 for an example from morphosyntax; also cf. Subsection 4.6.3.4).

According to Labov (2001a: 84), correlated variants usually occur within internal and external variables, but not across them. This is a rather strong claim, especially when based on experiences with a software that does not particularly support discovering and handling interactions (cf. Subsection 4.6.3.3). In Paolillo’s (2002: 66) illustration of some possible interactions, he emphasizes correlations between social and linguistic variables, especially when two different locales are compared. Within internal variables, the constraints for crossing them are not necessarily the same, as, for instance, there are no velar nasals in preceding consonant position of (non-emphatically) stressed vowels in

\[^{61}\text{Collinearity means that variants of two independent variables are highly correlated and can thus not be modeled independently (in two different independent variables, e.g. voicing and manner of articulation; cf. Chatterjee and Hadi 2006: 88, 114; Guy 1993: 242; Mendoza-Denton, Hay and Jannedy 2003: 117). Collinearity may also occur with more than two variables (multicollinearity; cf. Chatterjee and Hadi 2006: 95; Sigley 2003: 242).}\]
English, but in following consonant position. Crossing of linguistic factors in preceding environments does thus not follow the same rules as in following environments.

In the same fashion as empty cells, a low token number in cells\(^{62}\) and badly distributed data may pose problems to regression analyses. Distributional issues can arise when, for instance, 300 vowels from females in pre-nasal position versus 11 vowels from males in pre-nasal position were collected, or when females and males behave differently with regard to another independent variable. If possible interactions are disregarded, their effect may be a false significant finding (a Type I error) for one or both independent variables. Cross-tabulations and/or scatterplots are usually used to spot unknown or disregarded interactions (cf. Quinn and Keough 2002: 127; Tagliamonte 2006: 229).

There are several ways to take interactions into account, and each may have potential problems itself. First, it is easiest to enter the interaction as a predictor into the model in order to assess whether it is significant along with the other predictors and along with other interactions that may be more prominent (cf. Baayen 2008: 251). However, if many interactions have to be accounted for, this results in a more complex model, as more predictors are present, which might lead again to false positive results. Alternatively, two suspicious factor groups can be crossed, resulting in an ‘interaction group’ (cf. Sigley 2003: 242; also cf. Subsection 4.6.3.4). Second, the factors within the factor groups (similar to the variants of categorical independent variables) can be subsumed under one new factor, so that the cells will yield larger values. This, however, poses the problem that effects between two factors are deleted, once the factors are merged based on a priori theories (instead of, for instance, Bonferroni-adjusted multiple testing of mean values). Third, if one of two cells is empty or has very few tokens, both can be deleted, resulting in fewer tokens to analyze in total. Tagliamonte (2012: 136) suggests removal of low token numbers in cells for reliability reasons.

Which of these possibilities is ultimately used depends largely on the goal of regression. It may be inadvisable to rather blindly trim data in the way suggested above, especially when one of the two cells from which tokens are deleted contains a large number of tokens. Deleted tokens may also have indicated a tendency or direction of linguistic change (for social variables) that may be lost after data trimming (cf. Baayen 2008: 244; Chatterjee and Hadi 2006: 108). Likewise, the differences between two factors may have been of particular, but not primary, interest (e.g. Are vowels before nasals at a more advanced stage in an ongoing shift?), so that recoding them eradicates this difference. In such cases, adding the interaction as a predictor to the model may be the better alternative.

A further assumption of GLMs, which they share with linear models, is the independence of the data points entered as dependent variable in the model. Unlike outlined in the subsection above, this assumption cannot be met by using speaker means, as modeling

---

\(^{62}\) Many basic statistical tests such as the chi-square test for countable data assume expected frequencies of at least five (cf. e.g. Lomax and Hahs-Vaughn 2012: 221; Heiman 2013: 356). For regression, ten is a widely used threshold (Cedergren and Sankoff 1974; Guy 1993).
linguistic context along with social variables would thus be impossible (one observation per speaker equals one linguistic context). Further extension of GLMs allowed inclusion of random intercepts in addition to the predictors. Models including the option of evaluating fixed effects and random effects at the same time are referred to as Generalized Linear Mixed-Effects Models (GLMMs). These random effects account for the correlation of the data within the subjects (cf. Hedeker 2005: 729 for longitudinal study designs and Bergsma, Croon and Hagenaars 2009: 232 for a sociological example).

In that sense, my study design corresponds to a clustered one: on the first level are the observations per speaker (e.g. all the LOT vowels of one speaker), and on the second level are the speakers themselves. The first level is nested within the second level: The vowels of one speaker may correlate in their formant values more highly than between two speakers, because normalization merely minimizes physiological differences (cf. Subsection 4.5.2). Since the model calculates an intercept based on all vowels produced by all speakers, it may find a significant deviance from this intercept, which can be thought of as a kind of baseline mean (Baayen 2008: 244), in those cases where the formant values of one speaker cluster differently around/further away from the intercept than those of another speaker. As a result, a falsely significant finding might be assessed by the model (Type I error).

Random effects (or random intercepts) allow each speaker to have a slightly different intercept, so that significances based on correlations of formant two values within one speaker are disregarded (cf. Baayen 2008: 245 for an example from psycholinguistics), i.e. false positive results, especially when near the 0.05 threshold, are controlled for. The confidence intervals for the intercepts of the random (and fixed) effects are estimated by a common method referred to as Restricted/Relativized/Residual Maximum Likelihood (REML; Baayen 2008: 246). The estimated intervals (for any statistical test) indicate the true population intercept boundaries with a default probability of 95% that could have generated the intercepts found in the analyzed sample (Gries 2009: 125).

Put differently, I am not interested in the linguistic behavior of exactly those 34 speakers I analyze in order to compare their vowel qualities, but in the population which they are supposed to represent (i.e. from which they are sampled) – middle-class St. John’s English. If speaker-specific effects on formant values are disregarded, the result represents the linguistic behavior of the population speakers are sampled from. Social variables such as age and sex are then fixed effects in the model, because I am particularly interested in how young female speech differs from old male speech. Due to the possible correlation of formant values nested in lexical items, the latter should be regarded as random effect as well, especially in terms of St. John’s, Newfoundland, because vowel

---

63 Reading minimal pairs in a word list may then be thought of as an additional level hierarchically below the individual vowel tokens. The minimal pair vowel are then nested within the higher-level vowels, which are in turn nested within the higher-level speakers (cf. Hedeker 2005: 729).
4. Data and Methodology

quality has been reported to be distinct in some lexical items (e.g. raised and fronted ash in *catch*; Clarke 2010a: 30).

In the same fashion as the variants (factors) of independent variables (the fixed effects) may interact (collinearity/correlation; cf. Footnote 61), a fixed effect may correlate with a random effect: Phoneme label (the fixed effect; cf. Gorman and Johnson 2013: 232) may interact with speaker in terms of a possible merger, i.e. one speaker may produce lower second formant values for phoneme /o/ (*LOT*) and higher second formant values for /oh/ (*THOUGHT*), whereas another may produce second formant values the other way around. In a scatterplot of second formant values and phoneme label (in the order /o/ and /oh/) per speaker, the regression line will increase in the former case (positive slope) and decrease in the latter case (negative slope). In such cases, a random slope of phoneme label and speaker should be considered in the model in order to allow different slopes per speaker, and thus control for a possibly false positive result (cf. Baayen 2008: 248).

As indicated above, regression analysis is a trade-off between simple models that cannot explain the variation satisfactorily and complex models that capture richer conceptual pictures, but might be inaccurate in their estimates or hard to replicate (Wickens 2004: 1). Simple models with omitted important predictors (underfitted models) are reduced in their predictive power and bias estimates of effects for included predictors (Chatterjee and Simonoff 2013: 24). Complex models with unnecessary predictors (overfitted models) decrease in the accuracy of their predictions as well, but because of the additional unnecessary noise created by the those predictors; they are less likely to remain stable and are thus also less likely to be useful for future predictions than are simpler ones (Chatterjee and Simonoff 2013: 23). In addition, complex models can be computationally awkward or intractable, which is today often considered an issue of the past. Many previously impossible computations of algorithms or tests can be (nearly) routine today such as iterative algorithms, Monte Carlo methods and Bayesian approaches (Quinn and Keough 2002: xv; Wickens 2004: 1).

SPSS, for instance, takes all possible interactions for all fixed effects into account by default and attempts to yield respective *p*-values. This becomes problematic, however, when random effects and slopes are entered into the model, as it may result in ten to twelve hours of computation time per run (and at times) without a converged model finding

---

64 Monte Carlo methods are especially useful for small data sets in order to determine the upper bound for degrees of freedom in regression analysis (Baayen 2008: 248). Although ‘small’ is rarely explicitly translated into numbers, some authors suggest a threshold of 20 for number of subjects (e.g. Bergsma, Croon and Hagenaars 2009: 234) while others suggest one of 30 (e.g. Mannila, Nevalainen and Raumolin-Brumberg 2013: 357). Since my number of respondents exceeds both thresholds (*n* = 34), I will not use Markov chain Monte Carlo sampling for the estimation of *p*-values and confidence intervals (Highest Posterior Density intervals).

65 All possible interaction means including higher level interactions than just two, e.g. phoneme label*sex*place*manner*voice, etc.

66 On an average desktop computer with a quad-core processor of 3 GHz and 4 GB RAM. E.g. in one of my trial models, the two random effects and one random slope together were responsible for 314 out of 417 levels of the design matrix the model had to compute, whereas the 23 fixed effects and
(due to obvious collinearity). In other cases, after 10,000 iterations the software crashed, so that computational power still plays a role in such statistical modeling, especially with phonetic data (collinearity in simultaneous modeling of preceding and following consonant sounds and a random slope for speaker). The high number of iterations indicates that such statistical modeling involves multiple testing of the exact same variables, which inflates the alpha level for every factor group (cf. Subsection 4.6.2.2).

Since interactions are modeled together with the main factors (e.g. phoneme label, sex and phoneme label*sex), the factors are tested multiply (at least twice). In order to control for Type I errors (false positive results due to alpha error inflation), the Bonferroni approximation or the Šidák equation should be used in such modeling (Abdi 2007: 106; Seltman 2013: 327). The former is more common and by default implemented in SPSS and R (R Development Core Team 2006-13).

### 4.6.3 Selection of Statistical Tests

The selection of the tests for the analyses conducted in this study follows two main considerations: First, the assumptions of the tests are not or only justifiably violated, and second, the comparability to previous acoustic phonetic studies on St. John’s English or on Newfoundland such as Boberg (2010), Hollett (2006, 2007) and Labov, Ash and Boberg (2006b) is given. None of these studies provides results of statistical testing as to whether or not the assumptions of the employed tests are violated. This suggests two possible scenarios: Either the assumptions are not violated and it is thus taken for granted that the statistical tests are appropriate, or violation of assumptions was simply not tested.

Since the main research question of this study seeks to find an answer to whether the ongoing change on the Canadian mainland in the form of the Canadian Shift is also taking place in middle-class speakers of St. John’s, Newfoundland, age should be the strongest predictor in accounting for the shifts of TRAP, DRESS and KIT (in this order), after the merger of LOT and THOUGHT is in place (cf. Section 2.3). For this reason, I use Decision Trees with all the independent variables to assess whether age is the first predictor according to which nodes (the first decision) are found.

Boberg (2005: 141) and (2010: 218) uses multivariate analyses of covariance (MANCOVAs) to assess which of the social predictors has an effect on the formant values of all vowels that are understood to be part of the Canadian Shift. Boberg does, however, not state whether his data violates assumptions of the linear model underlying the MANCOVA: normality and homoscedasticity. The theory of the Canadian Shift states for TRAP and LOT that they primarily move in second formant space, but not necessarily in first formant space (at least in case of the former vowel). This means that their variances

---

their interactions were only responsible for the remaining 103 out of 417 levels. In the final analyses, I attempted to run simpler models whenever possible (with varying success).
in formant one should be very small compared to the variances in formant two. It is thus rather unlikely that the variances are homogeneous in those two cases.

In addition, the MANCOVA compares variances for all vowels per age group to assess whether there is a significant difference in variances for all vowels of one age group compared to those of the other two age groups. I am not certain whether a total variance of all vowel formant values is valid for the data set, as in theory these values do not belong to the same category (vowel class). In order to maintain the independence of samples, Boberg entered speaker means into the model, but hence only has a small number of observations (max. n = 35, Boberg 2005: 138, and n = 86, Boberg 2010: 144), which may render the MANCOVA results unreliable. Especially when the assumption of normality is violated, ANOVAs in general require at least 20 observations per variant of the category under investigation (e.g. age groups; cf. Kleinbaum et al. 2014: 485). For these reasons, I decided against using MANCOVAs as the primary source of my results. In addition, Boberg does not enter linguistic context and duration in the MANCOVA which are known to influence vowel formants greatly. This becomes problematic with regard to, for instance, the effect of /l/ following KIT and DRESS, as it is known to lower second formant values tremendously (cf. e.g. Di Paolo, Yaeger-Dror and Wassink 2011: 87; Thomas 2001: 50). In his Montreal article, he explicitly states that vowels before /l/ are included in his sample (Boberg 2005: 138-139).

Labov, Ash and Boberg (2006b: 146) used Principal Component Analysis in order to find regional differences in 21 selected mean formant values of all 439 acoustically analyzed speakers of North American English. The underlying idea of principal component analysis can be outlined in a simplified example of a data set that has three dimensions. In a cube with three axes, x, y and z, each single data point in a data set can be described by three values, one on each of the axes. A Principal Component Analysis attempts to reduce the number of axes or dimensions and checks whether the data points can still be approximately described using only two dimensions, possibly new ones. The analysis achieves this by rotating the axes in such a way that, for instance, two new axes can be found in the diagonal plane of the original, unrotated, axes (cf. Baayen 2008: 120 for more details).

Rotation is possible, if the data points in a cube are all located on a plane that extends gradually from large values on the z-axis and from small values on the y-axis to small values on the z-axis and to large values on the y-axes. The values on the x-axis range from small to large and form the width of that plane. When the values remain fixed in their location, the cube is then rotated in such a way that all of the data points are lying on the bottom of the cube. In this way, they can all be located within only two dimensions: the x- and y-axis (cf. Baayen 2008: 120 for more details). Principal Component Analysis will consequently choose the dimension with the most variation as its first axis and rename it

---

67 In this article, Boberg’s results suggest that KIT and DRESS are significantly retracting, but not lowering.
to principal component one. The dimension that it did not need to locate the data points is the dimension that does not account for any variability in the data (cf. Baayen 2008: 120). With real data sets, the analysis is complex, as the data points are rarely all on one plane and as there are usually much more than three dimensions, so that calculation takes a long time.

The problem with principal component analysis is that the variables in the model should be normally distributed and have no outliers, and that columns with the greatest ranges dominate the results, if the variables’ ranges differ greatly (Baayen 2008: 125). In the data of Labov, Ash and Boberg (2006b: 146), the variables first and second formant have very different ranges (the vocal tract’s width is always smaller than its length), and the numerical code for region has an even smaller range than the two formant values.

For this reason, I decided against using Principal Component Analysis to explore the relationship between age and formant values. I employ Decision Trees instead, as they do not make any of the above-mentioned assumptions about the data, but serve the same purpose as Principal Component Analysis (cf. Section 4.6.1). With the results confirming that age has a strong predictive effect on the behavior of the formant values, I employ statistics with age as the only social predictor variable and first and second formants as the continuous dependent variable, respectively. Figure 4.21 summarizes the choice of statistical tests based on their assumptions and the variables that ought to be tested.

I triangulate these statistical methods with linear and subsequently logistic regression using random effects/slopes, so that, first, assumptions possibly violated by the other tests are compensated for (e.g. normality), and second, more than one independent variable can be modeled simultaneously. Although these regression types only consider one dependent variable at a time, their advantage to MANCOVAs is the inclusion of random effects or random slopes. Which dependent variable I use for regression is based on the results of the statistical assessment of the correlation coefficients (cf. below and Subsection 4.6.3.2) and the established theory behind the Canadian Shift: For the vowels LOT and THOUGHT, I use second formant values, based on their more robust separation of the two vowels classes (Gorman and Johnson 2013: 232); for TRAP I use second formant values, as it is agreed to predominantly retract; for KIT and DRESS the issue will be evaluated in the respective Sections, 5.4.1.2 and 5.5.1.

The independent variables are comprised of internal or linguistic and external or social variables. Although none of the above-mentioned studies explicitly state in which fashion they explored the effects of linguistic environment, I employed Decision Trees again in order to find out whether the effect of the preceding and following phonological contexts results in clusters in the respective vowels’ formant means. Depending on the results, the contexts can be subsumed in groups according to either place or manner of articulation for linear and logistic regression. First and second formants were entered separately as the

---

68 I excluded the illustrations of MANCOVAs with their assumptions and the combination of variables they handle, because I do not use them in my analyses.
Data and Methodology

Number of dependent variables?
- paired t-test
- Wilcoxon paired samples test
- two sample t-test
- Mann-Whitney U-test
- Friedman test
- Kruskal-Wallis
- ANCOVA
- MANOVA

Type of dep. var.?
- continuous
- discrete
- binary

Type of indep. var.?
- discrete
- continuous
- discr./cont.

Does homoscedasticity and normality hold?
- yes
- no

Does independence of samples hold?
- yes
- no

Number of samples?
- one
- two

Number of indep. var.?
- one
- multiple

Does independence of samples hold? (linear regression w/ random effects)
- yes
- no

Figure 4.21: Flow chart of the statistical tests chosen for the analyses of this study. After assumption testing, the nature of the required variables and the type of variables are determined. Based on these settings and assumptions, the appropriate statistical test is to be selected.

For all continuous variables, the chart aids in deciding which statistical test to use in comparison of arithmetic means. I included only those combinations of dependent and independent variables that my data set is comprised of. Details and explanations are provided in the text.
4.6. Data Analysis and Statistical Modeling

dependent variables and the means per phonological context were assessed as to whether they significantly differ from each other. If they do not differ significantly, they can be subsumed in one group. The results of the Decision Trees showed that neither of the phonological contexts clustered, so that there was essentially one left – obstruents. For the regression analyses, this result indicates that linguistic context does not need to be entered into the model, as it seems to be controlled for. Nevertheless, I enter it separated by groups of glottal state and of place/manner of articulation (except for the analysis of TRAP), somewhat similar to Labov, Ash and Boberg (2006b: 220) and to a lesser degree to Clarke, Elms and Youssef (1995: 214). The results of the Decision Trees indicate that the output of the models should attribute an insignificant effect to these factor groups.

In order to transform the continuous variables, first and second formants, into categorical variables, I employ Optimal Binning (for age group), as implemented by SPSS, which reveals those mean values of the first and second formants per vowel (independent variables) at which the dependent variable age group, with its factors old, middle and young, behaves significantly different. Thresholds for the first and second formants were thus derived, which in turn serve as category boundaries for assessment of lowering/retraction of KIT, DRESS and retraction of TRAP in the logistic regressions. I determined whether KIT and DRESS are retracting and/or lowering via significance tests of coefficients of correlation between age (continuously) and each of the formant values separately. These tests triangulate the results from those tests in which age is modeled in groups (e.g. $t$-test and/or ANOVA results).

4.6.3.1 Merged Vowels

Mergers can be determined in numerous ways: via the Euclidean (or Cartesian, or Pythagorean) Distances between, for instance, LOT and THOUGHT as was done by Labov, Ash and Boberg (2006b: 62), or statistically via a MANOVA, linear regression or a $t$-test of two independent samples (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 102; Gorman and Johnson 2013: 230-234; Hay, Warren and Drager 2006). The Euclidean Distance (ED) is the length in Hertz between one vowel’s mean formant value and another vowel’s mean formant value in a two-dimensional coordinate system (here: an F1xF2 plot). The length is taken to be the hypotenuse (side $c$) of a rectangular triangle and is calculated – based on the Pythagorean theorem $c^2 = a^2 + b^2$ – via the square root of the sum of the squared values of the catheti (sides $a$ and $b$): $ED = \sqrt{a^2 + b^2}$. The lengths of the catheti are known when the difference between start and end points is known – the vowels’ (mean) first and second formant values (cf. Di Paolo, Yaeger-Dror and Wassink 2011: 101):

$$ED_{LOT-THOUGHT} = \sqrt{(F2_{LOT} - F2_{THOUGHT})^2 + (F1_{LOT} - F1_{THOUGHT})^2}.$$  

Labov, Ash and Boberg (2006b: 59) calculated the EDs for each of five minimal LOT/THOUGHT pairs, grouped the speakers per region and compared the regional dis-
stances in log-normalized Hertz (or ANAE Hertz) with each other. The ED metric thus corroborated the impressions of the analysts, that the vowels were either merged, namely when the EDs were very small (approximately 0 - 100 ANAE Hz), or that the vowels belonged to two phonemes, namely when the EDs were greater in value (e.g. 240 ANAE Hz; cf. Labov, Ash and Boberg 2006b: 62). The (means of the) EDs per speaker can then be statistically assessed via, for instance, a two sample t-test, ANOVAs or linear regression.

Since I am only interested in one geographic region, I can only calculate the EDs between LOT and THOUGHT for that region and decide subjectively whether or not the value is small or large. The only possible analogy to Labov, Ash and Boberg’s procedure would be a comparison of the EDs per age group, as this is the strongest predictor in apparent-time change. The EDs for my young speakers range from 30 ANAE Hz to 70 ANAE Hz and those of old speakers from 90 ANAE Hz to 250 ANAE Hz. I have, however, no a priori hypothesis that older speakers of St. John’s are more or less homogeneously non-merged (supported by the greater range in EDs of older speakers) in the two low back vowels, so that an ANOVA of age group and EDs does not necessarily determine whether or not one of the age groups is merged.

The earliest attestation of the low-back merger in St. John’s was in the 1970s when Scargill and Warkentyne (1972: 64) found that the merger was at a rate of 70%. According to Kirwin (1993: 74), the merger is complete on the Avalon peninsula (including St. John’s). Other traditional and more recent research did not question this finding (cf. Boberg 2010; Clarke 1985a,b, 1991, 2004b, 2008b, 2010a,b, 2012; Hollett 2006, 2007), except for Labov, Ash and Boberg (2006b: 60). They claim that the low-back merger is only in place before nasals, but it is transitional, as the pronunciation of LOT and THOUGHT was classified as ‘same’ by the respondents (2006b: 61), but as ‘close’ by the analysts (2006b: 63). It has to be kept in mind, however, that Labov, Ash and Boberg’s findings with regard to St. John’s are based on no more than two speakers (2006b: 217). I thus consider it scientifically sounder to compare the speaker means of the first formants for LOT and THOUGHT and the second formants for both vowels per age group with t-tests for independent samples. If the result is statistically insignificant, the two means per formant and age group are understood to be identical – the vowels LOT and THOUGHT are merged in formants one and two.

One of the problems of determining mergers with two sample t-tests was outlined in Subsection 4.6.2.3. The tests frequently return a significant result between the two means (i.e. the means are not identical), although the vowels were perceived to be merged (Gorman and Johnson 2013: 230). If Labov, Ash and Boberg were right to claim that the merger is only completed before nasals in St. John’s, the tests should yield a significant result when I lump all of the following phonological contexts together.

The results of the Shapiro-Wilk tests for normality of LOT and THOUGHT per speaker can be found in Appendices B.1 and B.2.
Another problem of using \( t \)-tests in that way is that they can only be used for one demographic predictor (here: speakers’ age group), in order to minimize the possibility of a Type I error (cf. Subsection 4.6.2.2). In many cases, the data is, however, unbalanced according to other predictors (Gorman and Johnson 2013: 231) that may favor or disfavor the merger (e.g. style, social class, linguistic context), especially when it is in a transitional phase. Other statistical tests such as linear regression and analyses of covariance (ANCOVAs) are more suitable in such cases, because they account for multiple predictors. For this reason, I triangulate the \( t \)-tests with linear and logistic regression (cf. Subsection 5.2.4), focusing on the second formant as dependent variable because of its more robust separation of the the two vowel classes. In addition, \( t \)-tests may be problematic, because only one of the dependent variables (first and second formants are treated separately) is considered at a time (Gorman and Johnson 2013: 232).

Two possible ways to account for both dependent variables at the same time are either EDs or multivariate analyses of variance (MANOVAs). EDs are to some extent problematic themselves as they take the relative contributions of first and second formants to be equal. This is particularly undesirable when one of the acoustic measures has a larger range or different variance (Gorman and Johnson 2013: 232). Another problem of EDs is that they do not take the correlation between first and second formant into account, so that “any correlative structure will be artificially inflated when they are combined in this fashion” (Gorman and Johnson 2013: 233).

Multivariate analyses of variance do take the correlative nature of two dependent variables into account. One study that employed MANOVAs to each speaker with phoneme as predictor and first and second formants as dependent variables was, for instance, that of Hay, Warren and Drager (2006). The outcome of the MANOVA they reported for each speaker was Pillai’s trace. This method was adopted by Hall-Lew (2009, 2010) in her study of mergers and vowel fronting. However, their work seems problematic: First, they calculate Pillai’s trace per speaker (Hall-Lew 2010: 1, 4; Hay, Warren and Drager 2006: 467), and second, they claim that Pillai’s trace indicates the distance of two vowels (Hall-Lew 2010: 3, Hay, Warren and Drager 2006: 467). As the names imply, multivariate analyses of variance (MANOVAs) analyze variances and Euclidean Distances express distances between vowels.

Pillai’s trace is simply the proportion of multivariate variance accounted for by the phoneme predictor (e.g. LOT and THOUGHT; cf. Mayers 2013: 320). Pillai’s traces range from zero to one: values near zero indicate that no variance is accounted for by the phonemes (i.e. they are merged), and values near one mean that all the variance is accounted for by the phonemes (i.e. they are unmerged; cf. Gorman and Johnson 2013: 233). In the erratum on her website to the (2010) article, Hall-Lew states: “I now think that the Pillai score (or any regression-type analysis of variance between two vowel clusters) may be a better measure of difference than of distance” (Hall-Lew n.y.; my emphasis). Unfortunately, she now does not ‘think’ that computing Pillai’s traces
requires the data to be distributed in a particular fashion: sample size and covariance have to be equal (Mayers 2013: 321). Especially for LOT and THOUGHT tokens, the sample size is usually not equal, as LOT tokens are much more frequent than THOUGHT tokens; this might also have a negative effect on the equality of covariance. In addition, the relationship between the calculated Pillai’s trace and its corresponding p-value remains unclear and unmentioned by the authors: In my replication of using Pillai’s trace for the low back vowels in the way described by Hall-Lew (2010) and Hay, Warren and Drager (2006), some of Pillai’s traces were near zero but had significant p-values, and some of them were near one but had insignificant p-values.

According to Gorman and Johnson (2013: 234), the ED measures for LOT and THOUGHT as used by Labov, Ash and Boberg (2006b: 59) produce results similar to the per-speaker Pillai’s trace, so that both measures can be used interchangeably in this case. The caveat for Pillai’s traces is that they require a large number of observations, because linear-model-based statistical tests such as MANOVAs will produce no or unreliable results when the number of observations is small.

In essence: As outlined in Figure 4.21, the statistical tests all have several assumptions that have to be met before the tests can be employed. The most important one is independence of samples since its violation cancels the suitability of MANOVAs, multivariate regression and two-sample t-tests. Although a paired t-test for dependent samples remains suitable after the independence assumption is violated, it is not at all clear, for instance, which LOT token should be paired with which THOUGHT token in styles other than minimal pair. The established procedure to arrive at independent observations is to calculate the arithmetic mean per vowel and speaker (cf. Subsection 4.6.2.3). This reduces my number of observations to 34 in total and to even fewer per age group, and thus cancels the suitability of MANOVAs and linear regression, as both of these tests require a large number of observations (cf. Gorman and Johnson 2013: 234).

Due to these contemplations, I base my interpretation of the results with regard to the low-back merger primarily on a t-test for independent samples to compare the means of LOT and THOUGHT, after the samples (per age group) have been tested for normality via the Shapiro-Wilk test and for homoscedasticity via an f-test. For every t-test, I additionally calculate the effect size Cohen’s d, so that my results are comparable to those of other studies (Gries 2009: 211). In order to answer the main research question (cf. Section 3.4; also cf. Hypothesis 2 in particular), an insignificant result of the comparison between LOT and THOUGHT first formant means and between LOT and THOUGHT second formant means of only young speakers suffices.

For the sake of completeness, however, I also outline my replication of statistically assessing the merger via Pillai’s traces, and I employ an ANOVA to age group and the merged LOT-THOUGHT second formant mean values only, due to its more robust separation of the phonemes (Gorman and Johnson 2013: 232), after I tested for the assumptions of a linear model with the global test in R (cf. Peña and Slate 2006). As mentioned above,
ANOVA\text{s rely on a linear model which requires a large number of independent observations, so that the per-speaker results of the ANOVAs are triangulated with linear and logistic regression, including random effects/slopes (cf. Figure 4.21). I assess the presence of the non-differentiation between the TRAP and BATH vowels in St. John’s English similarly via \textit{t}-tests between the speaker means of both phonemes per age group. The literature did not suggest a differentiation between these two vowels in the speech community of St. John’s (cf. e.g. Clarke 2010a: 29), so that I did not control for the relation of BATH to TRAP tokens in the process of segmentation and extraction of vowel tokens.

### 4.6.3.2 Shifted Vowels

Everything that I stated about the statistical determination of merged vowels in the subsection above also holds for shifted vowels, except that the results of the statistical tests should be significant. In order to calculate the EDs in the case of the remaining three vowels that are understood to be part of the Canadian Shift, i.e. KIT, DRESS and TRAP, a reference point stable in apparent time is necessary. Results from the pilot study suggested that STRUT behaves relatively stable in apparent time in comparison to literally all other vowels included in the analyses. In order to test this stability, I employ a MANOVA based on Wilks’ lambda with the two formants’ mean values per speaker as the dependent variables and age group as the predictor variable. Beforehand, I tested for the homogeneity of multivariate variance-covariance matrices with Box’s M test and for the assumptions of linear models in general via the global test separately per dependent variable (first and second formant means, cf. Subsection 4.6.2.4). The assumption of equal sample sizes is violated, as I have 12 young speakers, 13 middle-aged speakers and 9 old speakers (cf. Subsection 4.2.2).

Out of the four outcomes of MANOVAs in SPSS, Pillai’s trace, Wilks’ lambda, Hotelling’s trace and Roy’s largest root, Wilks’ lambda is the most powerful one when more than two variants of the independent variable are entered into the model (here: age groups old, middle and young) and when the assumptions of Pillai’s trace of equal sample size and covariance are violated. Hotelling’s trace can be used when only two groups are considered (e.g. age groups old and young) and Roy’s largest root focuses only on the first factor (e.g. old speakers; cf. Mayers 2013: 321).

With the stability of STRUT in apparent time in place, I calculate the EDs based on speaker means\textsuperscript{70} per age group for KIT, DRESS and TRAP, respectively. I employ ANOVAs to test whether or not the EDs are significantly different in apparent time. As mentioned above, ANOVAs and MANOVAs are based on linear models that require the numbers of observations to be large in order to yield reliable results. The total number of observations per vowel is limited to 34 speakers, so that I triangulate the results of the

\textsuperscript{70} The Shapiro-Wilk test results for KIT, DRESS and TRAP can be found in the Appendices B.7, B.6 and B.4.
Data and Methodology

As outlined in Section 2.3, the *a priori* hypotheses about the behavior of KIT, DRESS and TRAP when participating in the Canadian Shift on the mainland, are only undisputed with regard to the latter vowel. TRAP changes predominantly in second formant space (the dependent variable). With regard to the former two vowels, Boberg (2005: 141) maintains that the vowels display a movement in second formant space, but not in first formant space (i.e. retraction, but not lowering). According to Boberg (2010: 147), however, KIT and DRESS both lower and retract simultaneously in his PCE data. Later in his book, he suggests parallel retraction of KIT and DRESS rather than lowering and thus no classic chain shift (Boberg 2010: 227). Yet, he also finds that KIT displays a movement in first formant space, but does not provide the results for a possible movement in second formant space (i.e. lowering, but no retraction; 2010: 227). Fortunately, his students provide a less confusing summary of their results, which is based on similar data: They find that KIT and DRESS are lowered and retracted in apparent time (Sadlier-Brown and Tamminga 2008: 10).

One of the methods Boberg (2005) and his students (2008) employed to arrive at these observations are problematic: Both articles use Pearson product-moment correlation coefficients ($r$) with age as the continuous independent variable and first and second formant separately as the dependent variables. The coefficients are defined to fall in the range between –1 and 1. The sign before the $r$ value reflects the direction and the absolute value of the strength of the correlation (Gries 2009: 138). A negative correlation of the variables mentioned above thus indicates that the older the speakers are, the lower the values of first and second formant (i.e. raising in F1 and retraction in F2 in apparent time). A positive correlation indicates that the older the speakers are, the higher the values of first and second formant (i.e. lowering in F1 and fronting in F2 in apparent time). If the value of $r$ is approximately zero, there is no statistical correlation between age and first and second formants, respectively; if it is close to –1 or 1, there is a very high negative or positive correlation (Gries 2009: 139).

The correlation coefficient is calculated by dividing the covariance by the product of the standard deviations of age and first formant (and age and second formant; cf. Gries 2009: 141). The covariance is derived by computing the differences of each variable’s value from the variable’s mean (e.g. [age of individual speaker minus average age] times [first formant value of individual vowel token minus average first formant value]) and by dividing the computed difference by the total number of observations minus one (cf. Gries 2009: 140). Pearson’s $r$ thus requires the continuous input variables to be normally distributed (and possibly without outliers), as the $r$ values rely on mean and standard deviation (cf. Subsection 4.6.2.1).

As outlined in Subsection 4.6.2.4, Kendall’s $\tau$ should be used when the assumption of normality is violated, and the correlation should be at least intermediate ($0.3 < r < 0.5$),
preferably high or very high (0.5 < r < 0.9; cf. Gries 2009: 139; Mayers 2013: 322). If the data are not normally distributed and/or have few outliers, the \( r \) values will, for instance, increase when a few younger speakers have very high first formant values (as the \( a \ priori \) hypothesis of shifted KITs and DRESSES among young speakers suggests), so that the strength of the correlation is mistakenly interpreted as high or very high.

In my data, the input variables (speaker means of DRESS) include outliers. The values of Kendall’s \( \tau \) for DRESS with age as independent variable and first formant as dependent variable yielded a result of –0.3 (i.e. the older the speakers are, the smaller the values of F1);\(^71\) with age and second formant the result yielded a value of 0.37 (i.e. the older the speakers, the larger the value of F2).\(^72\)

Unlike Boberg (2005, 2010) and his students, I add a statistical test of the correlation for two reasons: First, the correlations between age and mean formant values are rather low and barely reach an intermediate level. This suggests that the differences in formant values is not only a function of age, which is why I triangulate the statistical tests with multivariate generalized linear mixed-models such as linear and logistic regression. Second, the mere comparison of two numbers with regard to similarity and difference appears to me too vague to verify or refute vowel movement in a classic chain shift-like fashion as Boberg (2010: 227) did. If the test yields significant results for both correlations, there is no justification in assuming that movement in one of the formants is absent in apparent time.

4.6.3.3 Triangulation via Generalized Linear Mixed-Modeling

Earlier variationist sociophonetic studies predominantly used GoldVarb or more generally VARBRUL for regression analysis which has several limitations: Most relevantly, it does not include random intercepts and slopes (cf. Johnson 2009; Saito 1999), it does not easily allow for discovering and handling interactions (Bayley 2008; Johnson 2009; Mendoza-Denton, Hay and Jannedy 2003; Sigley 2003), and the dependent variable has to be binary (ordinary logistic regression; cf. Bayley 2008; Johnson 2009). It might be due to these modeling constraints that, for instance, Labov, Ash and Boberg (2006b) report values for first and second formants that establish a threshold beyond which a certain sound change or part of a sound change is present (e.g. /ɛ/ in DRESS is lowered as part of the Canadian Shift if formant one is greater than 660 Hz; 2006b: 130).

Since formant values are continuous (infinite), linear regression should be the preferred analysis for this kind of dependent variable modeling. The disadvantage is that the residuals have to be normally distributed in linear regression (Wickens 2004: 4), which is most likely not the case with individual formant values from all speakers. Although normality can be achieved via data transformation, this may be undesirable when it comes to modeling several predictors and their relative contributions to the non-transformed

\(^71\) Smaller first formant values indicate raised KIT vowels (compared to those of younger speakers).

\(^72\) Larger second formant values indicate fronted KIT vowels (compared to those of younger speakers).
dependent variable simultaneously (Quinn and Keough 2002: 359). It might be due to these constraints that logistic regression is the preferred statistical model in variationist analysis (Paolillo 2002: 1).

Whatever the reason, logistic regression via GLMM is the better choice when it comes to determining how speakers behave with regard to a lowering/non-lowering threshold or a retraction/non-retraction threshold for vowels, especially when previous literature established different mean positions in the vowel space. In comparison, results from linear regression “only” reveal whether predictors affect formant values in that they are smaller or larger. The question that thus arises is whether a shifted vowel really has to be defined in incomparable (i.e. across different studies) formant mean values, especially since vowel shifts occur neither abruptly nor categorically, but gradually. Yet, in analogy to Labov, Ash and Boberg (2006b) and to a lesser degree Boberg (2010), and due to the different assumptions of linear versus logistic regression, I will also work with threshold values in logistic regression.

Furthermore, VARBRUL selects predictors which affect the dependent variable in a stepwise fashion: Out of all the predictors entered initially into the model, only those remain in it that exert a significant effect on the dependent variable (step-down regression). Stepwise predictor selection has been increasingly criticized in the past few years, so that full-model fits are now the preferred method in regression analysis (Harrell 2001: 58). Among the problems of stepwise regression are, for instance, $R^2$ values that are biased high, $p$-values that are too small/significant (severe problems of multiple testing) and that predictor selection is made arbitrary due to collinearity (Harrell 2001: 56-57; see also for more problems of stepwise variable selection algorithms). In addition, the full model fits I use “[...] have the advantage of providing meaningful confidence intervals [...]” (Harrell 2001: 59).

Instead of VARBRUL, I use SPSS 21 (cf. Bayley 2013b) and Rbrul 2.21 (cf. Johnson 2009) for the Generalized Mixed-Effects Modeling of all predictor variables. The most obvious difference between the two is their output. Johnson’s (2009) script provides an output very similar to the one of VARBRUL, with log odds, application rates, factor weights and a $p$-value for each factor group that is not part of an interaction. In direct comparison, SPSS shows the intercepts, statistics (e.g. Wald Z), $p$-values and confidence intervals. An additional difference between the two programs in their default settings lies in the definition of contrasts: Rbrul is set by default to sum contrasts, which compare the grand mean of a data set with a single treatment mean (Crawley 2007: 380). SPSS is set by default to treatment contrasts, which chose a random factor in a factor group and set it to the value of the baseline mean (intercept) and compare each of the other factors to this randomly chosen one (cf. Crawley 2007: 371, 377). According to Gries

73 This does not imply that my data is directly comparable to the thresholds (mean formant values) established by Labov, Ash and Boberg (2006b). However, Boberg (2010) has based his entire book on a comparison to those thresholds.
(2009: 278), using sum contrasts is hotly debated, but many ANOVA findings reported in the literature are based on them. In mixed-effects models, calculation of sum contrasts is difficult, especially when random slopes are additionally added to the model (cf. SPSS 2005: 1).

The baseline mean is the intercept of the null model when all predictors are mathematically set to a value of zero (cf. Baayen 2008: 244). In a two-dimensional coordinate system, where the dependent variable is $Y$ and the independent variable is $X$, the intercept is the value on the $y$-axis, when $X$ is zero, i.e. the regression line is extended to cut through the $y$-axis. For instance, when age as the only predictor in a model is set to zero ($x$-axis), the dependent variable will be of a certain intercept value ($y$-axis). This value is usually (falsely) larger than zero with linguistic data, because zero-value predictors are not part of a data set (i.e. we do not record 0-year-olds in sociolinguistic interviews).

For continuous variables it sometimes makes more sense to mean-center them, i.e. subtracting the mean age of the data set from the age of each respondent (cf. Baayen 2008: 254 for centering “trial” in a psycholinguistic experiment), so that the intercept is closer to the dependent-variable values (e.g. formant values) of the mean-aged respondents in the data set. Alternatively, continuous variables can be centered on the smallest or largest value. I centered age and the local-ness index total on their smallest values in each model (20 years of age and 1.5 LItotal value).

The fact that the dependent variable in linguistic data is often not zero when every predictor is set to zero (mathematically), is expressed by the intercept’s $p$-value, which is usually of high significance. In other words, the significance of the intercept is socially not meaningful. Setting predictors to zero does, however, not make much sense for categorical values, as a setting for sex to zero does not have any social meaning whatsoever. The absolute value of the intercept is thus meaningless, because it dependents too much on the coding of the categorical variables (alphabetical order) and the complexity of the model (i.e. number of variables).

Model complexity and/or model fit to the observed data is sometimes understood to be reflected in a model’s $R^2$ value (see below). Although it is possible to manually calculate $R^2$ values in R (R Development Core Team 2006-13) for the individual fixed and random effects, neither Johnson’s (2009) script (based on R’s lme4 package, version 1.1-06, for mixed-effects models) nor SPSS return them by default, at least at the time of writing, because the variances that are modeled jointly stem from different sources (cf. Baayen 2008: 258). In fixed-effects models (without random effects), $R^2$ usually informs about the variance explained by just the fixed effects, most commonly by subtracting the sum of squares of the residuals (error) from 1 (Kramer 2005: 150). In mixed-effects models (with random effects), $R^2$ would inform about the variance of both effect types, which is not equally desirable by all the different disciplines that use general statistical software such as SPSS: Whereas some linguists may be more interested in the variance explained by linguistic fixed and random effects, sociolinguists are more interested in the
variance explained by social effects (fixed) or the combination of social and linguistic effects (fixed; similar to psycholinguists; cf. Baayen 2008: 259).

Matters are complicated, however, in terms of variance explanation in mixed-effects models, because different analysts are faced with different problems, which necessarily emphasize the importance of different parts of a model. According to Kramer (2005: 151), this fundamental part of modeling a process cannot be resolved mathematically, so that there is no general definition of $R^2$ for every mixed-effects model. Although $R^2$ values for ordinary regression and ANOVA are similar, the involved philosophies or assumptions about what $R^2$ is supposed to represent differ in the statistical literature (cf. e.g. Nakagawa and Schielzeth 2013: 134). When these differences are applied to mixed-effects models, the $R^2$ values can differ as well (Kramer 2005: 148).

Very generally, $R^2$ is taken to be a measure of the explanatory value of the statistical model in terms of the observed data, but is sometimes also taken as a way of assessing how well a model fits the observed data (cf. Edwards et al. 2008: 6138; Hosmer and Lemeshow 2000: 167; Labov 1994: 51-53). With regard to model fit or suitability, other measures are available such as Akaike’s information criterion (AIC), Bayesian information criterion (BIC) or deviance.

The latter is a measure of discrepancy between the fitted values of the model and the values of the data. Deviance is defined as $-2$ times the difference in log-likelihood between the current model and the saturated model (i.e. a model perfectly fitting the data). The smaller the deviance, the larger is the likelihood, i.e. the more adequate is the model (cf. Crawley 2007: 516).

The former two, AIC and BIC, are penalized likelihood measures for model selection, which penalize the likelihood based on the total number of parameters (predictors) and the number of subjects (speakers) in the model (cf. Nakagawa and Schielzeth 2013: 134; Seltman 2013: 373). Their values thus increase with every parameter (predictors, interactions, etc.) added to the model. They decrease, however, when the model with the added parameters accounts better for the behavior of the dependent variables than the model with fewer parameters did. Both information criteria are thus better suited in terms of model fit, as $R^2$ will not decrease once more parameters are added and thus distorts the interpretation of actual suitability of the model (Kramer 2005: 152). The absolute value of AIC (and BIC) has no interpretation; instead, the smaller the value they return, the better the model is balanced between complexity and good fit (Seltman 2013: 373; they are also suitable criteria for nested models; cf. Chatterjee and Hadi 2006: 327).

AIC and BIC can also be used to determine whether a model differs in terms of balance of complexity and fit only in fixed effects or only in random effects but also in random versus fixed effects. In the former case (fixed effects), Seltman (2013: 374) suggests to use Maximum Likelihood (ML) as an estimation method, rather than Restricted/Relativized/Residual Maximum Likelihood (REML). Fitting models by ML is
known to cause variance components to be biased, especially in small samples, often underestimating random-effects variances (cf. Nakagawa and Schielzeth 2013: 136). Zuur et al. (2009: 119) state that the two estimation methods produce dissimilar results, especially when the number of fixed effects is large in relation to the number of observations, without further specification. If the information criteria are smaller and p-values significant for the interactions and random slopes, the model is better balanced and the interactions and slopes are needed (Seltman 2013: 372). However, if the factor groups that are part of an interaction and/or random slope are of interest, it may be better to resolve them via deletion of tokens or recoding of individual factors (see below), than entering them as an interaction/random slope in the model.

Unlike $R^2$, AIC and BIC do not inform about the amount of variation in the independent variables explained by the model (Kramer 2005: 153, Nakagawa and Schielzeth 2013: 134), they also do not inform about the absolute suitability of a model, but the relative fit of one model in relation to alternative models, and they are highly data-set specific, so that they are not necessarily comparable across different data sets (Nakagawa and Schielzeth 2013: 134). However, $R^2$ artificially inflates when several predictors are added that essentially measure the same in a different guise (multi-/collinearity; e.g. vowel duration and number of syllables after the vowel; cf. Mayers 2013: 323 for examples from psychology).

Despite the problems associated with $R^2$ outlined above, Nakagawa and Schielzeth (2013: 137) suggest a general and simple method for calculating variance explained ($R^2$) by generalized linear mixed models ($R^2_{GLMM}$), which remains to be tested for its empirical validity in further studies (2013: 140). The two $R^2_{GLMM}$ values they suggest are marginal $R^2_{GLMM}$, which returns the $R^2_{GLMM}$ for the fixed effects in a mixed-effects model, and conditional $R^2_{GLMM}$, which returns the $R^2_{GLMM}$ values for the fixed and random effects in a mixed-effects model (Nakagawa and Schielzeth 2013: 136).

Although I will use their method for obtaining $R^2_{GLMM}$ for the three lax vowels trap, dress and kit, I will not consider it or discuss its implications any further (except for a general comparison in Subsection 5.6.2) for the following three reasons: First, some of the problems associated with obtaining $R^2$ in mixed-effects models may not be sufficiently resolved (e.g. a possible decrease in $R^2$ values with additional predictors in the model; cf. Nakagawa and Schielzeth 2013: 137); second, it is at present unclear whether their

74 The $R^2_{GLMM}$ values are obtained via the following formulae: 1) $R^2_{GLMM(m)} = \frac{\sigma^2_f}{\sigma^2_f + \sum_{l=1}^{n} \sigma^2_l + \sigma^2_e + \sigma^2_d}$, where $R^2_{GLMM}$ is “variance explained on the latent scale rather than original scale” (Nakagawa and Schielzeth 2013: 137), $\sigma^2_f$ is the variance calculated from the fixed effects, $n$ is the number of random effects, $\sigma^2_l$ is the variance component of the $l^{th}$ random effect, $\sigma^2_e$ is the additive dispersion component and $\sigma^2_d$ is the distribution-specific variance (Nakagawa and Schielzeth 2013: 137); 2) $R^2_{GLMM(c)} = \frac{\sigma^2_f + \sum_{l=1}^{n} \sigma^2_l}{\sigma^2_f + \sum_{l=1}^{n} \sigma^2_l + \sigma^2_e + \sigma^2_d}$, where the variance components of the random effects are additionally added to the variance calculated from the fixed effects in the numerator.
proposal will gain wider acceptance among statisticians, biologists and most importantly linguists, especially since there is “no consensus for a definition of $R^2$ for mixed-effects models” (Nakagawa and Schielzeth 2013: 134; also cf. Kramer 2005: 151 and Xu 2003: 3527); and third, most studies published on the Canadian Shift do not provide any measure of the explanatory value of the statistical model used, regardless of whether these studies employed fixed-effects or mixed-effects models (e.g. Boberg 2005, 2008b, 2010, 2011; Clarke, Elms and Yousef 1995; D’Arcy 2005; but cf. Labov, Ash and Boberg 2006b for the contrary). Instead of $R^2_{GLMM}$, I will use deviance and AIC in order to choose the best suitable model among several alternatives in my data set.

In terms of including random effects in a regression model, Seltman (2013: 371) states that SPSS indicates the necessity of such an effect via a Wald Z statistic and its corresponding $p$-value. SPSS estimates the variance of the random effects and yields a Wald Z statistic and a $p$-value based on the null hypothesis that their variance is zero. If it is not zero, the $p$-value will be significant and the null hypothesis has to be rejected. That is, a significant $p$-value signals the random effect is needed in the model, because there may be important unmeasured predictors for each speaker that raise or lower their performance. The fashion in which the performance is altered appears to be random, because the values of the missing predictors are unknown (Seltman 2013: 372). Seltman (2013: 371) explicitly suggests that an insignificant $p$-value for the variance of a random effect would indicate that it is not needed.

It is certainly true that an insignificant $p$-value does not allow for rejecting the null hypothesis of zero variance. If the variance of a random effect is really estimated to be zero, the significance of the fixed effects should not substantially change, i.e. cross the alpha-level of 0.05 in either way, because zero variance suggests that the random effect introduces no new fixed effect that is missing in the fixed effects entered into the model. Put differently, zero variance suggests that the random effect does not affect the fixed factors in the model. However, even if the $p$-value is slightly greater than 0.05, i.e. insignificant at the 95% level, the fixed effects’ $p$-values may nevertheless change substantially in some of the subordinate models with fewer tokens than the main model (this is, for instance, the case in the subordinate models for KIT with a $p$-value for lexical item of 0.094; cf. Subsection 5.5.2). Hence, although the estimated variance of a random effect is insignificant at the 95% level, the fixed effects may receive false positive $p$-values when the random effect is excluded from the model. This may also suggest that the threshold for significance at $\alpha = 0.05$ is not appropriate for including and excluding random effects.

This becomes more of an issue when more than one random effect is part of a regression model (Seltman’s example only included one random effect, subject, in a psychological experiment; cf. 2013: 372). Jaeger (2008: 444), for instance, stresses that random intercepts (for speaker and/or lexical item) are necessary if any of the fixed factors reach significance without them, but fail to do so when these intercepts are included, regardless
of the significance of the random effects themselves. If only one of the two random effects is entered into the model, the fixed factors associated with the excluded random one will increase in their p-values. For instance, glottal state will receive much higher p-values than sex when speaker is taken into consideration as a random effect, but lexical item is not.

Including speaker and lexical item in a sociolinguistic regression model is the standard procedure (cf. Johnson 2009; Saito 1999); excluding them is not only dependent on their variances’ p-value, as Seltman (2013: 371) suggests, but also on their influence on the p-values of the fixed effects. If a fixed effect remains significant after the influence of speaker and lexical item is controlled for, it is rightly significant (cf. Jaeger 2008: 444).

Unlike SPSS, Rbrul only returns the standard deviations for the random effects entered into the regression model without their corresponding p-values. The necessity of the random effects can thus be indirectly assessed by a standard deviation greater than zero per random effect (the standard deviation is the square root of the variance), and of course by the stable p-values of the fixed effects when the random effects are excluded.

4.6.3.4 Fitting a Regression Model

The purpose of this subsection is to illustrate the steps that have to be taken to modify the original input data in such a way that they can be used as input for statistical modeling. This includes close inspection of the established variables and variants in terms of them being exhaustive and mutually exclusive (i.e. each token can be categorized as one and only one variant of each variable), orthogonal (i.e. each variant of a variable can occur with every other variant of every other variable) and logically independent (cf. Guy 1993: 242). The steps illustrated below are based on the runs for kit lowering (cf. Subsection 5.5.1) and focus predominantly on the factor groups following voicing and place/manner of articulation. All of the steps outlined below apply to all factor groups/predictors that are entered into the model.

The initial input consists of all coded tokens (n = 1016). The dependent variable is binary: F1 values are classified as either < 525 Hz (not lowered) or > 525 Hz (lowered; application value). The independent variables (and variants) to be included in the analysis are listed in Table 4.11. Addressing the issue of orthogonality first, it is obvious that not each place of articulation can be combined with every manner of articulation. Cross-tabulating the two groups shows that only eight combinations are mathematically possible, so that twelve cells will have zero combinations are actually realized in English, as shown in Table 4.12. In this example, 5 \times 4 = 20 combinations are mathematically possible, so that twelve cells will have zero

---

75 The data set used here serves an illustrative purpose only and does not correspond to the full data set that I ultimately used, since coding was still ongoing and more tokens were added at later points in time. The categories for following place and manner of articulation are based on those used by Labov, Ash and Boberg (2006b) and to a lesser degree by Clarke, Elms and Youssef (1995: 214) in their original study on the Canadian Shift in Ontario.

76 Note that glottal fricatives are not realized in postvocalic position in English.
4. Data and Methodology

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variants</th>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td>Young</td>
<td>Age (continuous)</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>Age &amp; Sex</td>
</tr>
<tr>
<td></td>
<td>Old</td>
<td>Young female</td>
</tr>
<tr>
<td>Sex</td>
<td>Female</td>
<td>Middle female</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>Middle male</td>
</tr>
<tr>
<td>Voicing of following sound</td>
<td>Voiced</td>
<td>Old female</td>
</tr>
<tr>
<td></td>
<td>Voiceless</td>
<td>Old male</td>
</tr>
<tr>
<td>Place of articulation</td>
<td>Apical</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glottal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interdental</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Labial</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Labiodental</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Velar</td>
<td></td>
</tr>
<tr>
<td>Manner of articulation</td>
<td>Affricate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fricative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stop</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tap</td>
<td></td>
</tr>
<tr>
<td># of following syllables</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11: Illustration of variables and variants in the initial analysis (here KIT lowering).

<table>
<thead>
<tr>
<th>Place / manner</th>
<th>Affricate</th>
<th>Fricative</th>
<th>Stop</th>
<th>Tap</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apical</td>
<td>22</td>
<td>292</td>
<td>126</td>
<td>114</td>
<td>554</td>
</tr>
<tr>
<td>Glottal</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>Labial</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>Labiodental</td>
<td>0</td>
<td>210</td>
<td>0</td>
<td>0</td>
<td>210</td>
</tr>
<tr>
<td>Velar</td>
<td>0</td>
<td>0</td>
<td>201</td>
<td>0</td>
<td>201</td>
</tr>
<tr>
<td>SUM</td>
<td>22</td>
<td>502</td>
<td>378</td>
<td>114</td>
<td>1016</td>
</tr>
</tbody>
</table>

Table 4.12: Cross-tabulation of place and manner of articulation (KIT lowering). Non-existent consonant sounds in English are shown in italics. Laterals, approximants and glides are generally excluded in these analyses.

tokens and thus most likely cause interactions. Combining those two factor groups deletes the twelve cells.

Once place of articulation and manner of articulation are crossed, further cross-tabulation of the new groups with the variable voicing reveals additional cells which are either categorical or contain too few tokens (cf. Table 4.13). Consequently, apical taps, labial stops, glottal stops and apical affricates should be excluded from the analysis, so that 827 tokens remain. Further cross-tabulation shows that variant three of the variable number of following syllables has only four tokens, which are subsequently excluded,
4.6. Data Analysis and Statistical Modeling

resulting in a total of 823 tokens, which would then serve as input for all future statistical runs.\textsuperscript{77}

Alternatively, rather than starting with cross-tabulations, the model could have been run as-is. In that case, inconsistencies between application totals and statistical significance would have pointed out interactions that would then be followed up on by cross-tabulation as illustrated above (cf. Walker 2010: 42-44). Table 4.14 shows the results from a one-level analysis for the variable (place/manner of articulation) before the four categorical/few tokens place/manner groups were removed.

Two things are apparent: first, the log odds and factor weights indicate that all but one factor – apical affricates at the bottom of the table – favor lowering (positive log odds favor the application value, negative ones disfavor; factor weights above 0.5 favor, those below 0.5 disfavor). However, looking at the column for the application rate, none of the 22 affricate tokens is lowered (application ratio 0.000) – i.e. this is a categorical context displaying no lowering, which is another reason for its removal. When investigating the column with the application ratios further, a relationship between application rate and (dis)favoring effect becomes obvious: the higher the application rate, the more favorable is the environment to lowering.

Yet, there is a problem in the expected linearity (in boldface): glottal stops show an application ratio of 0.167, which should come above the 0.133 of apical fricatives. Glottal stops, apical fricatives and most likely apical taps are interacting with another factor in the analysis, which can be determined by cross-tabulating place/manner of articulation with all other independent variables. As outlined above, voicing is likely to be responsible. Labial stop is the only factor that does not ‘misbehave’ in the fashion described above, showing that interaction spotting/anticipating should be done prior to analysis. Once the

\textsuperscript{77} More details on data inspection can be found in Tagliamonte (2006, 2012); Guy (1993); Paolillo (2002) and Walker (2010).

<table>
<thead>
<tr>
<th>Place/manner</th>
<th>Voicing</th>
<th>SUM</th>
<th>SUM</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Voiced</td>
<td>Voiceless</td>
<td>Adjusted</td>
<td></td>
</tr>
<tr>
<td>Apical affricate</td>
<td>4</td>
<td>18</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Apical fricative</td>
<td>51</td>
<td>241</td>
<td>292</td>
<td>292</td>
</tr>
<tr>
<td>Apical stop</td>
<td>47</td>
<td>79</td>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>Apical tap</td>
<td>114</td>
<td>0</td>
<td>114</td>
<td>0</td>
</tr>
<tr>
<td>Glottal stop</td>
<td>0</td>
<td>18</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>Labial stop</td>
<td>5</td>
<td>28</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Labiodental fricative</td>
<td>41</td>
<td>169</td>
<td>210</td>
<td>210</td>
</tr>
<tr>
<td>Velar stop</td>
<td>96</td>
<td>105</td>
<td>201</td>
<td>201</td>
</tr>
<tr>
<td>SUM</td>
<td>358</td>
<td>656</td>
<td>1016</td>
<td>827</td>
</tr>
</tbody>
</table>

Table 4.13: Cross-tabulation of following voice and crossed place and manner of articulation (KIT lowering).
<table>
<thead>
<tr>
<th>Factor</th>
<th>Logodds</th>
<th>Tokens</th>
<th>Appli. ratio</th>
<th>Centered factor weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apical stop</td>
<td>2.103</td>
<td>126</td>
<td>0.254</td>
<td>0.891</td>
</tr>
<tr>
<td>Apical fricative</td>
<td>2.085</td>
<td>290</td>
<td><strong>0.133</strong></td>
<td>0.889</td>
</tr>
<tr>
<td>Apical tap</td>
<td>1.994</td>
<td>114</td>
<td><strong>0.114</strong></td>
<td>0.88</td>
</tr>
<tr>
<td>Glottal stop</td>
<td>1.548</td>
<td>18</td>
<td><strong>0.167</strong></td>
<td>0.825</td>
</tr>
<tr>
<td>Velar stop</td>
<td>1.424</td>
<td>201</td>
<td>0.100</td>
<td>0.806</td>
</tr>
<tr>
<td>Labial stop</td>
<td>0.806</td>
<td>33</td>
<td>0.091</td>
<td>0.691</td>
</tr>
<tr>
<td>Labiodental fricative</td>
<td>0.574</td>
<td>210</td>
<td>0.043</td>
<td>0.64</td>
</tr>
<tr>
<td>Apical affricate</td>
<td>-12.441</td>
<td>22</td>
<td>0.000</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Table 4.14: Results for crossed place and manner of articulation with interactions (KIT lowering).

However, the loss of almost 200 tokens is not necessarily desirable, especially when they contain valuable information. Instead, simultaneous modeling of the predictors, including a crossed category of place and manner, voicing and an interaction of the two (place/manner*voicing) retains the 200 tokens and accounts for the interaction. This only becomes undesirable when the whole model with other social (and linguistic) predictors included becomes too complex or when the interaction turns out to be significant (with the Bonferroni approximation). Alternatively to removing the tokens, the factors can also be regrouped/simplified based on Bonferroni-corrected multiple comparisons of the dependent variable’s means (cf. Crawley 2007: 374-377).
In this chapter, I present the analysis of the vowels generally agreed upon to be involved in the Canadian Shift. As outlined in Chapter 2, I consider Labov’s (1991, 1994; and colleagues’ 2006b) Peripherality Hypothesis to be a suitable framework to account for the vowel chain shift, despite the objections that have been raised. In this hypothesis, the shift is triggered by the low-back merger, which vacates a space on the non-peripheral track in low-back position. The LOT vowel retracts toward the THOUGHT vowel and both merge. In a pull-chain, TRAP retracts toward the space formerly occupied by the LOT vowel. Consequently, the lax front vowels DRESS and KIT lower and retract toward the position of TRAP and DRESS, respectively.

Section 5.1 provides a first overview of all the vowels measured acoustically for the subsequent individual analyses. The graphical representation of these vowels per age group already exhibits the results of the following statistical analyses. Moreover, this representation shows many apparent time vowel shifts in the speech of St. John’s residents that go beyond the scope of this thesis. In the methodology discussed above, I stated that a movement or shift of any vowel is first examined via the Euclidean Distances of a vowel in question relative to a stable anchor point. In Subsection 5.1.1, I will show that STRUT fulfills the criterion of stability in apparent time, necessary for the vowel to serve as a reference point from which Euclidean Distances are measured. Another prerequisite for the analysis of each of the vowel is to verify the non-differentiation of TRAP and BATH words in St. John’s English. I consider this analysis a formality or prerequisite given the situation of these two vowels described in the literature. According to Clarke, Newfoundland English shares this non-differentiation with the vast majority of North American varieties instead of – as may be suggested by the settlement history of Newfoundland – the southern British varieties, “which in the late seventeenth and eighteenth centuries underwent lengthening and retraction of the vowel in many BATH words […]” (2010a: 29).
After these preliminaries are considered in some detail, I proceed with the analysis of the low-back merger in my data set in Section 5.2. The first two subsections are concerned with the statistical assessment of the merger per age group. There is general consensus in the literature that the low-back merger is present in Newfoundland, although there is no agreement as to whether it is homogeneously completed. In Subsection 5.2.1, I compare the perceptual data from my respondents concerning the presence of the merger before /t/ and before nasals to the production data. Statistical assessment of the latter will provide conclusive evidence that the shift is completed for the young and middle-aged speakers in my data set. The old speakers behave more heterogeneously in this regard.

After having analyzed and statistically assessed the stability of STRUT, the non-differentiation of TRAP and BATH and the low-back merger, the plot of all the vowels analyzed acoustically can be refined and redrawn. The modified graph marks the beginning of Subsection 5.2.2, in which I triangulate the test results of the low-back merger with different methods: While I based my results in the preceding subsection on simple t-tests of speaker means between the LOT and THOUGHT vowels, in this subsection, I use an ANOVA of the Euclidean Distances between the two vowels’ per-speaker means (cf. e.g. Baranowski 2007; Labov, Ash and Boberg 2006b; Wassink 2006) and calculate Pillai’s trace of the individual LOT and THOUGHT tokens per speaker (cf. e.g. Hall-Lew 2010; Hay, Warren and Drager 2006) in order to compare the results in terms of their suitability to assess the merger (as outlined in Subsection 4.6.3.1). The results naturally vary, but the youngest age group always shows presence of a complete merger, regardless of the methods I used.

The two final subsections of Section 5.2 provide another and more recent means for investigating merger – mixed-effects regression modeling (cf. Gorman and Johnson 2013). In order to categorize the continuous dependent variable (F2) into a binary one for the logistic regression, I employ the data exploration methods Decision Trees and Optimal Binning. In case of LOT and THOUGHT, the overfitted (non-pruned) Decision Tree revealed that old speaker 20SCOF behaves quite innovatively (i.e. her merged vowels are realized in a very back position), so that the results of the regression models including her LOT-THOUGHT tokens would most likely be skewed. For this reason, I discuss the Decision Tree in detail in Subsection 5.2.3 and ultimately exclude the speaker from regression modeling of LOT-THOUGHT. The linear and logistic regression models for the merger are outlined in Subsection 5.2.4. The insignificant results suggest that the low-back merger is not a change in progress in St. John’s, Newfoundland. That is, the merger is present in all age groups in my data set, whether in a front (higher F2 values) or back position (smaller F2 values).

The structure introduced in Section 5.2 is ultimately the same for each subsequent sections on each of the lax front vowels, with some minor adaptations. As outlined by Clarke, Elms and Youssef, the hypothesis that front lax vowel lowering as part of the Canadian Shift might be governed by similar phonological constraints for each lax vowel
was too simplistic (1995: 214). Therefore, the respective subsections differ primarily as a function of the varied phonological constraints per lax vowel participating in the shift. In case of the low-back merger, following nasals favor merging of LOT and THOUGHT as a first stage. Following /t/ provides a less favorable context and thus marks a later stage in the process. In terms of TRAP, which is detailed in Section 5.3, following nasals and /g/ provide linguistic environments in which the vowel is raised and fronted, rather than lowered and retracted. As I will discuss in Subsection 5.3.1.1, these environments should be excluded or treated as distinct units of investigation when analyzing the Canadian Shift (cf. Boberg 2005, 2008b, 2010; Labov, Ash and Boberg 2006b), although not every impressionistic/acoustic phonetic study on the Canadian Shift does so, especially not the two that examined TRAP retraction in St. John’s, Newfoundland (cf. D’Arcy 2005; Hollett 2006, 2007).

In Subsection 5.3.1.2, I use Euclidean Distances between TRAP and STRUT to statistically assess the movement of /æ/ in apparent time without assuming a particular direction of the shift. I answer the question of whether it lowers or retracts in apparent time in the subsequent subsection, 5.3.1.3, via coefficients of correlation (cf. e.g. Boberg 2005). While a t-test assesses statistical significance between a categorical variable and a continuous one, correlation coefficients quantify the correlation between two continuous (ratio-scaled) variables. The results of a significance test of the correlation coefficients confirm those of an additional t-test: TRAP is retracting in apparent time, but not lowering in my data set, so that formant two will be the dependent variable in the subsequent mixed-effects regression modeling. I discuss the linear and logistic regression models in detail in Subsections 5.3.2.1 and 5.3.2.2, respectively. The results they provide are quite compelling with regard to TRAP retraction. When combining the findings for merged LOT-THOUGHT and TRAP, the status of both vowels in St. John’s, Newfoundland, is such that the Canadian Shift is free to develop, i.e. DRESS and KIT can lower and retract in a fashion suggested by the Peripherality Hypothesis. This status is generally not contested in the literature (cf. e.g. Boberg 2010; Clarke 1991, 2012; D’Arcy 2005), with the exceptions of Hollett (2006, 2007) and Labov, Ash and Boberg (2006b).

The status of the mid and high front lax vowels, DRESS and KIT, in St. John’s, Newfoundland, has only been analyzed in one acoustic phonetic study (Hollett 2006, 2007), using a larger sample than six participants but generally yielding inconclusive findings. This situation leaves little room for discussing my results in light of Hollett’s study. Instead, I discuss my findings primarily with regard to the two other recent acoustic phonetics studies, Boberg’s (2005, 2010) and Labov, Ash and Boberg’s (2006b), due to the methodological similarities I aimed at when I designed my study. I further elaborate them in light of the pioneering research on the Canadian Shift, conducted by Clarke, Elms and Youssef (1995).

I commence my elaboration with the movement of DRESS in Subsection 5.4.1, based on the Euclidean Distances of this vowel in relation to STRUT. The result, confirming a
movement in apparent time, leads to the statistical assessment of the direction of that movement via correlation coefficients in Subsection 5.4.1.2. Boberg originally claimed in his (2005) article that the dominant movement of DRESS (and KIT) in Montreal, Quebec, was retraction rather than lowering as originally suggested by Clarke, Elms and Youssef (1995). He revoked this claim in his revised version of the Canadian Shift in Montreal published as part of his (2010) book on Canadian English. My results based on correlation coefficients confirm this position for St. John’s, Newfoundland, and the results of Labov, Ash and Boberg (2006b) confirm this position for Canada from coast to coast. My results show a significant movement of DRESS in both formants, one and two, in apparent time, i.e. DRESS is moving diagonally down and back toward the position formerly occupied by TRAP in the vowel space. In line with Labov, Ash and Boberg (2006b), I use formant one as dependent variable in the regression analyses of DRESS outlined in Subsection 5.4.2.

The restrictions and decisions made for the analysis of DRESS are also true for KIT. The participation of KIT in the Canadian Shift is generally contested: Among a few others, Labov, Ash and Boberg (2006b) maintain that a shift of /1/ could not be found in their data, while the majority of research reports a movement in apparent time for that vowel (cf. e.g. Boberg 2010; Clarke, Elms and Youssef 1995; Hoffman 2010). As I will repeat in detail in Section 5.5, Labov, Ash and Boberg’s claim seems contradictory to the age coefficient they present for KIT, so that the participation of the this vowel in the chain shift could be regarded as less contested as it may seem. The analysis of the Euclidean Distances between KIT and STRUT, discussed in Subsection 5.5.1, confirms a shift of the former in apparent time. The correlation coefficient significance, discussed in the same subsection, shows a diagonal movement of KIT towards the initial position of DRESS. In analogy to the regression models for DRESS, I use formant one as dependent variable in the linear and logistic regression, outlined in Subsections 5.5.2.1 and 5.5.2.2, respectively.

After each of the vowels participating in the Canadian Shift have been analyzed and discussed in detail, I summarize these results in Section 5.6 in order to provide a comprehensive picture of the Canadian Shift in St. John’s, Newfoundland. I structured the summary according to the methods I used in the discussion of the individual vowels, so that all of the findings from the Euclidean Distance metrics are summarized in Subsection 5.6.1. In addition, I provide the mean formant values of merged LOT-THOUGHT, TRAP, DRESS and KIT per age group, including the standard deviations (sds).

In Subsection 5.6.2, I summarize the significant predictors from each of the individual regression models in order to provide the perspective of my interpretation of the results of this study. As Clarke, Elms and Youssef (1995: 214) have pointed out for the shift on the mainland, the constraints for each of the lax vowels are not the same, which is also true for St. John’s. In the same subsection, I answer the main research question of this thesis. In addition to the significant predictors of the regression models, I discuss the summary of the explanatory values of each of the regression models ($R^2$ values) and compare them
to those outlined by Labov, Ash and Boberg (2006b). It will become apparent that the explanatory value of my models is similar to that of Labov, Ash and Boberg’s models, and that the different $R^2$ values seem to provide some support to my interpretation of the data.

The last subsection of the summary (5.6.3) provides three case studies representative of my three age groups old, middle and young. The case studies generally support the findings derived from the quantitative analysis, although the representative of the old age group shows strong idiosyncratic patterns in the position of his high front tense and lax vowels. I discuss my case studies primarily in light of those shown by Boberg (2010) and Labov, Ash and Boberg (2006b) for their speakers from St. John’s, Newfoundland. This qualitative complement thus provides some consolidation with the results of the quantitative analyses discussed above.

## 5.1 First Results

The results of 617378 ANAE normalized vowel measurements from the 34 speakers included in my analysis are shown in Figure 5.1. The individual vowels are represented by their mean first and second formant values and their respective standard deviations. The means and standard deviations are derived from the speakers per age group. The standard deviations are represented as ellipses around the individual vowel means. In order to avoid a visual overload in the graph by the ellipses that extend well beyond the range of the first and second formant values plotted, it shows one-half the size of the standard deviations. Before any impressionistic assumptions can be made from the graph, it is necessary to verify whether the vowel means provide a suitable measure of central tendency, i.e. whether the data is normally distributed, and whether age is the (strongest) predictor (independent variable) that can account for the difference in vowel means.

The Shapiro-Wilk tests for normality per age group and vowel yielded partially highly significant results (cf. Appendices B.8 to B.10), i.e. the mean values provided here are not accurately summarizing the vowels. However, a vowel plot with medians revealed the very same relationship between the vowels and very similar positions of the vowels in the plot (cf. Appendix C). One notable exception is provided by the fleece vowel, which has an almost identical median for both the young and the middle-aged speakers. The Decision Trees with social predictors and first or second formant as dependent variables identified age as the primary (statistically significant) reason for the difference in means (cf. Appendix E).

Due to the similarity of the vowel mean and median plot, I take impressions inferred from the plot presented here to be reasonably accurate with regard to the vowels belonging

---

78 This figure excludes 4558 vowels before approximants, glides and nasals (and in the case of trap and bath before /g/) and after approximants and glides.
Figure 5.1: ANAE normalized means of formants one and two in apparent time (n = 6173). The ellipses represent the standard deviation divided by a factor of two in order to make the graph visually more accessible. The age groups are coded by the symbols in the legend on the lower left. The large difference in the first formant means of the TRAP and BATH vowels of middle-aged speakers is due to the low number of BATH tokens. The movement of FLEECE, FACE, FOOT and GOOSE has to be interpreted with caution, because of the comparatively low number of tokens per vowel (total n ≈ 1100).

The ellipses represent the standard deviation divided by a factor of two in order to make the graph visually more accessible. The age groups are coded by the symbols in the legend on the lower left. The large difference in the first formant means of the TRAP and BATH vowels of middle-aged speakers is due to the low number of BATH tokens. The movement of FLEECE, FACE, FOOT and GOOSE has to be interpreted with caution, because of the comparatively low number of tokens per vowel (total n ≈ 1100).

to the Canadian Shift: THOUGHT, LOT, TRAP, DRESS and KIT. The other vowels were essentially only included for normalization, except STRUT, and are thus low in number. Consequently, any direct comparison of these vowels with the Canadian Shift vowels is misleading.

Despite these limitations, the vowel plot in Figure 5.1 suggests that the vowels not belonging to the Canadian Shift also move in apparent time. Especially GOOSE seems to be fronting in apparent time (cf. Boberg 2010: 240 and Hollett 2005 for St. John’s), a behavior similar to what is reported in the literature for standard Canadian (mainland) English: As outlined in Section 2.1, Canada as a dialect region is characterized by the presence of the low-back merger, GOOSE fronting and the Canadian Shift, among others (Labov, Ash and Boberg 2006b: 146). The roles of FLEECE and FACE have not been mentioned with regard to innovative features in mainland Canadian English. The movement of FLEECE in apparent time is misleading in this graph, as the median plot in Appendix C indicates its movement only between old and middle-aged/young speakers and not between the middle and young age group. FACE seems to be raised and tensed
5.1. First Results

<table>
<thead>
<tr>
<th>Test</th>
<th>$p$-Value F1</th>
<th>$p$-Value F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global stat</td>
<td>0.8</td>
<td>0.12</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.6</td>
<td>0.56</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.3</td>
<td>0.94</td>
</tr>
<tr>
<td>Link function</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>0.58</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Table 5.1: Assumptions of a MANOVA regarding the distribution of age groups and mean formant values of STRUT: results of the global test on four degrees of freedom ($n = 34$). Insignificant values are shown in square brackets. Level of significance is set to 0.05.

in conjunction with lowering and retraction of KIT. With regard to FOOT, Boberg (2005: 144) found early signs of centralization among his youngest speakers. This movement seems to be present in my data as well, with FOOT centralizing and lowering. Hollett (2006: 158) did not report such a finding in her data of St. John’s English. However, it should be stressed once more that all of the observations in terms of vowels that do not a priori belong to the Canadian Shift have to be interpreted with a grain of salt.

Of much more relevance are the vowels that have been reported to be involved in the Canadian Shift: As the plot implies, LOT and THOUGHT are very close in proximity, suggesting the presence of the low-back merger. Corroborating the results of the pilot study, STRUT appears to be relatively stable in apparent time, as it behaves more similarly to the low back vowels than to the short front vowels. BATH and TRAP seem to be relatively close to one another in age groups old and young, but not within the middle-aged speakers. This interpretation has to be made with great care, as the number of BATH tokens is very small, compared to the TRAP tokens (ratio of 1:5). The literature on Canadian English in general does not support the assumption that the two low-front vowels are differentiated as in British English varieties (cf. Subsection 4.6.3.1). In addition, the behavior of the middle-aged speakers is marginally important in the present analysis, as they are a priori taken to provide a more or less homogeneous intermediate picture between the old and the young speakers. Finally, DRESS and KIT seem to retract and lower in apparent time.

5.1.1 The Stability of STRUT in Apparent Time

In order to statistically assess the behavior of the short front vowels in apparent time, the relative stability of STRUT, the merger of TRAP and BATH, as well as that of LOT and THOUGHT, have to be tested. I employ a MANOVA based on Wilks’ lambda in order to assess the stability of STRUT in apparent time (across all three age groups), after I test the assumptions of linear models separately for the dependent variables, mean first and
second formant. These results are provided in Table 5.1.\footnote{Note that I put the insignificant findings in square brackets, following the convention of sociolinguistic studies (e.g. Boberg 2005). In the results presented in the appendix, I put the insignificant results in boldface and the significant/marginally significant results in italics for visualization purposes, since those sections are meant to highlight insignificant results rather than significant ones.} The global test on four degrees of freedom (Peña and Slate 2003, 2006) yields insignificant results for all the assumptions of linear models – skewness, kurtosis, the link function and heteroscedasticity – for first formant values, i.e. the assumptions are accepted, but the variances in second formant values are heterogeneous \((p = 0.009)\). More importantly, however, the global stat, which takes all four tests into account, did not yield any significant results. In other words, the global test, which takes account of the \(\alpha\)-error inflation in multiple testing, states that the assumptions are met. Box’s M Test of equality of covariance matrices yields an insignificant result \((F_{6,12578} = 1.31, p_{\text{two-tailed}} = 0.25)\), i.e. the null hypothesis that the observed covariance matrices of the dependent variables are equal across the three age groups cannot by rejected, so that the assumption of multivariate homoscedasticity is met as well. The unequal sample sizes (9 old speakers, 13 middle-aged speakers and 12 young speakers) should not affect the test’s reliability as its outcome is Wilks’ lambda (cf. Subsection 4.6.3.1).

The graphical representation of the distribution of STRUT speaker means per age group is shown in Figure 5.2. The boxplots reveal much information about the distribution of the data: The horizontal bold line represents the medians of the three samples. The ‘+’ signs, which are not part of boxplots by default, indicate the mean of each data set, and thus provide an easily accessible comparison of mean and median. The regular-typed horizontal lines – the hinges – above and below the median that constitute the upper and lower boundary of the boxes represent the 75%- and the 25%-quartiles of each of the data sets. The dashed vertical lines – the whiskers – extending from the box to the upper and lower limit represent the largest and smallest values that are not more than 1.5 interquartile ranges away from the hinges of the box. Each outlier would be represented by individual circles outside the range of the whiskers (cf. Figure 5.3 for an example). The notches that extend from the medians toward the left and right sides of the boxes include values within the range of 1.58 times the interquartile range divided by the square root of the number of tokens in each data set: \(\pm \frac{1.58 \times \text{IQR}}{\sqrt{n}}\) (Gries 2009: 119). If the notches of two or more boxplots overlap, as is the case in Figure 5.2, then these will most likely be insignificantly different (see below). I added the dashed horizontal line to represent the grand mean of the three data sets in order to show how far away the main body of each data set is from the grand mean. The sizes of the boxes indicate the interquartile ranges (similar to the variance) of the data sets, and thus allow to anticipate the result of \(f\)-tests or Bartlett’s test visually (homogeneity of variance). An equal length of the whiskers indicates that the distribution of the data may not be normal and that the distribution may be long-tailed or short-tailed (skewed), respectively. Figure 5.2b shows nicely that, although a visualization provides an excellent summary of the data, it is also misleading:
5.1. First Results

(a) Boxplots of first formant per-speaker mean values per age group.

(b) Boxplots of second formant per-speaker mean values per age group.

Figure 5.2: Boxplots of first and second formant mean values of STRUT per age group (n = 34).
The values of the young speakers’ second formants of STRUT seem to be negatively skewed (to the left) and thus may not be normally distributed. As I outlined above, the global test does not allow me to accept the alternative hypothesis that the data is skewed (cf. Table 5.1).

Since the tests yielded no significant results for violation of symmetrically distributed data and since the boxplots indicate the absence of outliers, I employ Pearson’s product-moment correlation coefficient in order to determine the strength of the correlation between the dependent variables in the MANOVA. The result of Pearson’s \( r = 0.29 \) indicates that the correlation barely reaches the threshold of 0.3, so that there may be no multivariate effect (cf. Subsection 4.6.2.4). Since I do not expect any effect, separate ANOVAs would suffice to determine the stability of STRUT in apparent time.\(^80\)

The speaker means of the first and second formants of STRUT are very similar to each other. As Figure 5.2a shows, the average first formant for old speakers is 711 ANAE Hz\(^81\) (standard deviation [sd] 36 Hz) while that for middle-aged speakers is 719 Hz (sd 26 Hz) and that for young speakers is 733 Hz (sd 28 Hz). For the second formant, the average for old speakers is 1343 Hz (sd 37 Hz) compared to the averages of middle-aged speakers of 1381 Hz (sd 92 Hz) and of young speakers of 1392 Hz (sd 65 Hz; cf. Figure 5.2b).

According to the MANOVA employed, the univariate differences are insignificant. The variances of the first formant means between age groups are not significantly different: \( F_{2, 31} = 1.57; p_{\text{two-tailed}} = 0.22 \); and the variances of the second formant means between age groups are not significantly different: \( F_{2, 31} = 1.31; p_{\text{two-tailed}} = 0.28 \). As a consequence, the variable age group explains only 3% of the variance in the first formant: multiple \( R^2 = 0.092 \); adjusted \( R^2 = 0.034 \); and it explains only 2% of the variance in the second formant: multiple \( R^2 = 0.078 \); adjusted \( R^2 = 0.019 \). The multivariate differences are insignificant as well: The variances of both formant values between age groups are not significantly different: \( F_{4, 60} = 1.15; p_{\text{two-tailed}} = 0.34 \) (Wilks’ \( \lambda = 0.86 \)). The effect size for this result is low: \( \eta^2 = 0.07 \). With one independent variable (here: age group) in the model, the effect size eta-squared is also \( R^2 \) (Gries 2009: 279). This means that age group explains only 7% of the overall variance in first and second formant mean values.

As I hypothesized in Section 3.4, this result confirms that STRUT is stable in apparent time (cf. Hypothesis 3). Consequently, I can use the STRUT vowel as a stable anchor or reference point in order to assess the difference in distances of KIT, DRESS and TRAP between young speakers’, middle-aged speakers’ and old speakers’ means. This result corroborates that of Labov, Ash and Boberg (2006b: 220) for Canada from coast to coast, but is not in line with that of Clarke, Elms and Youssef (1995: 212) for Ontario and Hollett (2006: 157) for St. John’s, Newfoundland.

\(^80\) In fact, \( t \)-tests of two independent samples (old and young speakers’ mean values of first and second formants separately) would be sufficient to assess this stability.

\(^81\) All of the reported Hertz values in this chapter are normalized values, based on Labov, Ash and Boberg’s (2006b) normalization method. Henceforth, I will stop indicating this explicitly for each individual value through its unit (ANAE Hz), and refer to the values in Hertz (Hz) only.
5.1. First Results

Figure 5.3: Boxplots of the per-speaker mean first formant values of middle-aged speakers for BATH and TRAP. The boxplot of TRAP reveals one outlying speaker mean at approximately 850 Hz (n = 34).

5.1.2 The Status of TRAP and BATH

As outlined in Section 5.1, the number of tokens for TRAP and BATH is imbalanced, which may cause the large difference in the mean first formant values of the middle-aged speaker tokens. It is highly likely that a t-test of the middle-aged speakers’ means of the first formants will yield a significant result. The difference between TRAP and BATH in second formant space for these speakers is marginal in Figure 5.1. The middle-aged speakers’ mean distribution of the first formant values can be inspected in Figure 5.3. The notches clearly do not overlap, so that the first formant means of TRAP and BATH are most likely significantly different here.

This behavior of BATH’s F1 mean cannot be convincingly explained: If we, for instance, assumed that TRAP and BATH are split in St. John’s, Newfoundland, because it has been settled by southwest Englishmen and southeast Irishmen, the old speakers should show an even larger distance in their first formant values than the middle-aged speakers. More importantly, however, in British varieties of English, BATH is realized far more centrally than low in comparison to TRAP – in the lower middle of the vowel quadrilateral. The second formant values should thus be much smaller for BATH than for TRAP, possibly accompanied by a less extensive increase in the first formant of BATH than visible in Figure
5. Analysis and Discussion

### Table 5.2: Overview of the t-test results for TRAP and BATH.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phoneme</th>
<th>T-test for independent samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Trap</td>
</tr>
<tr>
<td>Old</td>
<td>F1 (Hz)</td>
<td>774</td>
</tr>
<tr>
<td></td>
<td>Sd</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>1952</td>
</tr>
<tr>
<td></td>
<td>Sd</td>
<td>66</td>
</tr>
<tr>
<td>Young</td>
<td>F1 (Hz)</td>
<td>803</td>
</tr>
<tr>
<td></td>
<td>Sd</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>1725</td>
</tr>
<tr>
<td></td>
<td>Sd</td>
<td>41</td>
</tr>
</tbody>
</table>

5.1. I thus consider the first formant values of BATH for my middle-aged informants as an artifact of the data rather than a meaningful movement in apparent time.

Since the literature does neither support the differentiation of TRAP and BATH words in St. John’s English\(^82\) nor reports change in the formants individually (non-simultaneously) and abruptly in the speakers’ thirties to forties (cf. Subsection 4.6.3.1), and since the small token number of BATH does not allow for developing a theory from this behavior in apparent time, statistical assessment of the middle-aged speakers’ means – in fact, of all speakers’ TRAP and BATH means – is not informative but nevertheless conducted for the sake of completeness.\(^83\) It is, however, informative to confirm the a priori hypothesis of the merged status of TRAP and BATH, so that I employ a t-test to first and second formants separately for old and young speakers. Since the dependent variables and/or the independent variables differ per test, the $\alpha$-error ($p$-value) should not be inflated.

The t-test assumptions are met; their results can be found in Appendix D.1.1. According to a t-test for independent samples, the small difference of 5 Hz between the average first formant frequencies of TRAP and BATH produced by old speakers is statistically not significant ($p = 0.91$) at a low effect (Cohen’s $d = 0.07$). The difference of 29 Hz in average second formant frequencies of TRAP and BATH produced by old speakers is insignificant as well ($p = 0.71$); the effect size is low (Cohen’s $d = 0.23$; cf. Table 5.2). These results suggest that the vowels TRAP and BATH are merged for old speakers of St. John’s English.

The result of a t-test for independent samples of the average first formant frequencies of TRAP and BATH of young speakers suggests, that the difference of 16 Hz is not significant

---

\(^{82}\) Recall that Newfoundland English differs in this respect from southern British English, which in the late seventeenth and eighteenth centuries underwent retraction of the BATH vowel.

\(^{83}\) Due to the multitude of figures from test statistics, degrees of freedom, $f$- and $p$-values for the four $t$-tests to follow, I put these results in the appendix. See Appendix D.1.1 for the assumptions of two-sample $t$-tests. As expected, the $t$-tests themselves (cf. Appendix D.1.2) yield significant results for first formant mean values ($p_{two-tailed} = 0.009$), but not for second formant mean values ($p_{two-tailed} = 0.44$).
The result of a $t$-test for independent samples of the young speakers’ second formant means is insignificant as well, due to a difference of 0 Hz between TRAP and BATH. It yields a very high $p$-value (0.99) at a low effect (Cohen’s $d = 0.004$). As expected, these results indicate that TRAP and BATH are also merged among young speakers of St. John’s English. The results are summarized in Table 5.2.

5.2 The Low-back Merger

The low-back merger is the most crucial part of the chain that allows the vowels to shift in the fashion the Canadian Shift proposes. As mentioned in Section 3.2, almost all of the relevant literature attests that the low-back merger is present in St. John’s, Newfoundland (e.g. Kirwin 1993: 74), but not necessarily completed (cf. Boberg 2010: 240; Scargill and Warkentyne 1972: 64). The only qualification was provided by Labov, Ash and Boberg (2006b: 217) and consequently Clarke (2010a: 31), who maintain that the merger is in a front position. This location may hinder the development of the Canadian Shift, as it reduces the margin of security to TRAP so that it cannot unconditionally retract (cf. Boberg 2010: 239; Labov, Ash and Boberg 2006b: 220). The split values yielded by the Decision Trees and Optimal Binning suggest a front position for old males and a back position for young females (cf. Appendix E.1). The Decision Tree for LOT indicates that age exerts the strongest influence on formant two, when the latter is modeled as dependent variable. In that way, the Decision Tree yields second formant values which may serve as thresholds for logistic regression.

The following two subsections will analyze the merger in apparent time with age group as the only predictor. As outlined in Subsection 4.6.3.1, I first employ simple $t$-tests for independent samples per speakers’ formant means within each age group. The results indicate whether the means are statistically different from one another. Such testing does not account for differences between age groups. Subsection 5.2.2 is devoted to an ANOVA with age group as the independent variable and Euclidean Distances (EDs) as the dependent variable. EDs do not make any assumptions with regard to distribution of the data and express the length of the distance between two vowels in a two-dimensional coordinate system such as that outlined in Figure 5.1 above. EDs take the contribution of first and second formants equally into account when determining the distance between two vowels. This may be undesirable since the variances and/or ranges of the two formants differ. Because $t$-tests and EDs do not take the relationship of first and second formant into account, a MANOVA calculating Pillai’s trace was offered as an alternative (e.g. Hay, Warren and Drager 2006). This method is also problematic, because it makes a number of assumptions about the distribution of the data that are difficult to meet with linguistic data. The last subsection contrasts the findings of these tests with those of linear/logistic regression. Regression analysis allows multiple predictors to be included.
5. Analysis and Discussion

in the model (similar to MANOVAs) and speaker and word to be included as random effects/slopes (dissimilar to MANOVAs). It thus extends the findings with regard to the low-back merger as it outlines the relative contributions of the other predictors in relation to age.

The discussion of the results presented here is quite comprehensive, because of the differences between the results of the various methods, which, however, primarily concern the old speakers only. Young and middle-aged speakers, on the other hand, almost always have a merger of the low back vowels, regardless of the methods I use. I would like the reader to bear this behavior of particularly the young speakers in mind when reading through the discussion of the results concerning the old speakers.

5.2.1 The Merger within Age Groups

Labov, Ash and Boberg (2006b: 217) find that the merger before nasals was “almost uniform” throughout Canada, including Newfoundland and St. John’s. This result is corroborated by all of my respondents, as everybody answered the questions in interview section three (cf. Appendix A.4.3) of whether don and dawn sounded the same (rhymed), close or different with ‘same’ (n = 34). With regard to the low-back merger before /t/, only 29 out of 39 of Labov, Ash and Boberg’s Canadian respondents stated that the words hot and caught rhymed, whereas eight interviewees were ‘close’ in either production or perception and two stated the words sounded different to them, but said them as rhymes (Labov, Ash and Boberg 2006b: 217). Their two speakers from St. John’s, Newfoundland, differed in their perceptions of hot and caught: one was the “same in perception and production” and the other one was “different in perception” (2006b: 218). Except for two out of nine older speakers, my respondents stated that cot and caught rhymed (n = 32). The two that considered the minimal pair to sound different also had the largest Euclidean Distance values between lot and thought of 240 Hz (female speaker 08GCOF) and 170 Hz (male speaker 07HPOM) respectively.

As mentioned in Subsection 4.6.3.1, the second formant separates the vowel classes LOT and THOUGHT more robustly than the first formant, as LOT retracts more than THOUGHT lowers. A t-test of the second formant mean values of the two low back vowels should thus suffice to determine their status in St. John’s, Newfoundland. Due to the difference in advancement of the merger before nasals and other following consonants, I exclude LOT and THOUGHT tokens before nasals, yielding 651 LOT tokens and 338 THOUGHT tokens (total of 989 tokens). The statistical assessment is not only based on the minimal pair tokens, but all LOT and THOUGHT vowels in the three interview styles word list, reading passage and interview (cf. Subsection 4.2.2).

Figure 5.4 shows the distribution of the young speakers’ means in first and second

---

84 Tokens in pre-approximant (and post-approximant) position are generally excluded (cf. Subsection 4.2.2).
5.2. The Low-back Merger

(a) The first formant means of LOT and THOUGHT for young speakers.

(b) The second formant means of LOT and THOUGHT for young speakers.

Figure 5.4: Boxplots of the first and second formant means of LOT and THOUGHT for young speakers, excluding pre-nasal environments (n = 12).
formant space, excluding pre-nasal environments. The small differences in the mean values and the overlapping notches in the two boxplot pairs already strongly hint at the expected merged status for this age group. The differences in mean first formant frequencies of 7 Hz and in mean second formant values of 5 Hz will most likely be insignificant. Except for the mean second formant values of LOT, the data seems to be distributed normally. The boxes are of similar sizes (i.e. interquartile ranges), and no outliers are visible. Some of the values within the notches lie outside the hinges, so that it may have been better to draw the boxplots without the notches. However, I wanted to show whether or not the differences in means/medians between the two low back vowels are likely to be significant in the graphs. The largest mean second formant value, and thus the frontmost realization of LOT, among my youngest respondents was uttered by a 20-year-old male and is at 1420 Hz. Instead of graphs, I use statistical tests for middle-aged and old speakers only (cf. Appendix D.2).

The results of the $t$-tests regarding all mean formant values for all age groups are summarized in Table 5.3. The $t$-tests of speaker means of LOT and THOUGHT yield insignificant results for all three age groups: For old speakers, the difference of 108 Hz between the mean second formant frequencies of the two vowels is insignificant ($p = 0.08$) at a low to intermediate effect. The difference of 22 Hz for middle-aged speakers and of 5 Hz for young speakers is insignificant ($p = 0.5$ and $p = 0.84$, respectively) at intermediate and low effects as well. This means that the low back vowels are merged in production in all three age groups before non-nasal sounds. The large difference in $p$-values between old and middle-aged/young speakers is a direct result of the large range in second formant means of the old speakers. They behave quite heterogeneously in the realization of LOT and THOUGHT, which is also supported by the relatively large standard deviations of the second formants (133 Hz and 109 Hz, respectively). The heterogeneity of old speakers in my sample is caused by the two speakers mentioned above that seem to have no merger, neither in production nor perception, two speakers that have a completed merger and by the remaining five speakers that are distributed between these two extremes.

From these results in apparent time, it is not straightforwardly plausible to argue that the merger in St. John’s began in the generation represented by my old respondents. For instance, the large Euclidean Distances, which also take the first formant into account, between LOT and THOUGHT for the male speaker 07HPOM and the female speaker 08GCOF may also originate in differences in sex, style, education and in the varied degrees of their local-ness to St. John’s. However, the overall age-group patterns allow for the interpretation of a stepwise procedure of the merger. It increases in plausibility if the results of linear and logistic regression support this finding (cf. Subsection 5.2.4). Despite the heterogeneity in the group of old speakers regarding the second formant values of LOT and THOUGHT, the results of the statistical assessment suggest that the two speakers who consider cot/hot and caught as different are exceptions in contemporary middle-class St. John’s English, but not the norm.
5.2. The Low-back Merger

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phoneme</th>
<th>$T$-test for independent samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lot</td>
<td>Thought</td>
</tr>
<tr>
<td>Old</td>
<td>F1 (Hz)</td>
<td>Mean: 814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd: 38</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>Mean: 1402</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd: 133</td>
</tr>
<tr>
<td>Middle</td>
<td>F1 (Hz)</td>
<td>Mean: 808</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd: 28</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>Mean: 1364</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd: 77</td>
</tr>
<tr>
<td>Young</td>
<td>F1 (Hz)</td>
<td>Mean: 814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd: 37</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>Mean: 1326</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd: 41</td>
</tr>
</tbody>
</table>

Table 5.3: Overview of the $t$-test results for LOT and THOUGHT, excluding pre-nasal tokens.

Although the sound quality of the data used in the atlas is “clearly not comparable”\(^85\) to that of “face-to-face interviews” (Labov, Ash and Boberg 2006b: 36), the authors maintain that in the majority of cases the acoustic analysis of a signal with a bandwidth ranging from 300 to 3000 Hz was possible with a satisfactory degree of reliability. Boberg (2010: 145, 147) directly compares his findings to the mean values derived from Labov, Ash and Boberg’s analysis, showing that his interregional mean of 86 Canadian subject of 1214 Hz (sd 71 Hz) meets Labov, Ash and Boberg’s (2006b: 130, 151, 219) average threshold for the low-back merger at $F_2 < 1275$ Hz for Canada as a whole. The data they base their assumptions on consist of minimal pairs in their word list (don and dawn) and inquiry into rhymes (hot and caught) with 39 Canadian respondents (Labov, Ash and Boberg 2006b: 217).

Direct comparison of my young speakers’ average second formant means to those of Labov, Ash and Boberg reveals a difference of maximally 40 Hz ($1275$ Hz versus $1315$ Hz; sd 63 Hz). The 35-year-old male case study, David B.,\(^86\) from St. John’s, Newfoundland, presented with reference to the realization of START by Labov, Ash and

\(^85\) The audio recordings provided on the accompanying website for the two speakers of St. John’s English are of very poor quality and accompanied by mechanical noises (cf. Labov, Ash and Boberg 2006a: n.p.).

\(^86\) The number of Canadian respondents in Labov, Ash and Boberg’s data seems to have been altered throughout the course of their work. The total number of 39 Canadian speakers is more often mentioned than 41, so that I assume 39 to be correct (cf. Labov, Ash and Boberg 2006b: 28, 217). Out of these, 33 were analyzed acoustically.

\(^87\) The name is a pseudonym. The speaker is identifiable via the label of his interview ‘TS662’, which means that he was the 662$^\text{nd}$ respondent during the nine years of data collection for the atlas (Labov, Ash and Boberg 2006b: 25). The pseudonym used on the accompanying website is ‘Dean B.’ (Labov, Ash and Boberg 2006a: n.p.), and the one on the accompanying CD-Rom is ‘Duke B.’ (Labov et al. 2006: n.p.).
Boberg (2006b: 221), shows an average position of the merged LOT-THOUGHT vowel near 1530 Hz. According to Labov, Ash and Boberg (2006b: 218), the low-back merger is only uniform in St. John’s before nasals, so that David B.’s vowel tokens for the merger may include nasals. According to the accompanying website to the atlas, David B. was interviewed in 1997 (Labov, Ash and Boberg 2006a: n.p.). I conducted my interviews in 2011, i.e. 14 years have to be added to David B.’s age in order to assess the difference between his second formant mean to those speakers of my data set who are comparable to David B. in terms of age (49). I have three male speakers who are 46 and 47 years of age, respectively. The two 46-year-old respondents have a mean second formant value of merged LOT-THOUGHT before /t/ of 1307 Hz (17LEMM; sd 117 Hz; local-ness index total 1.5) and 1296 Hz (24PSMM; sd 95 Hz; local-ness index total 3). The 47-year-old respondent has a value of 1260 Hz (22KCMM; sd 103 Hz; local-ness index total 5). This juxtaposition of second formant means is based on LOT and THOUGHT tokens in all phonological environments but those in pre-nasal position, since the above-mentioned results suggest that the merger is present before these phonological environments in all three age groups in my data set.

Inclusion of pre-nasal environments should provide even more conclusive results with regard to the merger of the low back vowels. Adding the 170 tokens before nasals I coded yields a total number of 1156 LOT and THOUGHT tokens. The mean second formant values of LOT of the middle-aged speakers do not change much: 17LEMM’s value increases from 1307 Hz to 1320 Hz, 24PSMM’s value increases from 1296 Hz to 1339 Hz and 22KCMM’s value increases from 1260 Hz to 1304 Hz. None of the second formant means is similar to the one Labov, Ash and Boberg (2006b: 221) report for their respondent from St. John’s, Newfoundland (approximately 1530 Hz), with pre-nasal environments included. The increases are rather small (maximally 44 Hz) and will most likely be insignificant (see below). The increase in values does not mean that Labov, Ash and Boberg’s proposition that the merger is more advanced before nasals is wrong, because the second formant means of the individual LOT and THOUGHT vowels are higher for both vowels compared to their means before the other phonological environments. The distance LOT has to retract in order to merge with THOUGHT is thus the same before nasals, given the second formant’s more robust character to separate the two vowel classes (Gorman and Johnson 2013: 232). In addition, neither Labov, Ash and Boberg, nor Gorman and Johnson state any order of first versus second formant movement in the course of the merger of the two vowels.

Labov, Ash and Boberg’s (2006b: 217) claim that the low-back merger is in a front position and only present before nasals in St. John’s, Newfoundland, is thus based on one arguably representative speaker (more details below). Although Clarke’s (2010a: 31) confirmation is based on seven speakers (only in her 2010 book), she does not provide any acoustic measurements which would support it. The second formant mean of my youngest speakers is clearly 100 Hz fronter than Boberg’s (2010: 145) Canadian mean, but it is also
more than 200 Hz backer than Labov, Ash and Boberg’s (2006b: 221) mean of David B. Since my standard deviation (63 Hz) for young speakers in St. John’s is almost as large as Boberg’s (71 Hz; 2010: 145) for young speakers from eight Canadian regions, I assume that young speakers from St. John’s are about to further retract their merged vowel in LOT-THOUGHT words. This assumption is supported by the comparison of the means per age group: Young speakers have a mean value of merged LOT-THOUGHT, including nasals, of 1315 Hz (sd 63 Hz), middle-aged speakers of 1358 Hz (sd 89 Hz) and old speakers of 1360 Hz (sd 140 Hz; cf. Figure 5.5 below). The relatively small change in second formant values across age groups is most likely insignificant, so that this assumption is rather a tendency (but cf. Subsection 5.2.3; also cf. Subsection 5.3.1 which is based on EDs).

Baring the lack of validity of a comparison of means from the two different data sets in mind, I tentatively suggest that Labov, Ash and Boberg’s respondent from St. John’s may not be a typical middle-aged middle-class speaker. This suggestion is supported by two facts: (1) The second speaker from St. John’s, Newfoundland, presented in the atlas is a 24-year-old female, who was also recorded in 1997 (cf. Labov, Ash and Boberg 2006a: n.p.). She completed 14 years of education (Labov et al. 2006: n.p.) and was a student of civil engineering at the time she was recorded (Labov, Ash and Boberg 2006a: n.p.). Although Labov, Ash and Boberg (2006b: 30) had particular problems in assigning a socioeconomic index score to students, hers is as high as 86 in the spreadsheet on the CD (cf. Labov et al. 2006: n.p.) based on the occupation of the “breadwinner” in her family (Labov, Ash and Boberg 2006b: 30). (2) No reference was made as to the whereabouts in the social hierarchy of David B. (Labov, Ash and Boberg 2006b: 221). He has completed twelve years of education, indicating that he did not enroll in university after graduation from high school. According to Labov, Ash and Boberg (2006b: 30), the assignment of socioeconomic index scores was also problematic for occupations that did not match any of the 503 occupations on the list of the census data and the data of the National Opinion Research Council (NORC). David B.’s socioeconomic index for his occupation is set to zero in the spreadsheet on the accompanying CD-Rom (cf. Labov et al. 2006: n.p.), which is most likely due to the fact that he works in the military (Labov, Ash and Boberg 2006a: n.p.). The interviewer might not have asked for more details about his military rank or whether he attended university in the military. It is, however, likely that David B. is rather a member of the working class, unlike his female counterpart.

If the “major sound changes in progress” are “the main focus of ANAE” (Labov, Ash and Boberg 2006b: 27), it would seem more plausible to show the vowel pattern of a young middle-class female from St. John’s as a case study, rather than a middle-aged working-class male. This choice becomes even more dubious in light of map 15.2 (Labov, Ash and Boberg 2006b: 218) where 24-year-old Clara B. from St. John’s clearly states the merger to be the ‘same in production and perception’ before nasals and /t/ (my emphasis), which contradicts Labov, Ash and Boberg’s (2006b: 217) claim that Newfoundland only has a merger before nasals.
Boberg’s (2010: 240) case study of a young (year of birth 1981, corresponds to my 30-year-olds) middle-class female speaker from St. John’s fulfills the social prerequisites needed to study sound changes in progress – at least to a greater degree than David B. Her parents are both local to St. John’s, Newfoundland, but she has traveled quite extensively in North America, so that her pronunciation might be less representative than that of middle-class speakers who traveled less (cf. Boberg 2010: 239). Boberg does not detail much about her vowels in terms of the Canadian Shift, but states that her low-back merger is not completed. She shows some overlap of LOT and THOUGHT in the F1xF2 plot, but Boberg (2010: 239) finds significant differences between vowels in formants one and two (F1: \( p = 0.01 \), F2: \( p = 0.02 \)). He does, however, not state which test he used for these findings.

For all of his other results, he uses speaker means and compares these with regard to region, sex and city size (Boberg 2010: 201). Using speaker means in that way meets the assumption of independence of most tests. Those who do not assume independence mainly require the dependent observations to be paired. It is, however, not possible to pair his case study’s LOT and THOUGHT tokens, as it is unclear which token is dependent on which other token (cf. Subsection 4.6.3). Put differently, it is not clear whether the statistical findings have social meaning or are an artifact of the assessment. In addition, statistical assessment of individual vowel tokens is not directly comparable to those of mean vowel formant values: “[...] the low-back merger [...] is not characteristic of every Canadian speaker or community” (Boberg 2010: 240), i.e. speaker mean assessment between two Canadian regions may show that the merger is present in one of them, but individual speaker token assessment within this group may still contain speakers that do not show the merger to be present/completed.

With pre-nasal environments included, the \( t \)-test results\(^{88}\) of the mean formant values for all age groups yield insignificant results as well. The most striking difference compared to the \( t \)-tests between LOT and THOUGHT of speaker means of the formant values excluding pre-nasal environments can be found in the old age group. The \( p \)-value for the first formant decreased from 0.11 to 0.05, indicating a marginally significant result for the merged status of the low back vowels. The \( p \)-value for the second formant is unaffected. As outlined above, the second formants are more robust in separating the two vowels, so that their insignificant result provides stronger support for the merged status than the significant result for formant one. The \( p \)-values for the other two age groups have not decreased as much, lending strong support to the hypothesis that the differences in the mean formant values between LOT and THOUGHT are coincidental, rather than a function of following nasals. The most important result of the \( t \)-tests is that of the youngest speakers. Their low back vowels are clearly merged. Their mean formant two value of the merged LOT-THOUGHT vowel is at 1315 Hz, indicating that the low-back merger is

\(^{88}\) The assumptions of the \( t \)-tests are met (cf. Appendix D.2.2). The results of the tests can be found in Appendix D.2.3.
5.2. The Low-back Merger

The previous section has shown that the low back vowels among the oldest speakers behave quite heterogeneously, but the statistical analysis suggested that they are merged.

moving backwards in apparent time (mean for old speakers is at 1360 Hz), given that the merger formerly was in a front position as attested by Labov, Ash and Boberg (2006b: 220) and Clarke (2010a: 31) for St. John’s, Newfoundland, for male speakers older than 35 years. With LOT-THOUGHT in this position, TRAP can retract in the fashion reported for the Canadian Shift, as it does not enter the margin of security of the merged low back vowels. The positions of the remaining three vowels considered to be part of the Canadian Shift, which will be outlined in the sections to follow, are summarized in Figure 5.5. The individual vowels are represented by their means, including one standard deviation as ellipses around each of the vowel means. In addition, STRUT is included in the plot, as it will serve as the reference point for determining the distance of TRAP, DRESS and KIT between the three age groups. Although STRUT is stable in apparent time, as outlined in Subsection 5.1.1 above, I do not summarize the three means per age group to one mean of STRUT, because the distances of TRAP, DRESS and KIT per age group are measured from the individual means of STRUT per age group.

5.2.2 The Merger across Age Groups

The previous section has shown that the low back vowels among the oldest speakers behave quite heterogeneously, but the statistical analysis suggested that they are merged.
Figure 5.5 indicated that the difference in means between the three age groups is quite small, similar to the three STRUT means. The Euclidean Distances (EDs) between LOT and THOUGHT mean formant values indicate that they are quite small among young and middle-aged speakers, but quite large for some of the older speakers. A simple subjective assessment of the differences in distances may lead to the conclusion that the low back vowels are not merged within the group of older speakers. The ANOVA I employ with EDs in order to statistically test whether there are significant differences between the age groups cannot sufficiently determine the merged status of LOT and THOUGHT, but only the difference in variances across the age groups. Since there is no threshold ED value at which the vowels have to be regarded as merged (say 150 Hz), the ANOVA is not as accurate as the t-tests above in determining the merger.

The boxplots of the EDs per age group confirm the heterogeneity in the realization of LOT and THOUGHT among old speakers (cf. Figure 5.6). The values of the EDs range from 22 Hz to 240 Hz. Since none of these values lies further away from the interquartile range (box size) than 1.5 interquartile ranges (i.e. no outliers), the box size is quite large. For the middle-aged speakers, one ED value is much larger (approximately 120 Hz) than the rest of the data points, so that, instead of increasing the interquartile range,
this large value is considered an outlier. This is due to the fact that twelve speakers have ED values lying close together and one has a value deviating strongly from the rest. For the old speakers, the ED values are distributed more or less evenly between the low extreme (22 Hz) and the high extreme (240 Hz). Thus, the difference in EDs between the middle and young age group is much smaller than the difference in EDs within the old age group. For this reason, a monofactorial ANOVA with EDs between LOT and THOUGHT per age group does not reveal how close the two low back vowels are, but how different the variance in one age group is from that in another age group. In other words, such an ANOVA per age group – in analogy to Labov, Ash and Boberg (2006b: 62) of EDs per region (cf. Subsection 4.6.3.1)\(^{89}\) – does not assess the difference between LOT and THOUGHT, but the range of differences between the speakers (in terms of difference between LOT and THOUGHT) in one age group compared to the others.

Since the dispersion of the data sets in Figure 5.6 is indicated via the interquartile range (similar to the variance), the result of the ANOVA I employ is easy to anticipate. The variance in the old speakers’ group is most likely significantly different from the variances of the middle-aged and young speakers, provided the variances are homogeneous. The mean value of the old speakers’ EDs is more than twice as large as that of the middle-aged and young speakers, and it is far below the grand mean (65 Hz) of the EDs of the three age groups. The unequal sizes of the boxes suggest that the variances are not homogeneous; the large difference between mean and median and the unequal length of the whiskers in the boxplot for the old speakers suggest that their ED values may not be distributed normally. If those values are not distributed normally, chances are high that the residuals are also not distributed normally (cf. Subsection 4.6.2.4). However, Peña and Slate’s (2003, 2006) global test on four degrees of freedom suggests that the assumptions of linear models in general are met. As Table 5.4 shows, the results of the global test are insignificant.

The results of the monofactorial ANOVA I employed are unsurprisingly significant: \(F_{2, 31} = 9.11, p_{\text{two-tailed}} < 0.001\). The \(p\)-value even states that the differences are highly significant. The variable age group explains more than 30 per cent of the overall variance: multiple \(R^2 = 0.37\), adjusted \(R^2 = 0.33\). Pairwise post-hoc comparisons of the ED means (Tukey’s HSD) indicate that the difference in means between old and middle-aged speakers is very significant \((p = 0.005)\) and that the difference in means between old and young speakers is highly significant \((p < 0.001)\). The means of middle-aged and young speakers

\(^{89}\) Labov, Ash and Boberg (2006b: 62) did not employ ANOVAs to the Euclidean Distances. They juxtaposed the EDs and the per cent of the overall response of their speakers to the five minimal pairs contrasting LOT and THOUGHT. The responses ranged from ‘sounding the same before all allophones’ (/t/, /d/, /t/, /k/ and nasals; cf. Labov et al. 2006: n.p.) to ‘sounding different before all allophones’. The intermediate responses were labeled as ‘transient’ (cf. Figure 9.1 in Labov, Ash and Boberg 2006b: 62). In the next step, they juxtaposed the EDs and the per cent of the analyst’s judgment of the speaker’s pronunciation of the minimal pairs (cf. Figure 9.2 in Labov, Ash and Boberg 2006b: 62). They do not report any statistical test with regard to the EDs between LOT and THOUGHT (cf. Labov, Ash and Boberg 2006b: 58-65).
Table 5.4: Assumptions of an ANOVA regarding the distribution of age groups and EDs between LOT and THOUGHT: results of the global test on four degrees of freedom (n = 34). Level of significance is set to 0.05.

<table>
<thead>
<tr>
<th>Test</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global stat</td>
<td>0.19</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.09</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.08</td>
</tr>
<tr>
<td>Link function</td>
<td>1.00</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>0.73</td>
</tr>
</tbody>
</table>

differ insignificantly ($p = 0.73$). As outlined above, this result does not necessarily mean that the old speakers do not have a merger of the two low back vowels, but simply that their means of the EDs of the low back vowels are significantly different from the ED means of middle-aged and young speakers. If a mean ED threshold value for the merger was known or if there were *a priori* hypotheses that old speakers had no merger, the result would be much more informative. Put differently, a mean ED of 111 Hz in the old age group is not necessarily indicative of the absence of the merger: For instance, Labov, Ash and Boberg’s analyzed regions which showed a low-back merger had mean ED values of 100 Hz or less, those which did not show it had mean ED values of 250 Hz or more (2006b: 62).

As mentioned in Subsection 4.6.3.1, none of the statistical tests employed in this and the previous section take the correlation of first and second formant into account. One method that does so is calculating Pillai’s trace for each speaker’s total number of LOT and THOUGHT tokens. The result of Pillai’s trace is near zero when no variance is accounted for by the vowel classes and near one when all the variance is accounted for by the vowel classes. Since Pillai’s trace is calculated as an outcome of a MANOVA, the assumptions of a linear model in general have to be met: normal distribution of the residuals, independence of samples and homogeneity of variance. Pillai’s trace in particular assumes the number of observations per sample and their covariance to be equal (cf. Mayers 2013: 321). In addition, the number of observations has to be rather large (Gorman and Johnson 2013: 234). Hall-Lew (2010) and Hay, Warren and Drager (2006) assigned one Pillai score to each speaker, so that each speaker’s LOT and THOUGHT tokens have to be entered into the MANOVA. None of the assumptions are discussed in the work of Hall-Lew (2010) and Hay, Warren and Drager (2006). Moreover, the figure in Hall-Lew’s article shows that the number of observations in the speakers’ samples is not equal (cf. Subsection 4.6.3.1).

Despite these violations, I reproduced the method with my data set for the sake of illustration and triangulation. Assumptions of independence and equality of covariance are violated. The results are not straightforwardly interpretable: Pillai’s trace for speaker
07HPOM, one of the old speakers with a large Euclidean Distance of 170 Hz between LOT and THOUGHT, is 0.33. This value is much closer to zero than it is to one. This means that the variance that is accounted for by the vowel classes LOT and THOUGHT is rather small, suggesting that another factor might be responsible for the difference in variance, provided that Pillai’s trace accurately represents a speaker’s status in terms of the merger. Since some important assumptions are violated, the other factors could be non-social ones such as independence of samples or homogeneity of variance. Assuming that Pillai’s trace is accurately representing what it is supposed to represent – the difference in variance accounted for by vowel class –, the two vowel classes can be considered one, i.e. merged, because vowel class does not account for much of the difference in variance.

Figure 5.7 shows a scatterplot of the individual LOT and THOUGHT tokens for speaker 07HPOM. The diamonds represent all of his LOT tokens and the x’s all of his THOUGHT tokens. As LOT retracts more than THOUGHT lowers in the course of the merger, the smaller second formant values of LOT are a more reliable indicator of the merger than the first formant values. At least 50% of the LOT tokens are relatively front in position with second formant values greater than 1380 Hz. The other LOT tokens are not clearly separable from THOUGHT in the scatterplot. The two dashed lines indicate a clear overlap.
of the two low back vowels. Three LOT tokens are even produced as far back as the
majority of the THOUGHT tokens. This distribution (not the Pillai score) supports the
interpretation of the ANOVA results reported above: the speaker has a merger, but his
variation in realizing the two vowels is comparatively large. Pillai’s trace of 0.33 seems at
first sight to summarize the status of the merger for this speaker very well; however, visual
inspection of the data distribution is needed to confirm the validity of this summary. For
this speaker, Pillai’s trace indicates that a Euclidean Distance of 170 Hz is not large
enough to express the unmerged status of LOT and THOUGHT, if the EDs are accurately
representing the merger. In order to assess the accuracy of Pillai’s trace in relation to
ED, I plot all of the speakers’ Pillai’s traces against all of their EDs in Figure 5.8.

As mentioned above, a large ED value indicates a large distance between LOT and
THOUGHT, i.e. the vowels are not merged, and a large Pillai score indicates that the vowel
classes LOT and THOUGHT account for most of the variance of the single vowel tokens,
i.e. the vowels are not merged. This means that both values very highly correlate and
can thus be plotted in a linear-model fashion. The correlation is positive, as the larger
Pillai’s trace is, the larger is the ED. Since the residuals, i.e. the vertical distance between
5.2. The Low-back Merger

an observed Pillai’s trace and the $y$-value on the regression line for the corresponding ED ($x$-value; cf. Gries 2009: 143), are not normally distributed ($p < 0.001$), and since visual outliers are present in the data set, I use the correlation coefficient Kendall’s \( \tau \), which is based on the ranks of the values of ED and Pillai’s trace (cf. Subsection 4.6.3.1). The coefficient shows only a high positive correlation of 0.54, instead of the expected very high correlation ($\tau > 0.7$; cf. Gries 2009: 139) – a first indicator that one of the two methods may not be accurate enough (see below). If we set the threshold for the merged/split distinction of LOT and THOUGHT rather arbitrarily and subjectively to a Pillai’s trace of 0.5, and even more so to an ED of 150 Hz, we can divide the plot into four quarters, represented by the crosshairs in Figure 5.8. All of the speakers with a Pillai’s trace larger than 0.5 and/or an ED value larger than 150 Hz are understood to be unmerged with regard to LOT and THOUGHT (cf. e.g. Gorman and Johnson 2013: 233-234).

In order to interpret Figure 5.8, one of the two variables needs to be constant to assess its relation the other variable. Since EDs do not assume any particular kind of distribution of the data they are based on, I assume them to be “the constant”, i.e. the accurate measure of the low-back merger, in this example. Most of the young and middle-aged speakers cluster nicely around the regression line in the third quarter in the lower left of the figure, which means that both variables seem to account well for their merger and could thus be used interchangeably (Gorman and Johnson 2013: 234). The picture is, however, flawed by the values of the two middle-aged speakers 22KCMM and 24PSMM. The former has a low ED and should thus have a low Pillai’s trace as well. The latter behaves in the opposite direction, as he has a comparatively high ED and should thus have a higher Pillai’s trace. On the $y$-axis, their positions should be exactly the opposite of what they are in the plot. Speaker 22KCMM should have a Pillai’s trace at around 0.15 (below the regression line) and speaker 24PSMM should have one at around 0.25 (above the regression line). I take this to be another indicator of the inaccuracy of Pillai’s trace when applied to my data set.

The next important indicator is the fact that there is no speaker at all in the second quarter in the figure. Pillai’s traces of speakers 27WLMF and 29CCOF in quarter one are far too high, given that the EDs are the accurate measure of the merger, as their values are larger than 0.5 and thus suggest absence of the merger. Virtually all of the outlying old speakers are positioned in quarter four. As outlined above, the two speakers with the highest EDs within the old age group are 07HPOM and 08GCOF. Together with the large EDs of speakers 09PDOM and 32RROM, they are primarily responsible for the significant finding of the ANOVA discussed above. According to their Pillai’s traces, however, they are merged, as their values are below 0.5. This observation holds for virtually all speakers

---

90 The Euclidean Distances shown in this figure are based on the means of the speakers’ LOT and THOUGHT vowels. The mean requires the absence of outliers and the data set to be distributed normally. This assumption is not violated (cf. Appendices B.1 and B.2). However, the Euclidean Distances can also be based on medians, if the assumption of normality was violated, and would then not assume any particular distribution.
in my data set, except 27WLMF and 29CCOF, which means that 32 out of 34 speakers have the low-back merger when measured with Pillai’s traces. In addition, there is not a single speaker who clearly has no merger, as there is no Pillai’s trace value larger than 0.9 (speaker 29CCOF has a value of 0.64). In general, 8 out of 34 speakers deviate from the prediction of the linear regression in this example.

Based on these interpretations, I conclude that Pillai’s trace is not an accurate measure of the low-back merger in my data, because of the violations of the assumptions of (multivariate) linear models. In other words, we do not know whether the difference in Pillai’s trace is socially motivated (e.g. by speakers’ ages) or whether it is an artifact of the violations of the assumptions underlying Pillai’s trace calculation. Having stated that, Euclidean Distance is also only arguably useful to determine the merger of LOT and THOUGHT, as it is unclear which ED values of the two low back vowels should be considered merged and unmerged, respectively. This statement has to be qualified, as it is only valid in those cases where the ED is assessed on absolute grounds. In cases where there are a priori hypotheses or empirical evidence that suggest presence of the merger in one sample and absence in another, EDs can be well-suited for comparative statistical assessment. They may be equally well-suited when they are calculated in relation to a stable reference point for LOT and THOUGHT individually, as they can thus also be statistically assessed (but see below). However, the ED metric may be undesirable, as the second formants have a larger range than the first formants and the correlation between the two is disregarded (cf. Subsection 4.6.3.1). Unlike Gorman and Johnson (2013: 234) propose for determining the low-back merger in Labov, Ash and Boberg’s (2006b) atlas, the Pillai’s trace and Euclidean Distance metrics cannot be used interchangeably in my data set. Instead of Pillai’s trace, Hotelling’s trace or Wilks’ $\lambda$ (lambda) could be used, given that mergers usually only have two vowel classes that are merged (two samples) or that the sample sizes are usually not equal (cf. Subsection 4.6.3.1). Data could also be transformed to yield normality and homoscedasticity, but the violation of testing independent observations is not as easily managed. MANOVAs (in SPSS) do not have the option of adding speaker as a random effect or slope, which would account for the higher correlation of formant values within speakers than between speakers (dependence of observations) by disregarding it.

These better-suited alternatives are, however, not relevant for this study. Each of the methods illustrated above and in the previous subsection corroborated the suggestion of the literature in terms of presence of the low-back merger in St. John’s, Newfoundland. As outlined in the beginning of this section, the character of the juxtaposition of Pillai’s traces and EDs here is more informative than resultative, but overall supports the key findings of this thesis: In all three methods employed, the youngest age group has a fully-fledged low-back merger. Moreover, the second formant mean of the youngest speakers shows that the merger is in a back position (1315 Hz). The consequent vowel movements attributed to the Canadian Shift are hence not impeded.
5.2. The Low-back Merger

5.2.3 Non-pruned Decision Tree for LOT

Recall that Decision Trees explore data in order to find a suitable classification based on significance. I mainly used them in order to identify categories for grouping the various preceding and following environments together, but also in order to establish whether age has the strongest effect on the movements of the vowels under analysis. With regard to /o/ (LOT) and /oh/ (THOUGHT), the non-pruned Decision Tree is shown in Figure 5.9.

For the ‘overfitted’ tree (i.e. independent variables have been added multiple times to the leaves) and for the following regression models, a few tokens had to be excluded: Five THOUGHT vowels were uttered in syllable-final position, which is not possible for lax LOT tokens, so that 1151 tokens of merged LOT-THOUGHT remain for all 34 speakers, including pre-nasal environments. The presence of the merger established above is also indicated by the fact that phoneme label reaches the least significance in this data exploration (cf. The Decision Trees I employed to find suitable classifications for the linguistic environments of the vowels yielded no useful results. As mentioned earlier, I ultimately used the classification suggested by Labov, Ash and Boberg (2006b), which is in turn similar to the one used by Clarke, Elms and Youssef (1995).
What is of importance then is the back position of the merger in second formant space, so that TRAP can retract. The variables entered into the Decision Tree are: the individual second formant values (dependent variable), age, sex, style, local-ness index total (LI\text{total}), phoneme label, vowel duration, number of following syllables, preceding sounds’ voicing, place and manner of articulation (PoA and MoA) and following sounds’ voicing, place and manner of articulation (PoA and MoA). Since formant one is not independent of formant two, it has been disregarded in this Decision Tree. The importance of each variable can be found in Appendix E.1.1. Not all of the important independent variables are shown in the Decision Tree. Age exerts the strongest influence on formant two, which means that a change in the former is mainly responsible for a decrease in second formant values (i.e. a backer position).

The order of the splits in the graphic representation of the Decision Tree actually foreshadows the overall significance in the statistical analysis of preceding place of articulation, which will be discussed in the relevant subsections to follow. The split made for linguistic context shows that the two groups cannot easily be explained by linguistic (phonetic) theory: glottal (e.g. /?/) is the backest of consonantal realization in the vocal tract in English, followed by velar (e.g. /g/) and palato-alveolar (e.g. /Z/). Yet, they seem to influence formant two differently, as the backest realization groups with the frontest (e.g. /ð/). The splits made for sex and local-ness index after nodes one and two are more easily to comprehend. The 34.5% of the tokens in node one can be divided into 20% from female speakers and 14.5% from male speakers. As node 10 shows, male speakers who utter the low back vowels after voiced consonants do seem to have a merger in a front position, yielding an average formant two of 1542 Hz. However, this mean only holds for 5% of all tokens.\footnote{Although this mean value holds for so few tokens, similar mean results are obtained from Optimal Binning, which can be found in Appendix E.1.3.} After voiceless segments (cf. node 9), the mean is very similar to that of female speakers older than 38 years (cf. node 8). Within the female group, significant splits can be made according to speakers’ ages. With regard to local-ness, speakers with indices lower than 5.25 have a mean second formant value of 1299 Hz, indicating a backer position for them than for speakers with a higher local-ness index (cf. node 5). Among the former speakers, age behaves counter established theory: Speakers older than 58.5 years (statistically speaking) have lower second formant values than speakers younger than that, i.e. older speakers have a merger in a position far backer than that of 38-year-olds and younger (cf. node 12 and 7). This yields a heavily skewed number of tokens per node (cf. 38.1\% of all tokens versus 5.1\%). The statistical assessment above indicates the opposite behavior for older speakers: the older they are, the higher are their second formant values.

Closer examination of older speakers revealed that this skewed classification is largely due to one female (20SCOF). With regard to the low-back merger only, she behaves far more similarly to young females than to her contemporaries. Only in terms of LOT-
5.2. The Low-back Merger

thought, 20SCOF’s behavior is thus remarkably similar to that of old females reported by Hollett (2006, 2007) for St. John’s. In other words, her merger is in a position as back as expected for young female speakers. Her pro-hypothesis, but nevertheless outlying, behavior led me to exclude her tokens from the Decision Tree classification. The results of the pruned version of the Decision Tree excluding speaker 20SCOF can be found in Appendix E.1.2. Unlike the tree shown here, its first split is made for age, which is also indicated by the importance (significance) values for the independent variables. In node one, 48% of all tokens for speakers younger than 37 years of age display a mean second formant value of 1309 Hz. It will thus serve as the threshold at which formant two is categorized into values larger and smaller than that for the logistic regression discussed below. It is larger than the one reported by Labov, Ash and Boberg (2006b: 219) for Canada as a whole, so that more speakers will show presence of this position of the merger. At the same time, it is much smaller than the value of 1530 Hz reported by Labov, Ash and Boberg (2006b: 221) for David B. from St. John’s, Newfoundland.

The caveat of this data exploration is that the classifications are based on mean values of the dependent variable, which requires absence of outliers or normality of the distribution. This is not given for all the variables entered into the Decision Tree. Optimal Binning does not make such assumptions, but yields irrelevant results (cf. Appendix E.1.3). The significant threshold value derived from Optimal Binning shows that at least middle-aged and young speaker show a categorical behavior in that virtually all their vowels have second formant values lower than 1521 Hz.

5.2.4 Generalized Mixed-effects Modeling of the Merger

The results of the statistical assessment of the low back vowels in Subsections 5.2.1 and 5.2.2 showed that the low-back merger is definitely present among young and middle-aged speakers. They also showed that at least 50% of the old speakers are merged. Unlike indicated by the pruned Decision Tree, regression analysis may thus not necessarily return age as a significant factor, suggesting that the low-back merger is not a change in progress in St. John’s, Newfoundland, middle-class English, but rather a completed change. As we will see, this is reflected in the fact that linguistic rather than social factors are more significant than age alone.

The model was run with the Bonferroni approximation, and the estimates are based on the REML method. The outcome (dependent variable) is formant two, due to its more robust separation of the vowel classes (cf. Gorman and Johnson 2013: 232) and due to the interest in the position of the low-back merger (cf. Clarke 2010a: 31; Labov, Ash and Boberg 2006b: 221). Speaker and lexical item were set as random intercepts, including

93 As I will mention briefly in the Subsection below, only the application rates drop with Labov, Ash and Boberg’s (2006b: 219) threshold value of 1275 Hz, but the insignificance of the factor groups does not change.

94 Recall that five tokens of thought words had vowels in syllable-final position, which is not possible for lax vowels in general. They have been excluded from this model.
5. Analysis and Discussion

a random slope (an interaction with random effects) for phoneme label and speaker, allowing for the possibility that each speaker behaves differently in terms of participation in the merger. The fixed-effects predictors are age (continuous), sex, phoneme label, number of following syllables, LI_total (continuous) and the following consonant factor groups place, manner and voicing (cf. Labov, Ash and Boberg 2006b: 220 for Canada).

For all non-continuous fixed effects, I entered two-way interactions into the model (e.g. phoneme label*sex, syllable number*place, etc.), so that the model estimates the effect of phoneme label (vowel class) for the whole population and for each sex (male and female) and for each manner of articulation, etc. (cf. Gorman and Johnson 2013: 231). Since the low-back merger is not an ongoing change in St. John’s, Newfoundland, style as a fixed effect is excluded from the model.

5.2.4.1 Linear Regression

In terms of the fixed effects, the results of the linear regression analysis shows that phoneme label is not significant ($F_{1,1114} = 0.996, p = 0.319$), i.e. LOT and THOUGHT are merged because the differences in variance is not accounted for by phoneme label, and thus corroborates the results outlined above. In addition, age is not significant ($F_{1,1114} = 0.372, p = 0.542$), as expected, but the interactions between phoneme label and sex ($F_{1,1114} = 11.307, p = 0.001$) and between sex and following manner of articulation ($F_{2,1114} = 11.757, p < 0.001$) are highly significant. In addition to age, all of the other fixed effects and their interactions are insignificant, suggesting that they unnecessarily add noise to the model and reduce its prediction power (cf. Chatterjee and Simonoff 2013: 23).

With regard to the random effects, speaker and lexical item are highly significant: Wald Z = 3.414, $p = 0.001$ and Wald Z = 4.149, $p < 0.001$, respectively. The random slope specified for speaker and phoneme label is also significant (Wald Z = 2.074, $p = 0.038$), which suggests that speakers differ in their participation in the merger, as outlined above. The value is, however, close to the threshold of insignificance ($\alpha = 0.05$), supporting the notion that only very few speakers are not participating in the merger. The deviance is 13,749 (Akaike’s information criterion, AIC = 13,831).

As outlined in the previous subsection, speaker 20SCOF is responsible for skewing the behavior of old speakers in terms of the low-back merger and the position of the merged vowels. Including her is misleading with regard to the interpretation of the results for age, but also for sex, which has not been discussed yet. Her merged position is already as retracted as that of young female speakers, so that excluding her makes the overall

\[\text{Note that the implementation of GLMMs into SPSS does not allow for determining the exact } p\text{-value. The result shown is 0.000.}\]

\[\text{Higher-level interactions than two-way (e.g. phoneme label*sex*place*manner*voice) are technically possible and might logically be necessary, but modeling them in practice never resulted in a finding. In addition, the software needed ten to twelve hours to calculate in vain (cf. Sigley 2003: 252), and any finding would be difficult to conceptualize linguistically.}\]
results more conservative. Consequently, the following model is based on 1081 tokens, excluding this speaker.

The new model does not cause any change in the significance of the factor groups: Phoneme label \( (F_{1, 1044} = 1.082, p = 0.299) \) and age \( (F_{1, 1044} = 1.092, p = 0.296) \) remain insignificant (unlike indicated by the Decision Tree results discussed above), but their \( p \)-values are smaller; the interactions between phoneme label and sex \( (F_{1, 1044} = 10.079, p = 0.002) \) and between sex and following manner of articulation \( (F_{2, 1044} = 8.608, p < 0.001) \) remain significant. The other independent variables are not significant, as in the model including speaker 20SCOF (cf. Appendix F.1.1). The behavior of the random effects is similar to the model above as well. Speaker and word are highly significant: for speaker, Wald’s \( Z = 3.311, p = 0.001 \) and for lexical item, Wald’s \( Z = 4.171, p < 0.001 \), respectively. The random slope between speaker and phoneme label remains significant near the threshold of \( \alpha = 0.05 \) as well (Wald’s \( Z = 2.107, p = 0.035 \)). Since the \( p \)-value is a little smaller than with old speaker 20SCOF, some of the older speakers may produce LOT and THOUGHT distinctly. Excluding speaker 20SCOF resulted in a large, but expected decrease in deviance from 13,749 to 12,602 (AIC from 13,831 to 12,610), since the random-effects levels decreased, i.e. the model is much less complex (levels of the fixed effects remain unaltered).

The random effect, speaker, has an estimated intercept variance of 7414.777, i.e. a standard deviation of \( \sqrt{7414.777} = 86.11 \). Since 95% of the behavior of the data set in a normal distribution is scattered within -2 to 2 standard deviations, for any given fixed effect that is set to zero (which will be detailed below), each speaker has an intercept up to \((86.11 \times 2 =) 172.22\) higher or lower than the intercept estimated for the whole data set (cf. Seltman 2013: 371). The intercept is a kind of baseline mean for the dependent variable (here formant two continuous), which takes the value of the dependent variable when all predictors or set to zero (cf. Subsection 4.6.3.3). Thus, the meaning of the intercept in this model is that it is needed, because exclusion would falsely lead the model to assume the prediction line starts at \( y = 0 \) and \( x = 0 \).

The intercept helps to interpret the relation of each individual predictor (fixed and random). For instance, each speaker’s individual intercept deviates up to ±172 from the intercept per independent variable about 95% of the time, which is significant. The null hypothesis states that the speaker variance is zero, i.e. each single speaker behaves exactly the same way with regard to formant two values per independent variable. The standard deviation multiplied by two for word is 115.7, i.e. about 95% of the time each word’s individual intercept is up to 115.7 higher or lower per independent variable than

---

97 The 95% confidence intervals for the random effects can be found in Appendix F.1.1.
98 Since the LOT-THOUGHT merger is not a change in progress in St. John’s, Newfoundland (also cf. Subsection 5.2.4.2), I did not calculate \( R^2 \) (or \( R^2_{GLMM} \)) for any of the regression models (cf. Labov, Ash and Boberg 2006b: 220).
99 As outlined in Subsection 4.6.2.2, about 68% of the time data points are scattered between -1 and 1 standard deviations in a normal distribution.
5. Analysis and Discussion

<table>
<thead>
<tr>
<th>Model Term</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic $t$</th>
<th>$p$-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1366.356</td>
<td>43.616</td>
<td>31.327</td>
<td>0.000</td>
<td>1280.770 1452.942</td>
</tr>
<tr>
<td>Age</td>
<td>1.187</td>
<td>1.136</td>
<td>1.045</td>
<td>0.296</td>
<td>-1.042 3.415</td>
</tr>
<tr>
<td>Phoneme Label</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOT</td>
<td>63.070</td>
<td>39.177</td>
<td>1.610</td>
<td>0.108</td>
<td>-13.804 139.944</td>
</tr>
<tr>
<td>Thought</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-7.437</td>
<td>41.795</td>
<td>-0.178</td>
<td>0.859</td>
<td>-89.448 74.574</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phoneme Label*Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lot*Male</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thought*Female</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thought*Male</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lot*Female</td>
<td>-69.038</td>
<td>21.745</td>
<td>-3.175</td>
<td>0.002</td>
<td>-111.707 -26.368</td>
</tr>
<tr>
<td>Sex*Foll. MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*Fricative</td>
<td>105.658</td>
<td>28.031</td>
<td>3.769</td>
<td>0.000</td>
<td>50.655 160.660</td>
</tr>
<tr>
<td>Female*Stop</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Fricative</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Nasal</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Stop</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*Nasal</td>
<td>-30.264</td>
<td>18.801</td>
<td>-1.610</td>
<td>0.108</td>
<td>-67.156 6.628</td>
</tr>
</tbody>
</table>

Table 5.5: REML-based results of linear regression analysis of merged LOT-THOUGHT in SPSS via GLMM (treatment contrasts), including speaker and word as random effects and a random slope for speaker and phoneme label (n = 1081, deviance = 12,602). The $p$-values of the insignificant results are provided in square brackets.

The intercept estimated for the model. The standard deviation for the interaction (random slope) between speaker and phoneme label is $(31.88 \times 2 =) 63.77$.

The individual findings of the linear regression via GLMM for all factors per above-mentioned factor group are outlined in Table 5.5. In this model, I choose the alphabetically last factor (default in SPSS) in a factor group as the “baseline”, indicated by zero, (e.g. THOUGHT is the last factor in the group phoneme label) and compares the other factors (here LOT) with it (treatment contrast). Within phoneme label, THOUGHT then takes the estimated baseline value, i.e. the intercept in the first row (1366), and LOT is larger than that (1366 + 63 = 1429), i.e. LOT has a larger estimated second formant mean value than THOUGHT (cf. Seltman 2013: 369). The difference of 63 between LOT and THOUGHT is not significant ($p = 0.108$). Likewise, the difference between sexes is not significant, and females have an intercept which is smaller than the baseline intercept (1366 − 7 = 1359). In the interactions, more than one factor is set to zero, which is
5.2. The Low-back Merger

Phoneme Label | Sex | Total
--- | --- | ---
 | Female | Male | Total
LOT | 375 | 359 | 734
THOUGHT | 165 | 182 | 347
Total | 540 | 541 | 1081

Table 5.6: Cross-tabulation of phoneme label and sex.

likely either due to the fact that the model is intrinsically aliased\textsuperscript{100} or that there is no significant difference from the baseline intercept. The results for phoneme label and sex seem to corroborate the findings shown in Subsections 5.2.1 and 5.2.2, as the mean of THOUGHT is smaller than that of LOT and that of females is smaller than that of males.\textsuperscript{101}

Of additional interest is that only the interactions between sex and phoneme label and sex and following manner of articulation are significant. As outlined in Subsection 4.6.2.5, interactions can be caused by empty cells, low token numbers, unevenly distributed data across cells and/or different behavior between the interacting factors. With regard to the interaction of phoneme label and sex, cross-tabulation shows that there is a distributional problem in the cells (cf. Table 5.6).

This problem does most likely not arise between males and females, as the totals in the table show (in per cent: 50\% versus 50\%). The tokens in the cells between the two phoneme labels are skewed with roughly 70\% LOT tokens and 30\% THOUGHT tokens. If this uneven distribution was to cause any factor to be falsely significant in the regression analysis, this factor should, however, be phoneme label. An interaction plot for phoneme label and sex, as well as for sex and following manner of articulation will shed some more light on this (cf. Figure 5.10).\textsuperscript{102} Figure 5.10a shows that all females retract the vowel in LOT tokens beyond the acoustic position of their THOUGHT tokens, indicating an overshoot in the former (cf. Lindblom 1963: 1779). Overshooting of /oh/ in the low-back merger has, for instance, been reported by Hall-Lew (2013) for California. Males, on the other hand, behave quite conservatively with regard to the merger; judging from the graph, their behavior might even be interpreted as distinct between the two vowel classes, but the difference in intercepts was determined to be redundant by the model. In Table 5.5, the intercepts of LOT and THOUGHT for male speakers (in interaction group phoneme label*sex) do most likely not differ significantly from the model’s intercept, as they are both set to zero (the baseline intercept). Similarly, the females’ intercept for THOUGHT is also not significantly different from that of the males. The significant difference is the smaller second formant estimate for females’ LOT ($p = 0.002$): The slopes of both lines

\begin{itemize}
\item Intrinsic aliasing occurs when there are too many variables added to a model, so that a parameter cannot be estimated for a factor and it is thus set to zero (cf. Crawley 2007: 380).
\item The results for phoneme label and sex are merely insignificant tendencies.
\item Note that I intentionally inverted the $y$-axis in this figure to emphasize that females have lower formant two values (i.e. retraction) than males.
\end{itemize}
5. Analysis and Discussion

![Interaction plots from the linear regression of merged LOT-THOUGHT (n = 33).](image)

(a) Interaction plot of phoneme label and sex.

(b) Interaction plot of sex and following manner of articulation.

Figure 5.10: Interaction plots from the linear regression of merged LOT-THOUGHT (n = 33).
in Figure 5.10a show that females behave opposite to males and thus that the females are more advanced in retracting the merger, as the females’ deviance in LOT results in smaller formant two values for both vowel classes. If the interaction was only due to the unevenly distributed cell sizes between LOT and THOUGHT, males’ estimated second formant mean in the latter should be greater and thus closer in value to the former. Put differently, the females behave differently from the males, regardless of the uneven distribution of tokens between vowel classes. This interpretation is supported by running the analysis without the THOUGHT tokens, which neutralizes the skew between LOT and THOUGHT (cf. Appendix F.1.2): The difference between the two sexes naturally remains, and age becomes significant ($t = 2.144$, $p = 0.032$) with a positive intercept (the older speakers are, the higher are their formant values, i.e. fronting of LOT). Retraction of the vowel in LOT words seems to be an apparent-time shift. Investigating LOT in isolation is, however, not sufficient to determine the merger of both vowels in apparent time, as corroborated by the results presented above and those to be presented below.

The interaction between sex and following manner of articulation most likely has quite similar reasons: the distribution between female and male tokens is almost as even as the one for the previously discussed interaction (53% female tokens and 47% male tokens); the distribution of tokens between the manners of articulation nasal, fricative and stop is, however, rather uneven (15% nasals, 22% fricatives and 63% stops). Despite the gap between nasals/fricatives and stops (37% versus 63%), the males do not differ significantly in their estimated formant means between these three manners (cf. interaction group sex*following number of syllables in Table 5.5). Likewise, Figure 5.10b shows quite similar gaps for the males between the manners of articulation. The significant difference is caused by the estimated second formant means for females’ fricatives, as they behave in quite an opposite fashion regarding their LOT-THOUGHT tokens before nasals and stops. Figure 5.10b shows much more retracted vowels before the latter two manners and opposite behavioral patterns between males’ and females’ low back vowels before fricatives. The same patterns are true for the relationship between nasal and stop tokens for both sexes. Although the tokens are unevenly distributed between the manners of articulation, the interaction is rather caused by the difference in behavior between females and males. This notion is further supported by the results of those models with the factor group, following manner, reduced only to stops, as phoneme label remains insignificant (cf. Appendix F.1.3).

This result is quite interesting: According to Labov, Ash and Boberg (2006b: 217), St. John’s, Newfoundland, has a low-back merger before nasals only. Since the males do not pronounce the low back vowels significantly differently for any of the three manners, they are merged before nasals, although their pre-nasal tokens have the most fronted vowel positions. In that sense, my findings corroborate those of Labov, Ash and Boberg (2006b: 217). However, they seemingly contradict Labov, Ash and Boberg’s claim that St. John’s is not merged before /t/. Males insignificantly retract the vowels in pre-stop
position, and even when the (conservative) old speakers are included in the analysis, vowel class does not explain the variation between the low back vowels. I understand these findings in such a way that speakers do not necessarily merge before nasals first, but that they retract the merged vowel before nasals first (closely followed by stops), provided that young urban females from the middle of the social hierarchy are innovators of an ongoing change. This interpretation has to be understood with some caution, as age is not significant in the main model and as the number of my low-back tokens is skewed with regard to nasals and stops. Unlike outlined by Labov, Ash and Boberg (2006b: 221) and Clarke (2010a: 31), my results indicate that the low-back merger is in a back position, at least for female respondents. The results also suggest that interaction between social and linguistic variables does indeed happen, unlike stated by Labov (1994: 84; cf. Paolillo 2002: 66).

Since all of the interactions entered into the model are insignificant, except for the two discussed above (also cf. Appendix F.1), they have to be excluded in order to increase the predictive power of the model (cf. Chatterjee and Simonoff 2013: 23). All of the random effects and the random slope are significant and thus need to remain in the model. In comparison to all the fixed effects, which only added roughly 100 levels to the design matrix, the random effects caused the model to be rather complex (three times as many levels without speaker 20SCOF), so that excluding fixed effects will not necessarily simplify the model much. The independent variables entered are the same as above: phoneme label, age (continuous), sex, number of following syllables, LItotal, following voicing, place and manner of articulation and the two significant interactions (phoneme label*sex and sex*following manner of articulation).

The only difference between the complex model and the simple model in terms of significant factor groups is that following voicing becomes significant ($F_{1,1066} = 6.651, p = 0.01$). The model suggests that an estimated second formant value of 1344 Hz for merged LOT-THOUGHT before voiceless consonants is significantly different from an estimated value of 1368 Hz before voiced consonants (i.e. a difference of 24 Hz in formant two space; model intercept 1232.39). This result may be due to the disregarded interaction between voicing and following place of articulation, which was close to significance in the complex model ($p = 0.08$). However, deviance substantially increased from 12,602 to 12,828 (AIC 12,610 to 12,837), suggesting that the simpler model is a worse fit to the observed data. The levels for the fixed effects decrease to 28, those of the random effects remain unchanged. With all other predictors being equally in-/significant, phoneme label is not significant, despite a smaller $p$-value ($F_{1,1066} = 2.706, p = 0.121$), which corroborates the result that the low back vowels are merged in St. John’s, Newfoundland (95% confidence interval for LOT: 1028.49 and 1389.49; for THOUGHT: 1003.79 and 1300.869).
The Rbrul results for both the complex and simple model are very similar. In the complex model, the same two interactions are significant: phoneme label*sex ($p = 0.003$) and sex*following manner of articulation ($p < 0.001$). The factor groups that are part of an interaction are not assigned any $p$-values, because it is unclear whether significance is due to the interaction or not. The simple model showed an increased deviance (12,877) and following voicing becomes significant ($p = 0.014$).

5.2.4.2 Logistic Regression

The predictors of the logistic regression in SPSS are the same as those of the linear regression; the dependent variable is the major difference, as it is not continuous but binary. The difficulty lies in the appropriate estimation of the two values for a discrete or binary formant two as dependent variable. Results from Decision Trees outlined in Subsection 5.2.3 suggest a threshold value of 1309 Hz, so that the two outcome possibilities are greater than and smaller than 1309 Hz, respectively. This threshold and the one suggested by Labov, Ash and Boberg (2006b: 219) for the Canadian Shift in Canada in his data (1275 Hz) result in models with no significant factor groups whatsoever. The random effects and the fixed effects are the same as outlined for the linear regression above. This result corroborates the one from linear regression and suggests that the deviation of some speakers from a back position of merged LOT-THOUGHT in formant two space is insignificant. In other words, speakers are merged, and especially females retract the merger, which is most advanced before nasals and stops.

Traditionally, studies did not include random effects and random slopes (cf. e.g. Gorman and Johnson 2013). The following paragraphs are thus additions to the regression model presented above in order to serve two purposes: 1) I want to shed more light on the behavior of the predictors when random effects are excluded, which in turn 2) eases comparability to such studies. If the random effects are excluded, we get a nice change-in-progress picture as repeatedly reported in the literature for other vowels in the past. The model was run in SPSS with a threshold of 1309 Hz. The deviance is 1033 (Akaike’s information criterion = 1061) and the model’s design matrix includes only the approximately 100 levels for the fixed effects. Age is modeled continuously and the only significant single predictor ($F_{1, 470} = 13.059, p < 0.001$), i.e. the merger supposedly is a change in progress. In descending $p$-value hierarchy, sex is the next social factor that is significant ($F_{1, 470} = 4.94, p = 0.027$), followed by its two interactions sex*voicing ($F_{1, 470} = 4.727, p = 0.03$) and sex*manner of articulation ($F_{2, 470} = 3.398, p = 0.034$). Within the

Using treatment contrasts as in SPSS, Rbrul returns an AIC of 12,686 and a baseline intercept of 1299 Hz for the complex model. Sum contrasts return an AIC of 12,755 and a baseline intercept of 1255 Hz for the complex model. Rbrul changed the estimation method from REML to ML automatically, so that the absolute numbers will be different, but the relations between factor groups are the same.

Note that Rbrul does not provide the $F$ statistic and the degrees of freedom for the factor groups. The models exclude speaker 20SCOF and were calculated with SPSS, based on treatment contrasts.
5. Analysis and Discussion

<table>
<thead>
<tr>
<th>Model Term</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic $t$</th>
<th>$p$-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.686</td>
<td>1.647</td>
<td>-1.631</td>
<td>0.104</td>
<td>(-5.922, 0.551)</td>
</tr>
<tr>
<td>Age</td>
<td>0.021</td>
<td>0.006</td>
<td>3.614</td>
<td>0.000</td>
<td>(0.010, 0.231)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.130</td>
<td>0.894</td>
<td>-0.145</td>
<td>0.885</td>
<td>(-1.887, 1.627)</td>
</tr>
<tr>
<td>Sex*Voice</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*Voiced</td>
<td>0.782</td>
<td>0.360</td>
<td>2.174</td>
<td>0.03</td>
<td>(0.075, 1.489)</td>
</tr>
<tr>
<td>Female*Voiceless</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Voiced</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Voiceless</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex*Foll. MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*Fricative</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Nasal</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Stop</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*Fricative</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female*Stop</td>
<td>-0.260</td>
<td>0.712</td>
<td>-0.366</td>
<td>0.715</td>
<td>(-1.660, 1.139)</td>
</tr>
<tr>
<td>Female*Nasal</td>
<td>-1.621</td>
<td>0.858</td>
<td>-1.891</td>
<td>0.059</td>
<td>(-3.307, 0.064)</td>
</tr>
<tr>
<td>Place of Articulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labiodental</td>
<td>1.917</td>
<td>1.852</td>
<td>1.035</td>
<td>0.301</td>
<td>(-1.723, 5.556)</td>
</tr>
<tr>
<td>Apical</td>
<td>0.726</td>
<td>0.755</td>
<td>-0.961</td>
<td>0.337</td>
<td>(-0.758, 2.210)</td>
</tr>
<tr>
<td>Labial</td>
<td>0.672</td>
<td>0.718</td>
<td>0.934</td>
<td>0.351</td>
<td>(-0.740, 2.082)</td>
</tr>
<tr>
<td>Velar</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.7: Results from logistic regression analysis of merged LOT-THOUGHT in SPSS. The binary dependent variable is $< 1309$ Hz (application value). All random effects and their interactions are excluded ($n = 1081$, deviance = 1033).

linguistic predictors, only following place of articulation is significant ($F_{3, 470} = 2.677$, $p = 0.047$).

The behavior of the individual factors within these significant factor groups is outlined in Table 5.7. The fact that most of the individual factors are insignificant although their factor group is significant is most likely due to disregarded interactions (e.g. speaker*phoneme label). These are false positive findings, so that the only truly significant results in this model are those for age and the interaction of sex and voicing. Age being significant in the logistic regression does, however, not mean that the low-back merger is a change in progress, but rather that most of the vowels are not realized at the position of 1309 Hz in formant two space. When speaker is set as a random effect, each individual speaker is allowed to deviate from the position of 1309 Hz. Even though this model is inaccurate in terms of disregarding random predictors, the result seems to corroborate the one from linear regression outlined above: some younger speakers are
Figure 5.11: Application rates for merged LOT-THOUGHT in a back position (application value: F2 < 1309 Hz; n = 33).

retracting their merged LOT-THOUGHT vowel – most likely females (and before voiceless consonants; see below). However, the inconsistency in the significance of the results for factor group (variable) versus factor (variant) makes this model meaningless. It serves as an illustration of what happens when random effects are not included in the model. As mentioned above, inclusion of the random effects does not yield any significant factor, which supports the notion that the low-back merger is present in the speech community although some individual speaker differences remain.

The same picture emerges from the fixed-effects model in Rbrul: the results of logistic regression without random effects support those yielded by SPSS without random effects. Continuously-modeled age is significant \( (p = 0.01) \); the interactions of sex and following voicing and of sex and following manner are (marginally) significant \( (p = 0.048 \) and \( p = 0.046 \), respectively); recall that no \( p \)-value is assigned to the individual factor groups that are part of an interaction. Modeling age as a categorical variable and crossing it with sex allows for visualizing the behavior of the corresponding speakers with regard to the position of the merger. However, the resulting model fits the observed data much worse (deviance and AIC increase), and nothing is significant anymore. Figure 5.11 outlines the number of tokens per age/sex group in per cent which are below the threshold of
1309 Hz. These have to be understood as tendencies, rather than significant differences. The percentages show that 56% of the LOT-THOUGHT tokens uttered by young females are below 1309 Hz in formant two space, as opposed to 34% of the young males. The males seem to retract their merged vowel to a lesser degree in apparent time than their female counterparts: 37% of the middle-aged males produce the merged vowel in a back position, compared to 44% of the middle-aged females. Whereas the females display a continuing decrease for second formant values in apparent time, males seem to stop at roughly 33% of retracted tokens. The same model with a threshold of 1275 Hz as suggested by Labov, Ash and Boberg (2006b: 219) yielded lower application rates and the same insignificant factor groups (predictors).

As mentioned above, this interpretation has to be understood with the caveat that age and sex are not significant when each is modeled as a categorical variable. Consequently, the only useful piece of information from this modeling procedure with the present data is that it reveals the application rates of retracted LOT-THOUGHT. The significant results in the three models (one in SPSS, two with Rbrul) without random effects are most likely a consequence of the inter-speaker and inter-word differences in the data set and thus false positives. This receives further support from entering age group instead of age continuously into the model, since the significant results disappear. When the necessary random effects and the random slope for phoneme label are added to the model, no predictor and no interaction are significant in the logistic regression (SPSS). This result also shows that the significance of the interactions phoneme label*sex and sex*manner of articulation that I discussed in the previous subsection for the linear regression does not survive triangulation. Recall that the only difference between the logistic regression and the linear regression in SPSS with random effects was the distribution of the dependent variable, formant two. Consequently, the cause of the interactions in the linear regression is most likely unimportant, because it seems to be limited to the statistical model and does thus probably not have social meaning.

By way of summarizing this whole section, the low-back merger is present in St. John’s, Newfoundland, but for a few old speakers it is still variable, and the retraction of the merged low back vowel is largely due to interactions between sex and linguistic variables, if at all significant. The tendency is that females in general and young females in particular retract the merged vowel more than males do.

### 5.3 Trap Retraction

Recall that with regard to this vowel, the literature’s statements are not as homogeneous as with regard to the low-back merger, but much more studies report a movement of TRAP than do not. Some of the studies that attested no TRAP retraction to young females from St. John’s included only a small number of respondents from St. John’s (cf. e.g. Clarke 2010a: 29; Hollett 2006: 155, 2007: 52-53; Labov, Ash and Boberg 2006b: 221; Reckling
5.3. Trap Retraction


In the first subsection I outline the analysis of TRAP retraction in apparent time in my data via an ANOVA of age group and the Euclidean Distances (EDs) between TRAP and STRUT. As discussed in Subsection 5.1.1, STRUT is stable in apparent time, so that the difference in EDs between TRAP and STRUT per age group can be taken to reflect the shift of TRAP. The indication of movement in apparent time via the EDs is detailed by an assessment of whether first or second formant are responsible for the shifting of TRAP via t-tests and Kendall’s correlation coefficient. The second subsection provides the results of multiple regression analysis followed by those of the logistic regression, based on the categorization of the dependent variable (formant two) via Optimal Binning (and Decision Trees; cf. Appendix E.2.2). Overall, the analysis thus triangulates the results of TRAP retraction in apparent time and extends the findings in terms of the relative contributions of the other (social) predictors to the retraction of TRAP.

5.3.1 Retraction of TRAP across Age Groups

Although all of the vowels that participate in the Canadian Shift are governed by phonological as well as social constraints, the former plays a more important role with regard to TRAP than for any other vowel. In many North American varieties, the low lax vowel is split into two phonemes: a tensed one before nasals and /g/ and a lax one that may participate in the Canadian Shift. This split will be outlined in detail in Subsection 5.3.1.1, including the implications for the analysis of the TRAP vowel in this thesis. As mentioned, Subsections 5.3.1.2 and 5.3.1.3 will establish the movement of TRAP in apparent time and confirm retraction of the vowel as the overarching direction of its movement via t-tests and the significance of Kendall’s τ (tau).

5.3.1.1 Following Phonological Environment

In Subsection 5.1.2, the non-differentiation of TRAP and BATH in St. John’s English was confirmed (in second formant space) in accordance with the literature (e.g. Clarke 2010a: 29). The analysis of TRAP retraction presented here thus includes both phoneme labels and subsumes them under the TRAP lexical set, although the vowel in BATH words is usually lengthened before fricatives and nasals plus consonant (cf. Subsection 3.2.2). The analysis of TRAP is comprised of a total number of 646 tokens, excluding pre-nasal tokens and vowels before /g/, because TRAP tends to be tensed and raised before these
environments in North American varieties of English (cf. e.g. Boberg 2000: 5, 2010: 146; Labov 1991: 5, 1994: 503). According to Labov, Ash and Boberg (2006b: 220), Canadian TRAP (short-\(a\)) before nasals is raised by approximately 100 Hz \((n = 1467)\) compared to non-nasal environments and thus in the same region as in those North American dialects that do not have ash retraction (Labov, Ash and Boberg 2006b: 221). With regard to the distance between TRAP vowels before nasals (and \(/g/\) and non-nasals, Newfoundland has a similar Euclidean Distance to that of the western provinces of Canada (Labov, Ash and Boberg 2006b: 223) where the Canadian Shift is present (2006b: 222). In fact, the distance between the two following environments may even be as great as 300 Hz for speakers from St. John’s, Newfoundland (2006b: 176).

Among others (e.g. De Decker 2002), Boberg (2008b: 135, 2010: 146) confirms the phonetically distinct behavior of TRAP words before nasals and \(/g/\) and suggests separate analyses in order to avoid skewing of the main distribution. Clarke (2010a: 29) does not make this differentiation and states that \(/æ/\) is raised and fronted in Newfoundland English in all environments and most apparently in pre-nasal position. She further maintains that among her two younger speakers, raising and fronting of TRAP is “particularly obvious” in pre-nasal environments (Clarke 2010a: 29). This behavior of ash is precisely the same as that Clarke, Elms and Youssef (1995: 214) found to be consequently hindering the retraction of TRAP in the Canadian Shift in Ontario, although they found TRAP tokens before other environments to be retracted (also cf. D’Arcy 2005: 330). In her (1991) and (2012) publications, Clarke stresses that ash is retracted (and lowered) for some speakers of St. John’s, Newfoundland.

As mentioned in Subsection 4.1, Hollett (2006: 149) excluded tokens in TRAP words before \(/l/\) and \(/r/\), but not those before nasals, \(/g/\) and glides. Although she does not detail in how far her tokens are balanced between undesirable (pre-nasal, before \(/g/\) and glides) and desirable phonological contexts (other), inclusion of undesirable contexts may have skewed the main distribution towards a mean second formant value indicating absence of TRAP retraction \((F2 \approx 2070 \text{ Hz}; \text{Hollett 2006: 154, Table 4, 2007: 52-53})\). Inclusion of these environments thus renders her findings incomparable to those of Boberg’s (2010), Labov, Ash and Boberg’s (2006b) and my analyses. For instance, Boberg’s mean second formant value of \(/æ/\) before nasals is at 2085 Hz \((sd 151 \text{ Hz})\), whereas it is at 1727 Hz \((sd 97 \text{ Hz})\) before other environments (2010: 145, Table 3.12). Labov, Ash and Boberg (2006b: 219) consider TRAP words to have a retracted vowel when the second formant mean is 1825 Hz and less. For all Canadian regions combined, they found a second formant value of 1725 Hz (2006b: 220).

In her impressionistic study of St. John’s, D’Arcy (2005: 332) excluded TRAP tokens before nasals, as her tokens categorically showed neither retraction nor lowering in this environment, which yielded results that corroborate the findings of, for instance, Boberg (2010), Clarke, Elms and Youssef (1995), De Decker (2002) and Labov, Ash and Boberg (2006b). D’Arcy’s results for TRAP in other environments (2005: 337) corroborate those
of Clarke (1991: 30) in that the vowel is lowered and retracted in St. John’s English among young urban females. Striking about D’Arcy’s (2005: 338) results is, however, that ash before /g/ is more often retracted than before its voiceless counterpart and other environments. This contrasts sharply with the suggestions made by Boberg (2010: 146) and Labov, Ash and Boberg (2006b: 223) for other Canadian regions. In addition to the fact that D’Arcy’s analysis is impressionistic, inclusion of ash before /g/ renders her results incomparable to mine.

5.3.1.2 Euclidean Distances

Since the main focus of this analysis is the confirmation of TRAP retraction in St. John’s, Newfoundland, and since Labov, Ash and Boberg (2006b: 223) propose that Newfoundland and the western Canadian provinces behave similarly in terms of the difference between ash in pre-nasal position (and before /g/) versus all other environments, I do not analyze TRAP in those environments where it has not been reported to retract in regions with presence of the Canadian Shift. The 646 tokens I analyze are represented by the Euclidean Distances (ED) of the speaker means between TRAP and STRUT per age group, as shown in the boxplots in Figure 5.12. The EDs thus indicate whether the
5. Analysis and Discussion

<table>
<thead>
<tr>
<th>Test</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global stat</td>
<td>0.38</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.76</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.97</td>
</tr>
<tr>
<td>Link function</td>
<td>1.00</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 5.8: Assumptions of an ANOVA regarding the distribution of age groups and EDs between TRAP and STRUT: results of the global test on four degrees of freedom (n = 34). Level of significance is set to 0.05.

length or distance in a two-dimensional coordinate system (F1xF2) between TRAP and STRUT in each age group is different. If so, an ANOVA can assess whether the difference in variance between the two vowels is statistically significant. Unlike the EDs between LOT and THOUGHT outlined in Subsection 5.2.2, the EDs in relation to STRUT, which is stable in apparent time (cf. Subsection 5.1.1), are a much more direct measure of the distance of TRAP in apparent time, because the literature-based a priori hypothesis (cf. Hypothesis 4 in Subsection 3.4) states that TRAP is shifted at least for the youngest speakers in my sample. That is, the literature suggests that differences in the distances between the vowels TRAP and STRUT are to be expected between younger and older speakers. For LOT-THOUGHT, the EDs were used to establish the merger of the two vowels, which is present but not a change in progress. Consequently, a literature-based a priori hypothesis concerning the merger would state that non-merged vowels can only appear within my old age group, if at all, but not between two age groups such as my old and middle-aged speakers. If there is no change in progress, an apparent-time study design cannot accurately capture differences between speakers or groups of speakers. The Canadian Shift vowel movements are changes in progress, according to the literature. All the limitations of EDs outlined for LOT and THOUGHT, however, hold here as well: despite their difference in range, first and second formant values equally contribute to the EDs, and further, their correlation is disregarded.

The boxplots in Figure 5.12 show that the mean ED for my youngest respondents is almost half the value of my oldest respondents, which clearly shows a shift of TRAP in relation to STRUT in apparent time. The absence of an overlap in the notches between the young and old/middle-aged speakers supports this interpretation. The difference in means between old and middle-aged speakers is comparatively small with a value of roughly 100 Hz, accompanied by an overlap in the notches between these two age groups. The unequal interquartile ranges (sizes of the boxes) of the three data sets suggest heteroscedasticity. The unequal lengths of the whiskers for old and young speakers indicate non-normality of the distribution of the EDs within each age group. However, Peña and Slate’s (2003, 2006) global test for linear model assumptions does not support
the visual impressions. Table 5.8 shows that all of the assumptions are acceptable, except for homogeneity of variances. The \( p \)-value indicates significance at the 95\% level, but is relatively large at 0.04 (i.e. close to insignificance). It is likely to be a false positive result, especially since the global stat, which controls for \( \alpha \)-error inflation, does not yield a significant result.

Results from a monofactorial ANOVA confirm the expectation that the difference in ED means between the age groups is significant: \( F_{2, 31} = 13.99; \ p < 0.001 \). The variable age group explains almost 50\% of the overall variance: multiple \( R^2 = 0.52 \), adjusted \( R^2 = 0.48 \). Pairwise post-hoc comparisons of the means (Tukey’s HSD) show that two out of three means are significantly different from each other: the youngest speakers’ ED mean differs very significantly from the middle-aged speakers (\( p = 0.01 \)) and and even more so from the old speakers (\( p < 0.001 \)). The difference in means between the latter two age groups is insignificant (\( p = 0.21 \)). These results clearly show that TRAP is shifted in relation to STRUT in apparent time in my data set. The vowel plot in Figure 5.5 above and the relative position of STRUT in relation to TRAP suggest that young middle-class speakers of St. John’s English retract ash in positions other than before nasals and /g/. This is not to say that ash is exclusively retracted, but that this movement in formant space is predominant.

These results do not per sé contradict those of Hollett (2006: 155, 2007: 52-53) with regard to TRAP; instead they seem to corroborate hers despite the apparent incomparability outlined in Subsection 4.1 and inclusion of male speakers in my analysis. Hollett’s young speaker cohort recorded in 2003 corresponds to my middle-aged speakers and her other two cohorts correspond to my old respondents. Thus similar to Hollett (2006: 155, 2007: 52-53), I find no significant movement in vowel space between my middle-aged and old speakers, but only when EDs between TRAP and STRUT are statistically tested between age groups via an ANOVA. TRAP retraction thus seems to be a very recent development in middle-class St. John’s English, according to this method (but cf. Subsection 5.3.2).

D’Arcy (2005: 338) found small percentages of TRAP retraction among young females in her impressionistic study on St. John’s (cf. Subsection 4.1), which may be understood in such a way that the sound change does not occur abruptly but gradually within the speech community. The apparent-time analysis presented in this subsection cannot account for the gradual movement of innovative features into the community, i.e. small percentages of innovative feature use per speaker. Such a more fine-grained analysis is provided in Subsection 5.3.2 below.
5.3.1.3 TRAP Retraction versus Lowering

In order to assess whether TRAP retracts more than it lowers, simple \( t \)-tests for independent samples\(^{106} \) of the first and second formant means between my youngest and oldest speakers suffice, given the results from using EDs and from other studies outlined above. The first formant mean of the vowel in TRAP words realized by my old speakers is 774 Hz, the average first formant mean of TRAPs produced by my young speakers is 802 Hz. The difference of 28 Hz between the means of these two age groups is not significant: \( t_{\text{Welch}} = -0.875, \ df = 17.568, \ p_{\text{two-tailed}} = 0.4, \) Cohen’s \( d = 0.45 \). With regard to the second formant means of TRAP, the old speakers produced the vowels at 1952 Hz on average and the young speakers at 1725 Hz on average. The difference of 227 Hz between these means is statistically highly significant and strong: \( t_{\text{Welch}} = 6.839, \ df = 13.943, \ p_{\text{two-tailed}} < 0.001, \) Cohen’s \( d = 0.953 \). Lowering of TRAP among the young speakers is thus rather a tendency that accompanies retraction, as only that is highly significant between the two age groups (see below).

Since lumping all speakers’ means into two groups makes the result rather general, statistical assessment of the correlation coefficients between the ages of the individual speakers and their first formant mean values sheds some more light on possible fine-grained differences between individual speaker means. Figure 5.13 shows that there is a low negative correlation between the individual speakers’ mean first formant values and their ages.\(^{107} \) The older speakers are, the smaller are their first formant values, i.e. their ashes are not lowered. The regression line shows that it does not summarize the distribution of the data, i.e. age does not explain the range in different means. If we extended the data set by speakers older than 60 years and add them on the right of the figure, we could continue the line in order to predict the formant values for those older speakers. Extension of the regression line suggests that those speakers will probably have first formant values of 770 to 760 Hz. This prediction is, however, inaccurate, considering the great spread in formant values above and below the regression line, so that any value between 700 and 900 Hz is possible. The negative correlation is not significant: Kendall’s \( \tau = -0.1, \ z = -0.789, \ p = 0.43 \). These results corroborate those of the \( t \)-test for independent samples per age group above. TRAP does lower in apparent time to some extent in my data, but not significantly so.

Having stated that, the test result seems to be a direct result of the three older speakers’ behavior, one male (32RROM) and two females (20SCOF and 29CCOF), as a visual inspection of the graph suggests. This does, however, not mean that Clarke’s (2012) and Hollett’s (2006, 2007) statements of old females from St. John’s being most innovative in terms of TRAP are corroborated, because the dominant movement and the

\(^{106} \) The assumptions are met and can be found in Appendix D.3.1. The variances between the young and old speakers are homogeneous, indicating that the global tests result was largely due to the influence of the variance in the middle-aged speaker EDs.

\(^{107} \) The assumption of normality in the distribution of age is violated (\( W = 0.857, \ p < 0.001 \)), so that I use Kendall’s \( \tau \) instead of Pearson’s \( r \) as correlation coefficient.
defining criterion of TRAP in order to be part of the Canadian Shift is its retraction (cf. Section 2.3). The positive correlation between second formant means and age is highly significant: Kendall’s $\tau = 0.51$, $z = 3.4$, $p < 0.001$. If the three old speakers really cause the correlation test of age and F1 (lowering of TRAP) to be insignificant, then they do not do so for F2 (retraction of TRAP). Further visual inspection of the graph shows that none out of the 13 middle-aged speakers shows mean F1 values higher than approximately 810 Hz. Since out of 22 older speakers (13 middle-aged and 9 old) only three old speakers show mean F1 values greater than 810 Hz, they seem to be outliers in my data set, rather than reflections of socially meaningful differences. Consequently, it may well be that lowering of TRAP was also significant in my data if these three speakers were excluded, which would be in line with the findings of other studies on the Canadian Shift in other communities. Whether lowering of TRAP is significant or not is, however, unimportant in this study, because, once more, the major role of TRAP in the Canadian Shift is its retraction.

For Boberg (2010: 147), it is of particular importance to emphasize that his mean value for TRAP (1727 Hz, sd 97 Hz; 2010: 145) conforms to Labov, Ash and Boberg’s
(2006b: 130, 151, 219) threshold of 1825 Hz. At 1725 Hz (sd 45 Hz), my youngest respondents have a second formant mean value identical in number to that proposed by Labov, Ash and Boberg (1725 Hz; 2006b: 220) which they report for mainland Canada from coast to coast. Retraction of TRAP, with (or without) accompanying lowering, is consistent with Labov’s peripherality theory derived from his principles of vowel chain shifts (1991: 4-12, 1994: 116-284; Labov, Ash and Boberg: 16-20; also cf. Section 2.2). Before any sound changes commence, TRAP is initially positioned at the lower end of the non-peripheral track, so that its possible (unidirectional) movements are: lowering and retraction to an [a] value before linguistic environments other than nasals and /g/ while remaining in its position on the non-peripheral track (Principle II); raising and fronting (Principle I) after TRAP has shifted from the non-peripheral to the peripheral track when realized before nasals and /g/ (Principle IV; cf. Boberg 2000: 5; Labov 1991: 7-8, 1994: 176, 280; Labov, Ash and Boberg 2006b: 16,18).

5.3.2 Generalized Mixed-effects Modeling of TRAP Retraction

Although a total of 646 tokens of TRAP is not too low for modeling factor groups or independent variables simultaneously, this proved to be particularly troublesome with the following consonant classifications place/manner of articulation and glottal state, as there were some empty cells. Cross-tabulation and Bonferroni-corrected multiple comparisons of between-factor formant two means helped to reduce the number of factors in each group (cf. Subsection 4.6.3.4). For some consonants, the means were insignificantly different according to their manner/place, for others according to their glottal state. The ultimate factor group which I entered into the model is an interaction group of all three classifications, referred to as ‘Voice PoA MoA’ (Glottal State, Place of Articulation, Manner of Articulation). It consists of only five remaining factors: voiced fricatives, voiced stops, voiceless apicals, voiceless labiodentals and voiceless velar stops.

Linear and logistic regression models are run with the Bonferroni approximation. The estimates in linear regression in SPSS are based on the REML method (Residual Maximum Likelihood) and the model shows treatment contrasts (one factor versus all other factors per factor group; Johnson 2009: 361). The dependent variable is formant two, as the discussion in Section 2.3 has shown TRAP retraction to be the dominant movement in the Canadian Shift. The independent fixed effects are: age (continuous), sex, local-ness index total (LItotal; continuous), number of following syllables, style and the crossed category of voicing, place and manner of articulation (Voice PoA MoA; cf. Labov, Ash and Boberg 2006b: 220). Age and LItotal are centered on the minimum value of each variable (age = 20, LItotal = 1.5). None of the interactions between these effects were significant in the full model (deviance 7526, Akaike’s information criterion 7534), so that I do not discuss them any further below. The random effects taken into consideration are speaker and lexical item.
Logistic regression analysis consists of the same factor groups as the ones outlined above. The difference to the linear regression is the dependent variable, which I categorized according to the Optimal Binning result (and to a lesser degree according to the Decision Tree result; cf. Appendix E.2.1) outlined in Appendix E.2.2. The procedure identified three bins at which a categorization of the continuous dependent variable F2 differs significantly with regard to age group. The first bin is marked by an upper end point of 1786 Hz, which includes most of the tokens of the youngest age group. The second bin is marked by an upper end point of 1935 Hz, which in turn serves as the lower end point of the last bin. Most of the tokens of middle-aged and old speakers fall within the latter two bins. The upper end point of the first bin, < 1786 Hz, consequently serves as the threshold (application value) in the logistic regression discussed below. This threshold is more conservative than the one suggested by Labov, Ash and Boberg (2006b: 220) of 1825 Hz, but it is also not as back as the mainland Canadian formant two of 1725 Hz in their data. Although the independent fixed effects in the model are the same as outlined above, age is modeled categorically in groups, which is in turn crossed with sex in order to trace the change in progress more accurately. The estimates of Rbrul (Johnson 2009) are based on Maximum Likelihood (ML; automatically altered), instead of REML. The logistic regression is further based on sum contrasts (each factor is compared to an optimal zero-sum that represents the deviation from the mean; Johnson 2009: 361).

5.3.2.1 Linear Regression

The simple model’s deviance is higher than that of the full model (7755, AIC 7761; Nakagawa and Schielzeth’s [2013] marginal $R^2 = 0.394$; henceforth $R^2_{\text{GLMM}(m)}$) as in the case of merged LOT-THOUGHT. The regression analysis of the merger showed, however, that none of the fixed effects were significant. For TRAP, the REML-based results indicate a change in progress with age as the most significant fixed effect: $F_{1, 634} = 26.772$, $p < 0.001$. Sex is the only other significant effect with a much smaller $p$-value than that of age: $F_{1, 634} = 6.041$, $p = 0.014$. The crossed category of voicing, place and manner of articulation is close to significance ($F_{4, 634} = 2.154$, $p = 0.073$), whereas local-ness, number of following syllables and style are unimportant predictors for the model ($p$-values of 0.55 to 0.99). Both of the random effects are significant: for speaker, Wald’s Z is 2.709, $p = 0.007$, and for lexical item, Wald’s Z is 2.836, $p = 0.05$.

Table 5.9 shows the individual factors based on treatment contrasts: The factor with the alphabetically first letter is set to zero, and the other factors are compared to the selected one. In terms of the social predictors, age shows a strong and significant positive coefficient, i.e. the older speakers are, the greater are their second formant values for

Note that in the model with following glottal state, place of articulation and manner of articulation as separate predictor variables and not as an interaction group, manner and voicing are significant. Since I cannot rule out the possibility that this significance is owed entirely to the empty cells (e.g. TRAP before voiced velar stops is excluded), I do not present this model here.
Table 5.9: REML-based results of linear regression analysis of \textit{trap} in SPSS via GLMM (treatment contrasts), including speaker and word as random effects (n = 646, deviance = 7755). The \textit{p}-values of the insignificant results are provided in square brackets.

<table>
<thead>
<tr>
<th>Model Term</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic $t$</th>
<th>\textit{p}-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1720.272</td>
<td>47.067</td>
<td>36.550</td>
<td>0.000</td>
<td>1627.847 - 1812.698</td>
</tr>
<tr>
<td>Age</td>
<td>5.484</td>
<td>1.060</td>
<td>5.174</td>
<td>0.000</td>
<td>3.403 - 7.565</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>77.811</td>
<td>31.659</td>
<td>2.458</td>
<td>0.014</td>
<td>15.641 - 139.981</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voice PoA MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced Stops</td>
<td>0.339</td>
<td>26.372</td>
<td>0.013</td>
<td>[0.990]</td>
<td>-51.448 - 52.126</td>
</tr>
<tr>
<td>Voiced Fricatives</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless Apicals</td>
<td>-36.819</td>
<td>30.129</td>
<td>-1.283</td>
<td>[0.200]</td>
<td>-87.510 - 13.872</td>
</tr>
<tr>
<td>Voiceless Labials</td>
<td>-38.658</td>
<td>25.814</td>
<td>-1.426</td>
<td>[0.154]</td>
<td>-97.822 - 20.506</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word List</td>
<td>44.221</td>
<td>50.433</td>
<td>0.877</td>
<td>[0.381]</td>
<td>-54.815 - 143.258</td>
</tr>
<tr>
<td>Reading Passage</td>
<td>9.983</td>
<td>14.588</td>
<td>0.684</td>
<td>[0.494]</td>
<td>-18.663 - 38.628</td>
</tr>
<tr>
<td>Interview</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li_{total}</td>
<td>4.457</td>
<td>12.426</td>
<td>0.359</td>
<td>[0.720]</td>
<td>-19.944 - 28.857</td>
</tr>
<tr>
<td>Number of Foll. Syllables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-1.435</td>
<td>14.607</td>
<td>-0.098</td>
<td>[0.922]</td>
<td>-30.119 - 27.249</td>
</tr>
<tr>
<td>2+</td>
<td>-3.137</td>
<td>17.303</td>
<td>-0.181</td>
<td>[0.856]</td>
<td>-37.115 - 30.840</td>
</tr>
</tbody>
</table>

In other words, the younger speakers are, the more retraction they show (smaller F2 values) of that vowel. In addition, females have significantly lower second formant means than males. These results corroborate those for mainland Canadian English, e.g. Boberg (2005: 142) for Montreal, Clarke, Elms and Youssef (1995: 217) for Ontario, Esling and Warkentyne (1993: 240) for Vancouver, Hoffman (2010: 131) and Hoffman and Walker (2010: 56) for Toronto, Roeder (2012: 484) most recently for Toronto, Ontario, and Woods (1979: 152-153) for Ottawa, although the effect of age was not very strong in the former two studies. More importantly, my results are also in line with those of Boberg (2010: 240) for his 32-year-old female case study from St. John’s, Newfoundland (also cf. Clarke 2004b: 371, 2012: 514).

To some extent, my results also corroborate those of D’Arcy (2005: 346) for St. John’s, Newfoundland, although she finds higher retraction rates for her adolescent females than for her pre-adolescent females. She explains this difference with the instability of the pre-adolescents’ sociolinguistic personae, i.e. they are more influenced by their parents than adolescent respondents, who are rather influenced by their peers (social identity).
This explanation is, for instance, supported by Bailey (2002: 319) and Cukor-Avila and Bailey (2013: 246), but only for linguistic change at the level of morphology-syntax and discourse-pragmatics, and only for speakers of 20 years of age and older. In contrast to those linguistic levels, the phonological system is already stable for speakers younger than 20 years, according to Labov (1994: 111-112). D’Arcy’s adolescents are 16 to 17 years old (2005: 332). The (exclusively) conservative nature of TRAP realizations for young and middle-aged speakers suggested by Clarke (2010a: 30), Hollett (2006: 154-155) and Labov, Ash and Boberg (2006b: 221) for St. John’s, Newfoundland cannot be supported with my data.

With regard to the fourth social predictor, the local-ness index (which consists of the birth places of the respondents and their parents, as well as schools and colleges they visited in St. John’s), the analysis does not yield significant findings. The evidence for a classic change in progress is compelling: Young (urban) females lead in the use of the innovative retracted TRAP variant in St. John’s, regardless of whether their parents are from other provinces or whether they have visited schools outside Newfoundland for a longer period of time. TRAP retraction is not a function of non-local-ness, i.e. TRAP is retracted although the majority of my respondents is local to the speech community. In this regard, my results do not confirm D’Arcy’s (2005: 341-342), as she found a significantly higher application rate of TRAP retraction (and lowering) among her young females of non-local parentage (21%) than among those of local parentage (9%). Since she exclusively analyzed pre-adolescent and adolescent females, a different picture might emerge when my young female respondents are analyzed in isolation.

The result of such a model shows that the only significant effect is the third social predictor style ($F_{2, 163} = 4.246, p = 0.016$), with word list having a significantly higher second formant mean, i.e. TRAP is in a significantly more fronted position in word-list style (cf. Appendix F.2.2). This finding suggests that the innovative retracted TRAP variant is present in the vernacular of respondents from St. John’s, after it might have originally entered the phonological system via the formal styles, as maintained, for instance, by Clarke (2012: 514). However, this analysis includes only 174 tokens, so that D’Arcy’s 801 tokens may provide a more reliable result (2005: 338). And yet, her analysis was impressionistic only. In addition, as shown in Table 5.9, although style is not significant in the full model, the tendency is the same: the most retracted variants of /æ/ are produced in spontaneous speech/interview style, with a stepwise decrease in second formant values per increase in stylistic formality. Since this is merely a tendency, it cannot serve as the only basis for the interpretation that the innovative TRAP variant was not introduced via formal styles into the system of St. John’s English.

In the original study on the Canadian Shift in Ontario, Clarke, Elms and Youssef (1995: 215) found that the only significant linguistic factor to affect the shift of TRAP was following manner of articulation. This is not completely replicable in my analysis, as I combined the individual factors to a crossed factor group. However, similar tendencies
(following Voice PoA MoA is not significant) are visible in Table 5.9. As mentioned above, testing of significant differences between the mean second formant values per voicing, place and manner of articulation showed that for two remaining factors, stops and fricatives, manner was the most prominent determinant for the similarity of the means. Since nasals are generally excluded in my analysis, the factor stops consists of oral stops only.

In Clarke, Elms and Youssef’s (1995: 215) data, oral stops slightly disfavored retraction of /æ/, whereas fricatives favored it. Following glottal state and place of articulation was not significant in their data, but the tendency for voicing was that voiced consonants disfavor and voiceless ones favored retraction. The same picture emerges in my analysis in Table 5.9 for voiced oral stops and contradicts the findings of D’Arcy (2005: 339) for St. John’s. Her results indicate that manner was not significant; instead, glottal state was significant, with voiced sounds favoring TRAP retraction. In terms of glottal state, voicing was primarily responsible for the classification into voiced fricatives and stops in my data, i.e. it did have a stronger influence than manner alone. In line with Clarke, Elms and Youssef (1995: 215), voiceless sounds have lower second formant values of TRAP (i.e. they favor retraction) in my data. With regard to following place of articulation, prominence was exhibited by apical and labial sounds; velar pronunciations were only prominent for oral stops. Since fricatives in English occur only in alveolar and fronter positions, and since (voiceless) apical and labial sounds have lower second formant values of /æ/ in my data, they offer some support for the findings of Clarke, Elms and Youssef (1995: 215), with fricatives favoring retraction. Although not significant, (labio- and inter-)dental and velar sounds also favored retraction in their data, unlike alveolar ones. My data seems to offer some support for this finding with (voiceless) velar, apical and labiodental sounds having lower second formant values. D’Arcy’s (2005: 339) findings differ insofar that following place of articulation is significant in her data of young female St. John’s English speakers, and that interdentals disfavor innovative TRAP. In terms of velar, alveolar and labial sounds, my results are in line with hers.

None of the studies mentioned above take vowel duration into account. Since unstressed and thus short vowels are known to centralize, they would conform to the pattern of TRAP discussed here. However, each of the studies discussed above included only impressionistically stressed vowels in their analyses. In addition to that, Labov, Ash and Boberg (2006b) included number of following syllables as a factor group in their analysis, which substitutes vowel duration insofar as vowels are shorter in multisyllabic lexical items, when they are stressed before following syllables, i.e. vowels in monosyllabic environments are usually much longer than those in multisyllabic words. This factor group is not significant in my data, but as Table 5.9 shows, TRAP has continuously lower second formant means the more syllables are following the stressed vowel; i.e. the shorter the vowel, the more retracted it is.

---

109 Laterals favor retraction of TRAP even more in Clarke, Elms and Youssef’s (1995: 215) data and are excluded in my analysis for this reason, among others (cf. Subsection 4.4.5).
### 5.3. Trap Retraction

#### 5.3.2.2 Logistic Regression

The model’s deviance is 519 (AIC 551; \( R^2_{\text{GLMM}(m)} = 0.434 \) and Nakagawa and Schielzeth’s conditional \( R^2 = 0.659 \); henceforth \( R^2_{\text{GLMM}(c)} \)) with 16 degrees of freedom and a grand mean of 0.395, i.e. 39.5% of all tokens in the model (646) are below the threshold of 1786 Hz in value. The intercept of the model is relatively small at \(-1.166\); it represents the grand mean of the predictions for all cells (Johnson fc.: 37). Similar to the results of the linear regression outlined above, the only and highly significant factor group is the crossed predictor of age group and sex \((p < 0.001)\). The interaction group Voice PoA MoA is much further away from the significance threshold of \( \alpha = 0.05 \) than in the linear regression \((p = 0.172)\). The remaining predictors are even less significant: style \((p = 0.6)\), number of following syllables \((p = 0.74)\) and LItotal \((p = 0.953)\).

Table 5.10 shows the individual factors of each predictor with its corresponding log odds and factor weights. The factor weights in the last column primarily serve the purpose of comparability to the VARBRUL outcomes. The log odds are shown in the first column, followed by the number of tokens and their application rates in the third column, i.e. the

<table>
<thead>
<tr>
<th>Factor</th>
<th>Log Odds</th>
<th>Tokens</th>
<th>Application Rates</th>
<th>Centered Factor Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age and Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young Females</td>
<td>2.694</td>
<td>175</td>
<td>0.794</td>
<td>0.937</td>
</tr>
<tr>
<td>Middle Females</td>
<td>1.475</td>
<td>121</td>
<td>0.537</td>
<td>0.814</td>
</tr>
<tr>
<td>Young Males</td>
<td>1.293</td>
<td>54</td>
<td>0.481</td>
<td>0.785</td>
</tr>
<tr>
<td>Old Females</td>
<td>-1.649</td>
<td>77</td>
<td>0.091</td>
<td>0.161</td>
</tr>
<tr>
<td>Middle Males</td>
<td>-1.790</td>
<td>141</td>
<td>0.085</td>
<td>0.143</td>
</tr>
<tr>
<td>Old Males</td>
<td>-2.024</td>
<td>78</td>
<td>0.077</td>
<td>0.117</td>
</tr>
<tr>
<td>Voice PoA MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless Labials</td>
<td>0.549</td>
<td>44</td>
<td>0.545</td>
<td>0.634</td>
</tr>
<tr>
<td>Voiceless Velar Stops</td>
<td>0.336</td>
<td>193</td>
<td>0.420</td>
<td>0.583</td>
</tr>
<tr>
<td>Voiceless Apicals</td>
<td>0.230</td>
<td>191</td>
<td>0.382</td>
<td>0.557</td>
</tr>
<tr>
<td>Voiced Stops</td>
<td>-0.510</td>
<td>146</td>
<td>0.390</td>
<td>0.375</td>
</tr>
<tr>
<td>Voiced Fricatives</td>
<td>-0.605</td>
<td>72</td>
<td>0.278</td>
<td>0.353</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>0.101</td>
<td>503</td>
<td>0.396</td>
<td>0.525</td>
</tr>
<tr>
<td>Formal</td>
<td>-0.101</td>
<td>143</td>
<td>0.392</td>
<td>0.475</td>
</tr>
<tr>
<td>Number of Foll. Syllables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2+</td>
<td>0.138</td>
<td>98</td>
<td>0.439</td>
<td>0.534</td>
</tr>
<tr>
<td>0</td>
<td>0.049</td>
<td>313</td>
<td>0.396</td>
<td>0.512</td>
</tr>
<tr>
<td>1</td>
<td>-0.186</td>
<td>235</td>
<td>0.374</td>
<td>0.454</td>
</tr>
<tr>
<td>LItotal +1</td>
<td>-0.016</td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Table 5.10: ML-based results of linear regression analysis of TRAP in Rbrul via GLMM (sum contrasts), including speaker and word as random effects \((n = 646, \text{deviance} = 519)\). The factor weights within the insignificant factor groups are provided in square brackets.
5. Analysis and Discussion

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Application Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old</td>
<td>20</td>
</tr>
<tr>
<td>Middle</td>
<td>40</td>
</tr>
<tr>
<td>Young</td>
<td>60</td>
</tr>
</tbody>
</table>

Figure 5.14: Application rates for TRAP retraction (application value: F2 < 1786 Hz; n = 34).

number of tokens per factor that conform to the application value of < 1786 Hz. The log odds indicate whether a factor favors (positive log odd) or disfavors (negative log odd) the application value.

Within the crossed Voice PoA MoA category, an interaction is visible in the non-linear decrease in application rates. 38.2% of the tokens before voiceless apicals fall within the range of the application value, compared to 39% of the tokens before voiced stops, although the coefficient indicates that the latter disfavors retraction of TRAP. The interaction is most likely caused by the relatively large number of factors in the two factor groups: six factors in age and sex and five factors in Voice PoA MoA. Cross-tabulation shows that in four cells there are fewer than 10 tokens, especially for young male respondents and voiceless labials. Fortunately, the interaction does not seem to skew the results too badly, as they are quite similar to those outlined above for the linear regression. In the latter, age was modeled continuously and sex only has two factors, male and female. Due to the fact that following glottal state, place and manner of articulation are not significant when crossed to one predictor, I do not exclude those tokens/factors that cause the interactions. I did, however, recode style as a binary variable (interview
5.3. Trap Retraction

style and formal style) to reduce the total number of cells and thus the potential for further interactions.

The crossed factor group age and sex provides a better picture of the ongoing change in progress with regard to TRAP retraction than the one painted by the results of the linear regression: It appears that female speakers are clearly one apparent-time generation ahead of their male counterparts. D’Arcy’s (2005: 338) retraction rates of 9% for young females of local parentage and 21% for young females of non-local parentage in St. John’s cannot be supported. Young females show a retraction rate of almost 80%, i.e. more than three-fourths of their tokens have a second formant value of less than 1786 Hz. The young females are closely followed by their middle-aged like-gendered predecessors (54%) and their young male counterparts (48%). In stark contrast to this, the middle-aged male speakers show an application rate of retracted tokens of 8.5% and the old speakers of 9% (females) and 8% (males), respectively.

This explains the findings in Subsection 5.3.1.2: middle-aged speakers patterned with the old speakers, so that only young speakers significantly differed in their Euclidean Distances from the other two age groups. Although middle-aged females are clearly more advanced in retraction of TRAP, the conservative behavior of middle-aged males is too strong for the whole age group in order to yield significant differences to the old speakers. The relationship between the six speaker groups is visualized in Figure 5.14. The generational gap between middle-ages females and young males is most apparent in that the former group is more advanced in retracting TRAP than the latter. However, it has to be kept in mind that these application rates are a direct result of the application value, which in turn is more or less arbitrarily selected. This is not only true for the present analysis, but also for the the atlas of Labov, Ash and Boberg (2006b). More importantly, the mean second formant value for TRAP of young females in my data is 1722 Hz (sd 52 Hz), which is nearly identical to that shown by Boberg (2010: 145) of 1727 Hz (sd 97 Hz) and by Labov, Ash and Boberg (2006b: 219) of 1725 Hz for the entire Canadian mainland. Young males have an average second formant mean of 1736 Hz (sd 47 Hz) in my data.

With regard to following consonants, a similar picture to that of the linear regression discussed above emerges. In general, the findings are insignificant, so that they can only be understood as tendencies rather than significant correlates. A direct comparison of the order of the individual factors in the logistic regression with that of the linear regression reveals that they differ: In the linear regression, voiced stops had a slightly higher F2 value than voiced fricatives, and voiceless apicals had slightly lower F2 values than voiceless velar stops. In the logistic regression (Table 5.10), voiced fricatives disfavor the application value of < 1786 Hz more than voiced stops, and voiceless velar stops favor the application value more than voiceless apicals. Since this factor group is insignificant, these differences are tendencies; they may, however, also be due to the method, i.e. sum contrasts, or to the low token number in some cells. Very generally, TRAP before voiceless
consonants has lower F2 values than before voiced consonants, and the former favors the application value, the latter disfavors it. In this regard and in terms of significance, both regression models yield the same results for this factor group.

Style and LI_total are insignificant in both regression analyses, but there is a tendency for interview style (spontaneous speech) to favor retraction of TRAP. Solely on the basis of my data, a claim that the innovative production of TRAP is introduced to the system of St. John’s English via the formal styles is too strong a claim to make. Since my data consist largely of spontaneous speech tokens, and since style is not significant, it seems that TRAP retraction has entered the vernacular of young and middle-aged females as well as young males. If we add Clarke’s results (1991: 116) to this finding, a claim that TRAP retraction has originally entered the system of St. John’s young middle-class via the formal styles seems less doubtful. Local-ness does not play a role with regard to ash retraction; speakers very local to and thus more representative of St. John’s, Newfoundland, retract their TRAPS. I do not consider this finding to contradict D’Arcy’s, but rather to reflect the lack in comparability between my study and hers. Finally, the result for the variable number of following syllables differs slightly from the result of the linear regression: Both analyses yield insignificant results for this predictor, but the order of factors differs: zero and more than two following syllables favor retracted /æ/; one following syllable disfavors retracted TRAP in the logistic regression. I understand this result as that it proves this predictor to be insignificantly contributing to the explanation of the binary dependent variable, i.e. number of following syllables does not have any effect.

When the mainland Canadian mean second formant value of 1725 Hz (cf. Boberg 2010: 145; Labov, Ash and Boberg 2006b: 220) is set as the application value in my data, almost none of the tokens below this threshold is uttered by males: old males utter one token, middle-aged males two and young males 16. Similarly, the old females only produce six tokens that have a second formant value of less than 1725 Hz. The majority of the TRAP tokens for young females, however, easily reaches this application value (116 versus 59, application rate of 66.3%). The skewed distribution results in interactions when this model is run, so that it is unreliable in its results. It does, however, show that younger females approach the average mainland Canadian version of retracted TRAP, since they would otherwise categorically show no tokens below the threshold of 1725 Hz (cf. Appendix F.3).

5.4 The Shift of DRESS

Recent literature on the innovative status of DRESS (and KIT) in St. John’s English is sparse, unlike that on the low-back merger and the retraction of TRAP. Neither Clarke (1991, 2010a, 2012) nor Labov, Ash and Boberg (2006b) support any shift of DRESS in the fashion proposed by the Canadian Shift in St. John’s. Boberg (2008b, 2010) im-
5.4. The Shift of Dress

Plicitly acknowledges the possibility of innovative Canadian mainland behavior of dress in Newfoundland, but does not explicitly position himself with regard to his data from St. John’s, Newfoundland. Clarke (2004b: 370) likewise acknowledges a “lowered and somewhat retracted pronunciation of [...] dress [...]” in the speech of upwardly mobile younger urban Newfoundlanders”, which “reflects the influence of the innovative CanE tendency described as the ‘Canadian Shift’ [...]”. As indicated above, this statement leads, however, quite a solitary, if not at least ambivalent, life within Clarke’s publications ever since the 1980s. Hollett (2006: 156, 2007: 54) finds some support for dress lowering among her young speakers recorded in 2003 (first formant mean of 668 Hz), but not for dress retraction (second formant mean of 2136 Hz). This finding for St. John’s is diametrically opposed to that of Boberg (2005: 137, 143) for Montreal, who finds only retraction of dress.

The latter finding also runs counter the predictions of the principles of vowel chain shifts, given the position of dress in the middle of the non-peripheral track. Based on these findings, Boberg (2005: 136, 2010: 227) suggests that dress retracts parallel to trap, indicating parallel shifts rather than a classic chain shift. Among the other Atlantic Provinces in Canada, some studies do indeed find Canadian Shift patterns. Contrary to Kieft and Kay-Raining Bird (2010: 64) statements, Sadlier-Brown and Tamminga (2008) find lowering and retraction of dress in Halifax, Nova Scotia (and Vancouver). In a conference paper, Gardner and Childs (2011) report a Canadian Shift pattern in Petty Harbour, Newfoundland. The results outlined below will be discussed in light of those studies that used a methodology similar to the one I employed.

The following subsection will thus establish the movement of dress in apparent time via an ANOVA of age groups and Euclidean Distances (EDs) between dress and strut. The Decision Tree results suggest that age is the most important variable to influence formant one (cf. Appendix E.3.1). In order to assess the relative contribution of first and second formants, correlation coefficients will indicate which formant has a statistically significant relationship with age in order to establish a trajectory of the change (cf. Boberg 2005: 135). In the final subsection, the results for the age groups will be extended and further triangulated via multiple regression analyses. The dependent variable for these analyses is chosen on the basis of the significance of the correlation between age and formant values as well as the literature. The subsequent logistic regression is based on the categorization of the dependent variable via Optimal Binning and Decision Trees (cf. Appendix E.3).

Note that Boberg’s article (2008b) and his book (2010) are essentially based on the same data set and thus offer similar results and draw similar conclusions. The vowel plot of his case study from St. John’s, Newfoundland, does not include the positions of either dress or kit, and he does not mention their roles at all (cf. Boberg 2010: 240).
5.4.1 DRESS Movement across Age Groups

The apparent-time analysis of DRESS presented here includes a total number of 1046 tokens in all phonological environments, except pre-nasal positions and the other categorical exclusions (cf. Labov, Ash and Boberg 2006b: 77), which is roughly the same amount of tokens Labov, Ash and Boberg (2006b: 221) used for their statistical assessment of the 33 acoustically-analyzed Canadian speakers (n = 949). While following nasal environments generally inhibit retraction of TRAP (cf. Subsection 5.3.1.1), they tended to particularly promote DRESS lowering in the study of young Ontarians by Clarke, Elms and Youssef (1995: 216). Labov, Ash and Boberg may have additionally decided to exclude those tokens, because they were interested in the dialect regions of the entire North American continent. Some of these regions show presence of conditioned mergers before nasals (the *pin-pen* merger), predominantly in the southern dialect areas of the United States, but continually spreading northwards by diffusion (as far north as Saint John, New Brunswick, and Halifax, Nova Scotia; Labov, Ash and Boberg 2006b: 220; also cf. Labov 1994: 197 and Thomas 2001: 52 for effects of coarticulation of vowels before nasals). In the course of the merger, epsilons in pre-nasal position are raised and fronted, whereas short-i (K IT) is lowered and retracted.

Boberg’s data set from Montreal, Quebec, does not include tokens before liquids for the final analysis (2005: 152), and neither does that of Labov, Ash and Boberg (2006b: 77). However, with regard to DRESS vowels in pre-nasal position, it is not made explicit whether he included them (cf. Subsection 4.1): In footnote nine, he states that each vowel under analysis, /i, e, æ, o, ʌ, u/, was “also represented by a token before /n/”, except FOOT (Boberg 2005: 152). Only for ash, he explicitly states that he excluded pre-nasal tokens; for K IT, DRESS, (TRAP) and LOT/THOUGHT, he maintains that their representation “was augmented by including tokens before /p/ and /k/ as well as /t/” (Boberg 2005: 153, fn. 9). Hollett’s study of word-list data from St. John’s did not exclude DRESS in pre-nasal position either (2006: 149, 2007: 29, 35). Due to the difference in terms of inclusion/exclusion of DRESS before pre-nasal environments, my analysis is neither directly comparable to the findings of Boberg (2005, 2010) nor those of Hollett (2006, 2007).

The 1046 tokens included in this analysis are first discussed in terms of the Euclidean Distances (EDs) between the speaker means of DRESS and STRUT. As shown in Subsection 5.1.1, the vowel in STRUT words is stable in my data in apparent time, so that a significant difference between the respective EDs of the three age groups provides strong evidence of a shift of DRESS in apparent time. The previously discussed shortcomings of EDs hold here as well: first and second formant values contribute equally to the ultimate distance

---

111 As mentioned in Section 4.1, his data from Montreal, Quebec, is part of a larger data set called the *Phonetics of Canadian English* and the *Phonetics of Montreal English* projects. All of his below-mentioned publications are to some extent based on these projects and thus have the same underlying methodology (cf. Boberg 2010: 199).
5.4. The Shift of Dress

Figure 5.15: Boxplots of the Euclidean Distances between DRESS and STRUT per age group (n = 34).

Figure 5.15 shows the boxplots of the ED values per age group. Most obviously, the middle-aged and young speakers have mean and median EDs above the grand mean as opposed to the old speakers. The mean values show that the difference in distance between the old speakers and the middle-aged speakers is three times as large (almost 200 Hz) as that between the middle-aged and young speakers (about 50 Hz). Hence, the significant difference does not seem to be between the middle-aged and young speakers as for TRAP, but between the old and middle-aged speakers, as their notches do not overlap. The unequal lengths of the whiskers for old and young speakers seem to indicate violation of normality. Likewise, the differences in box sizes hint at heteroscedasticity between the variances of the three groups. Since the interquartile range (box size) for the middle-aged speakers is largest, they seem to metaphorically build a bridge between the oldest and youngest speakers. Anticipating a result from Subsection 5.4.2, females are most likely responsible for the relatively large interquartile range in the data points of the middle-aged speakers.
Test & $p$-Value
\hline
Global stat & 0.82 \\
Skewness & 0.28 \\
Kurtosis & 0.94 \\
Link function & 1.00 \\
Heteroscedasticity & 0.54 \\
\hline

Table 5.11: Assumptions of an ANOVA regarding the distribution of age groups and EDs between DRESS and STRUT: results of the global test on four degrees of freedom ($n = 34$). Level of significance is set to 0.05.

Peña and Slate’s (2003, 2006) global test for linear model assumptions does not support the visual impressions of violations of the assumptions of a linear model (ANOVA). As Table 5.11 shows, all of the assumptions for an ANOVA are acceptable. The results of a monofactorial ANOVA support the impressions indicated by Figure 5.15: The difference between the age-group means of the EDs is very significant: $F_{2, 31} = 7.032$, $p = 0.003$. However, the variable age group explains only 30% of the overall variance: multiple $R^2 = 0.31$, adjusted $R^2 = 0.27$. Tukey’s HSD showed that two out of three pairwise post-hoc comparisons of the means are significantly different from each other: the mean of the old speakers differs very significantly from the mean of the middle-aged speakers ($p = 0.01$) and from that of the young speakers ($p = 0.003$). The difference in ED means between middle-aged and young speakers is insignificant ($p = 0.72$). For young and middle-aged speakers, DRESS is significantly closer to STRUT than for old speakers.

5.4.1.1 Correlation of Age and Formant Values

These results seem to be at odds with the findings for TRAP in apparent time, where the middle-aged speakers grouped more closely with the old speakers than with the young speakers. With regard to DRESS, they have moved their vowels significantly away from those of the old speakers in relation to STRUT. Since the position of the latter is behind that of DRESS in the vowel space, the difference might suggest that young and middle-aged speakers retract their DRESS vowels more than they lower them, especially since this has been reported by Boberg (2005) for Montreal. However, he did not find any movement for DRESS in first formant space in apparent time at all (ANOVA result: $F = 2.502$, $p = 0.98$; Boberg 2005: 144), which led him to conclude that DRESS is retracting rather than lowering and thus retracting parallel to TRAP rather than filling the vowel space vacated by it.

In his criticism of Clarke, Elm and Youssef’s (1995) study of the Canadian Shift, Boberg maintains that they had no suitable age groups to compare their impressionistic findings for young respondents from Ontario to, and that they did not acoustically analyze their whole data set (2005: 135). If they had done so, they could have shown, e.g.
5.4. The Shift of Dress

preliminarily via Pearson product-moment correlation coefficients of formant values and age, that there is a trajectory of the shift in apparent time (cf. Boberg 2005: 137). As I have shown previously for TRAP, Pearson’s correlation coefficient determines the strength of the correlation between two continuous variables. Boberg (2005) outlines the correlation coefficients Pearson’s $r$ for first and second formant values of all vowels of the Canadian Shift from a small sample of ten speakers from Ontario recorded for the Atlas of North American English (Labov, Ash and Boberg 2006b; cf. Boberg 2005: 137). For DRESS, the first formant values show a low correlation of $r = 0.16$ and the second formant values show an intermediate correlation of $r = 0.47$. This interpretation of the coefficients is, however, subjective, as Boberg (2005: 136) considers the low correlation of speaker means and age as non-existent.\footnote{A result of statistical significance would have shed more light on whether the correlation of 0.16 has to be considered or whether it can be disregarded.} Gries (2009: 139), by contrast, interprets a correlation of up to 0.2 as low. Sullivan and Feinn (2012: 280) interpret a correlation of $r = 0.04$ as small and up to 0.25 as medium. In case of a positive decimal number, the correlation coefficient indicates that with increasing age, the first formant value increases as well, i.e. the older speaker are, the more lowered are their DRESS vowels. This finding contradicts, of course, the Canadian Shift pattern in general, which is problematic, because it was identified first in Ontario.

The inherent problem of such subjective interpretation of a number is that each interpreter has different ideas about one value in comparison to another value. Boberg’s students Sadlier-Brown and Tamminga (2008: 9), for instance, decided in their comparison of Vancouver, British Columbia, with Halifax, Nova Scotia, in terms of the Canadian Shift on the basis of Pearson’s $r$ that DRESS was not only retracting, but also lowering. Their coefficients for both localities were admittedly very close between formants one and two (e.g. Halifax, DRESS F1: $r = 0.48$, F2: $r = -0.51$ with birth year). However, Pearson’s $r$ assumes the two continuous variables to be distributed normally. In my data, age is not distributed normally\footnote{The Shapiro-Wilk test for age was highly significant: $W = 0.857$, $p < 0.001$.}, so that I use Kendall’s $\tau$ (tau) instead: Both correlations between age and per-speaker mean formant values (cf. Boberg 2005: 146) are intermediate: F1: $\tau = -0.3$ and F2: $\tau = 0.37$\footnote{Pearson’s $r$ yielded values quite similar to the ones reported by Sadlier-Brown and Tamminga (2008: 9): F1: $r = -0.47$, F2: $r = 0.55$. The correlation coefficient for the second formant is also close to the one reported by Boberg (2005: 137) for Montreal. If the data are normally distributed, the two correlation coefficients (Pearson’s $r$ and Kendall’s $\tau$) yield very similar values for the same correlation.}, i.e. the lower the ages are, the higher is formant one (DRESS lowering), and the lower the ages are, the lower is formant two (DRESS retraction). Young speakers thus both retract and lower their pronunciations of DRESS in comparison to older speakers, as the difference in the coefficients is subjectively marginal (0.07).
5. Analysis and Discussion

5.4.1.2 Dress Retraction versus Lowering

Unlike shown for TRAP (cf. Subsection 5.3.1.3), the correlation coefficients for DRESS both yield significant results. Age and formant one correlate very negatively with each other: \( \tau = -0.3, z = -2.458, p = 0.014 \). The correlation of age and formant two is highly positive: \( \tau = 0.37, z = 3.024, p = 0.002 \). These results show that there is no justification in claiming that one of the two movements is much stronger than the other in my data. Boberg’s studies that suggest so are based on findings where there was no change at all in one of the two formants. In my data, the movement of DRESS in apparent time is thus a diagonal one (simultaneous lowering and retraction), as outlined by Labov, Ash and Boberg (2006b: 220) for Canada and Sadlier-Brown and Tamminga (2008: 10) for Halifax, Nova Scotia, which is consistent with Labov’s Principle II (non-peripheral nuclei fall; 1991: 7, 1994: 176; Labov, Ash and Boberg 2006b: 16). The mean first formant value of DRESS for my youngest speakers is 689 Hz (sd 33 Hz), which not surprisingly does not reach the mean Boberg (2010: 145) outlines for Canada (732 Hz, sd 44 Hz) “as a national average against which the systems of particular regions or individuals can be compared” (Boberg 2010: 146; also cf. Boberg 2008b: 135). As he further makes clear (2010: 147), his mean first formant value “conforms to the threshold established by Labov, Ash and Boberg (2006: 219) for the shift: F1 of /e/ [E] in dress is greater than 650 Hz [...]”.

The same is true for my youngest respondents.

Falling of non-peripheral nuclei does not necessarily mean that only a strict increase in formant one is to be expected, as the positioning of the short front vowels on the two tracks in relation to one another is much more diagonal than vertical (cf. Labov, Ash and Boberg 2006b: 17). Thus, my result contradicts that of Hollett (2006: 154, 2007: 54-55), namely that old females lower their DRESS vowels without retraction in St. John’s English. Boberg’s (2005) strong emphasis on retraction in Montreal, Quebec, cannot be corroborated for St. John’s, Newfoundland. Since DRESS lowers and retracts simultaneously, a plot of the second formant values per age of DRESS and TRAP will naturally reveal a parallel shift as suggested by Boberg (2005: 146). The fact that parallel retraction is present in my data does not indicate that the Canadian Shift is not a chain shift. However, the fact that the middle-aged speakers are more advanced in terms of DRESS lowering than they are in terms of TRAP retraction seems contradictory.

In chain shifts, one vowel vacates phonemically the space of another by maintaining adequate margins of security between neighboring phonemes (e.g. cf. Boberg 2005: 136; Labov, Ash and Boberg 2006b: 130). If TRAP has not retracted far enough, the margin of security to DRESS is not maintained, so that the latter should not be able to move in the

Note that Labov, Ash and Boberg’s (2006b) thresholds for the Canadian Shift differ throughout the atlas. On page 130, the Canadian Shift is “defined by quantitative measures. [E.g. by] the lowering of /e/ [E] in DRESS] by a mean F1 greater than 660 [Hz].” On page 151 in footnote 5, the definition for DRESS is “F2(e) < 650 Hz”, i.e. the mean second formants have to be lower than 650 Hz, which is clearly due to editing errors. On page 219, the definition for DRESS is a mean first formant greater than 650 Hz. My youngest respondents conform to both thresholds: 689 Hz > 660 Hz and 650 Hz.
direction of the former. Yet, the middle-aged speakers are insignificantly different from the young speakers in their distance of DRESS to STRUT, but significantly different from the young speakers in their distance of TRAP to STRUT. This result may be explained by four possible scenarios: First, the Canadian Shift is not a chain shift, as suggested by Boberg (2005). Second, speakers are able to retract and/or lower their vowel realizations in a subsequent rather than simultaneous manner. Third, the result is an artifact of the method. Fourth, the difference in token number between TRAP and DRESS vowels in my data skews the result.

With regard to the first explanation, Boberg (2005: 136) suggests that his findings indicate that the mental process underlying parallel retraction of DRESS and TRAP rather than lowering is “a kind of analogy that produces identical alterations in the production of phonetically similar vowels”. If these alterations are really to be identical, the movement of the vowels involved should thus be identical in both formant spaces. Unfortunately, this does not make sense for his findings, as there is no significant alteration in first formant space for DRESS, but there is one for TRAP (2005: 141). If it is possible to retract a front lax vowel without an accompanying lowering (or vice versa), it is also possible for a vowel to be first retracted and subsequently lowered (or vice versa), which leads to the second explanation. It may be the case that respondents from Montreal retract their vowels first, and maybe lower them at a later point in time. If this was true, then the chain shift is not marked by a movement of one vowel at a time in both formant spaces, but first in one formant space for all vowels and then in another formant space for all vowels. This would make sense for Boberg’s (2005) findings in Montreal, as TRAP has begun to lower, but DRESS and KIT have not, while at the same time they are all retracted. Admittedly, this explanation sounds a little far-fetched, as it questions the whole nature of established chain-shift theory, whereas Boberg’s (2005) interpretation acknowledges this theory and maintains that the Canadian Shift is simply not a chain shift. This explanation is, however, also doubtful, because it would imply that the shift is more advanced in middle-class St. John’s than it is in Montreal. It is additionally doubtful, Boberg (2010: 147) emphasizes that there are “significant effects of speaker age on both the F1 and F2 of /e/” in his “apparent-time analysis presented [in Chapter 5 for Montreal]”, “suggesting that the shift involves both lowering and retraction of that vowel”. Interestingly, he does not reiterate his original explanation for these new findings.

Note that in the Northern Cities Shift (NCS), the possible movements for DRESS are either lowering to [æ] or retracting to [ə]: While Labov, Ash and Boberg (2006b: 197) stress that the latter is the dominant movement, the discussion of individual schematics representing the advancement of the NCS reveals that DRESS is moving in both formant spaces (2006b: 198-199), i.e. DRESS is retracting and lowering simultaneously. Only KIT seems to be moving merely in one formant space in the course of the NCS: “[...] /i/ has shifted back, almost to central position” (2006b: 198), but is not further discussed.
although the movements of DRESS and TRAP are now truly “identical alterations” (Boberg 2005: 136).

I consider the third explanation as the most likely one: Euclidean Distances take formants one and two equally into account, although these have different ranges. Since TRAP only changes significantly in second formant space, the relative contribution of the first formant to the ultimate distance metric is much smaller than for DRESS (and KIT), so that the fact of a movement in both formant spaces for the latter results in smaller distances for middle-aged speakers. While both vowels, TRAP and DRESS, are retracted for middle-aged speakers, the alteration in the first formant is much smaller for TRAP, i.e. zero, than in the second formant, i.e. more than zero. The alteration in first formant space for DRESS is about the same as that in second formant space. The regression analyses below will shed more light on this. Although not unlikely, the fourth scenario does not require further elaboration – more TRAP tokens would have to be analyzed to exclude this possibility.

5.4.2 Generalized Mixed-effects Modeling of DRESS Lowering

As outlined above for St. John’s English and by Labov, Ash and Boberg (2006b: 220) for mainland Canada, DRESS shifts in first and second formant space, i.e. it shows a diagonal movement towards the position of TRAP. Labov, Ash and Boberg (2006b: 219) defined the behavior of DRESS in the Canadian Shift by its first formant value, due to its mid position on the non-peripheral track (Principle II; cf. Labov 1991: 7, 1994: 176; Labov, Ash and Boberg 2006b: 16). If epsilon values reached a threshold greater than 650 Hz (660 Hz; Labov, Ash and Boberg 2006b: 130), the vowel was determined to be lowered via participation in the shift. I take the same approach here, since both formant values show a significant change in apparent time. The dependent variable of the multiple regression analysis is thus formant one.

The regression analyses consist of a total of 1046 DRESS tokens, excluding those in pre-nasal and other environments (cf. Subsection 4.4.5). Both linear and logistic regression are conducted with the Bonferroni approximation. The estimates in linear regression in SPSS are based on the REML method (Residual Maximum Likelihood), and the significances are based on treatment contrasts.

I combined the categories following place and manner of articulation as outlined in Subsection 4.6.3.4, yielding a factor group with apical stops, apical fricatives, labials and velar stops, which I refer to as ‘PoA MoA’. The predictors taken into consideration for the linear regression via GLMM are age (continuous), sex, LItotal (continuous), style, glottal state, PoA MoA and number of following syllables. Age and LItotal are centered on their minimum value (age = 20, LItotal = 1.5). None of the interactions between the fixed effects were significant, so that I will not address them any further below. Random intercepts were given to speakers and lexical items.
As for TRAP, I conducted the logistic regression in Rbrul, based on the first formant threshold suggested by Optimal Binning (cf. Appendix E.3.2). The procedure identified two bins for formant one as the dependent variable. The first bin has an upper end point of 669.5 Hz, which in turn serves as lower end point for the second bin. I use 669 Hz as the threshold for determining lowered DRESS tokens, which is more conservative than the one(s) suggested by Labov, Ash and Boberg (2006b: 130, 219). The independent variables are the same as outlined above for the linear regression, but age is modeled in groups. The estimates of Rbrul are based on Maximum Likelihood (ML; automatically altered), instead of REML. Logistic regression analysis is based on sum contrasts.

5.4.2.1 Linear Regression

The linear model’s deviance is 11,221 (AIC 11,227; $R^2_{\text{GLMM}} = 0.147$) without interaction groups. In line with the results discussed in Subsection 5.3.1, age is the most important predictor in the model ($F_{1, 1036} = 9.789, p = 0.002$). Among the social variables, it is the only significant factor group in the model. LItotal, sex and style are not significant whatsoever ($p = 0.515$, $p = 0.62$ and $p = 0.908$, respectively). At a first glance, it seems that DRESS lowering is a very recent development in St. John’s English, which is just about to gain social relevance. This is also supported by the linguistic predictors: following voicing is the only significant factor ($F_{1, 1036} = 4.901, p = 0.027$). It is closely followed by the crossed category of place and manner of articulation (PoA MoA) in terms of the $p$-values, but the variable fails to reach significance ($F_{3, 1036} = 2.245, p = 0.081$). The same is true for number of following syllables ($p = 0.189$). In comparison to the relatively large $p$-values of the social predictors, the linguistic ones seem to be determining or partially condition the lowering of DRESS. Yet, it is age that dominates the regression model in terms of significance. The random effects are both highly significant: for speaker, Wald’s $Z = 1361.06, p = 0.001$, and for lexical item, Wald’s $Z = 1317.91, p < 0.001$ (cf. Appendix F.4).

The individual factors are shown in Table 5.12. Age negatively correlates with the first formant, i.e. the younger speakers are, the greater their first formant values (lowering). The lowering process in apparent-time is significantly influenced by voicing of the following consonant sound. Voiceless sounds are more likely than voiced ones to promote lowering of DRESS vowels.

This result is in line with that reported by Clarke, Elms and Youssef (1995: 216) for young Ontarians, for whom only the lowering of DRESS was significantly promoted by following voiceless consonants (cf. Subsection 5.6.2). Acoustic phonetic literature generally supports the notion that higher first formant values occur before voiceless consonants rather than voiced ones (cf. e.g. studies by Mermelstein 1978: 332 for perception and Summers 1987: 858, 1988: 485 for production of vowels before voiced versus voiceless consonants). In terms of age, my result corroborates those of Hoffman (2010: 131), Hoffman and Walker (2010: 56) and Roeder (2012: 484) for young speakers from Toronto,
Ontario. However, these studies did not include linguistic factors in their statistical analyses and found variable support for significant sex differences with regard to dress in the Canadian Shift. For instance, De Decker and Mackenzie (2000: 5) attest young Ontarian females a significant effect for lowering of dress in interview style; Boberg (2005: 147) does not find one for retraction of dress among Montrealers in word-list style.

Among the few studies that analyzed varieties in the Atlantic Provinces, similar results were reported in terms of age by Boberg (2010: 147) for Vancouver, British Columbia, and Halifax, Nova Scotia; Gardner and Childs (2011) for Petty Harbour, Newfoundland; and Hollett (2006: 156, 2007: 54) for St. John’s, Newfoundland. The observation that dress is not lowered in St. John’s, Newfoundland, as offered by Labov, Ash and Boberg (2006b: 221), and that the Canadian Shift is not a classic chain shift (because dress is not lowering) as offered by Boberg (2005: 136) can thus not be corroborated with my data.

The interplay of the predictors age and glottal state suggests that following voiceless environments favor lowering per se, but for the youngest speakers the lowering effect for

Table 5.12: REML-based results of linear regression analysis of dress in SPSS via GLMM (treatment contrasts), including speaker and word as random effects (n = 1046, deviance = 11,221). The p-values of the insignificant results are provided in square brackets.
DRESS is significantly higher in this environment than for speakers of older ages (cf. Figure 5.16). In fact, young speakers' first formant means of DRESS before voiceless consonants is 697 Hz, compared to 654 Hz before voiced consonants. In contrast, middle-aged speakers have a mean of 666 Hz before voiceless and of 620 Hz before voiced consonants. The difference in means between the two phonological environments is smallest for the old speakers: 631 Hz before voiceless and 615 Hz before voiced consonants.

I understand this combination of factors to show that lowering of DRESS has only recently started in the speech community of St. John’s. It may not have progressed far enough yet to display the usual social picture of women (significantly) leading the change in progress. In other words, both sexes have more or less simultaneously begun to lower their ejectives in favorable linguistic environments. This becomes particularly obvious when the first formant means of middle-aged and old speakers between the two glottal states are compared. The difference of 4 Hz is minimal, but the difference of 35 Hz before voiceless consonants in contrast to that supports the idea that lowering started within the

118 Note that the Hertz value of 697 is closer to Boberg’s (2010: 145) Canadian mainland mean of 732 Hz than to Labov, Ash and Boberg’s (2006b: 219) threshold of > 650 Hz (also cf. Boberg 2010: 147).
middle-aged group before voiceless consonants. Only the youngest speakers have already begun to lower DRESS in voiced environments as well.

Although PoA MoA is not significant, velar stops additionally favor higher first formant values, which supports this interpretation. Within the factor group, labials and apical stops have significantly smaller first formant values, i.e. the DRESS vowels before these phonological contexts are at a higher position in the vowel space. Apical fricatives show a similar effect on preceding DRESS, but the difference to velar stops is not significant. Since the difference in first formant mean values of labial stops and labiodental fricatives was insignificant (655 Hz and 664 Hz, respectively), manner of articulation seems to play a minor role with regard to lowering of DRESS (i.e. higher F1 frequencies). Likewise, labials and apical stops cause the first formant frequency of DRESS to be significantly lower (i.e. a non-lowered DRESS vowel) than velar stops do. This further supports the assumption that place seems to play a more important role than manner of articulation, unlike suggested by Clarke, Elms and Youssef (1995: 216) for DRESS in Ontario. However, such a differentiation between the two linguistic variables is not possible with my data.

The finding from linear regression analysis contradicts the result discussed in Subsection 5.4.1: It is not the case that middle-aged and young speakers are more advanced in lowering DRESS than in retracting TRAP. Even under most favorable conditions, the difference in means is evenly distributed among the three age groups with roughly 30 Hz more lowering as the ages per group decrease. Middle-aged speakers in fact pattern with old speakers before voiced environments in terms of their mean F1 values; it is only the youngest age group that shows more lowering before voiced consonants than the other two groups (by 34 Hz). Unlike in the analysis of the Euclidean Distances (EDs) above, only the first formant is taken into consideration in this analysis. It may have been the combined favorable linguistic environments that indicated a significant change for young and middle-aged speakers together in apparent time. I rather see confirmation in the interpretation offered above: it is most likely that, due to their different ranges, the combined first and second formant values led to those results: For DRESS, lowering and retraction are significant in apparent time, and for TRAP, only retraction is significant in apparent time. Since the ED metric disregards the different ranges between F1 and F2, the lack of a significant change in F1 for TRAP may have caused the pattern of old and middle-aged versus young speakers for this vowel, while the additional significant change in F1 for DRESS may have caused the pattern of old versus middle-aged and young speakers for that vowel. If we looked at this ED result in isolation, it would suggest that middle-aged speakers are more advanced in terms of DRESS lowering than in terms of TRAP retraction. If we assumed a temporal connection between the two vowels, i.e. TRAP retracts first,

---

119 Due to the insignificant difference in first formant mean values between labial stops and labiodental fricatives, I coded them together as labials (e.g. cf. Crawley 2007: 374-377). The multiple statistical testing of formant means was conducted in SPSS with the Bonferroni approximation (cf. Abdi 2007: 106).
then young speakers lead TRAP retraction and middle-aged speakers are possibly already quite advanced, while only young speakers may show DRESS lowering. The interpretation of the current regression analysis suggests exactly that: DRESS lowering is not as advanced in St. John’s, Newfoundland, as TRAP retraction, since, for instance, the age coefficient of the former (−1.498) is much smaller than that of the latter (5.484), i.e. younger speakers lead much less in the lowering of DRESS than in the retraction of TRAP (cf. Labov, Ash and Boberg 2006b: 197). The minus-sign only indicates the direction of the correlation between each formant value and age; it is not indicative of the strength of the correlation. In summary, the fine-grained results of the linear regression analysis clearly support a chain shift notion for the two vowels, DRESS and TRAP, whereas the ED metric is not able to yield such a fine-grained result.

The other factors do not play a role in the shift of DRESS (cf. Table 5.12). What appears to be noteworthy is that within number of following syllables, the vowels are more lowered when they are followed by another syllable, i.e. when they are shorter, but this effect is not significant. The significance of sex shown by Clarke, Elms and Youssef (1995: 217) cannot by replicated with my data from St. John’s, Newfoundland, but the tendency for females rather than males to lower DRESS is the same. Similar to the effect of style on TRAP, style’s effect on DRESS seems to lead to a behavior of the vowel opposite to what is expected in St. John’s, Newfoundland (cf. Clarke 1991: 116 for TRAP): Tokens uttered in spontaneous speech seem to have higher first formant values (more lowering) than those uttered in the formal styles. However, this factor group is the least significant one with the greatest p-value.

5.4.2.2 Logistic Regression

The model’s deviance is 990 (AIC 1020; $R^2_{\text{GLMM}(m)} = 0.233$ and $R^2_{\text{GLMM}(c)} = 0.566$) with 15 degrees of freedom and a grand mean of 0.393, i.e. 39.3% of all tokens in the model possess a first formant value beyond the threshold of 669 Hz. The application value is established via Optimal Binning (cf. Appendix E.3.2). The grand mean for the predictions of all cells (the intercept) is 0.459. The Rbrul model is based on sum contrasts and its estimation method is Maximum Likelihood (automatically altered by Rbrul).

The overall results are quite similar to those of the linear regression detailed above, but the order of some of the predictors and their significances differ quite substantially. Age group and the glottal state of the following consonant are most significant ($p = 0.014$ and $p = 0.016$, respectively). Unlike the results of linear regression, style is significant as well ($p = 0.019$), and sex is only marginally insignificant ($p = 0.083$). Local-ness index, number of following syllables and place/manner of articulation are all clearly insignificant ($p = 0.256$, $p = 0.269$ and $p = 0.332$, respectively).

The reasons for this difference may be manifold: It could be due to the sum contrasts, the estimation method, the categorization of the continuous independent variable age, an interaction that I could not find via cross-tabulation, different coding of the variable
Table 5.13: ML-based results of logistic regression analysis of DRESS in Rbrul via GLMM (sum contrasts), including speaker and word as random effects (n = 1046, deviance = 990). The factor weights within the insignificant factor groups are provided in square brackets (n.s. = not significant).

The results of the individual factors are shown in Table 5.13. In this model, middle-aged and old speakers disfavor the application value of a first formant greater than 669 Hz. The young speakers reach the application value in almost 60% of their uttered tokens. This is about 20% less than the retraction of TRAP for young females. Since females favor lowering of DRESS (though insignificantly), it is likely that the females are primarily responsible for the application rate of 60% within the youngest age group.

With regard to the middle-aged speakers, categorizing age into three groups and categorizing the dependent variable seems to have the middle-aged speakers pattern with the
old speakers in terms of DRESS lowering. This supports the interpretation of the linear regression result above insofar as it clearly shows that the ED metric discussed in Subsection 5.4.1 is inaccurate in patterning middle-aged and young speakers together. However, it is accurate in its major function, namely determining a shift of the vowel in apparent time. In other words, EDs are not intended for a socially meaningful interpretation in terms of detailed differences between age groups to be based on them. As mentioned previously, the inaccuracy of the ED result is most likely caused by taking formants one and two equally into account although they have different ranges. Yet, one caveat remains: all the findings depend on the accuracy of the application value in this analysis.

A comparison of application rates per age group and glottal state is not as revealing as a comparison of the first formant means as shown above. Roughly 50% of the epsilon tokens before voiceless consonants are lowered. With every second token before voiceless consonants lowered, the application rates of the middle-aged and old speakers could be interpreted in that every second token was before a voiceless consonant. This is, however, not possible, since the ratio per voiceless token within the age groups is not known from the analysis outlined in Table 5.13. Regardless of factor groups’ significance, when further investigating those factors that disfavor the application value, it becomes apparent that even in these cases, at least 25% of the DRESS tokens have the application value of $> 669$ Hz. This may suggest that the threshold might be too low, although it is already quite a large value in absolute terms.

Despite the high application rates in non-favoring environments, the general picture painted by the linear regression does not change substantially. Instead, the additionally significant social factors support the interpretation of the linear regression outcome: Sex is not significant, but following voicing is and age groups even more so. This is the main reason for the interpretation offered in the subsection above. In addition, style becomes significant: reading passage and word list favor the lowered DRESS tokens, whereas interview style (spontaneous speech) does not. At first glance, the hierarchical patterning of reading passage and word list seems wrong. Literature suggests the reverse pattern, since word-list style is believed to be the most formal of the three, eliciting the least natural speech data (cf. Labov 1972b: 84).

Apart from the criticism of Labov’s definition of style outlined in Subsection 0.3.4, I will consider another reason why this may not necessarily be so: Reading a word list, regardless of its type such as elicited days of the week, simple words in isolation or minimal pairs, is the most unnatural and thus the rarest speech behavior. Most, if not all, of my respondents – the old ones remembered quite well – told me rather repetitively how they were drilled in their early years of school not to speak in a strong form of the local variety, but a more standard variety of English. Drilling is not equally successful for each participant. Usually, children from all social classes are required to go to school. The early years of school crucially consist of learning how to write, speak and most importantly read. Reading tasks in school do,
however, not require the students to read word lists, but texts. Reading a text is thus closely linked to speaking a standard version of English in the minds of respondents, which was more or less simultaneously learned in the early years of school.

It is natural that they will speak most formally in their reading passages, since the link between reading a text and formal speech has become an automatism – at least when compared to reading a word list. Such reading requires more concentration on the part of the reader, because they cannot anticipate the next word on the basis of context. In addition, nouns and verbs are usually collocating in natural texts, whereas in word lists, the grammatical status of a lexical item may be ambiguous. In English, most pronunciations are based on the grammatical function of the lexical item, not the orthographic realization (e.g. *lead*). In essence, reading a text formally is much easier for most speakers of a language than reading a word list formally, as the former is usually internalized in the early years of school and the latter is never practiced. Even the days of the week may be crystal-clearly stored in each speakers’ brain, but uttering them consecutively (on demand) is not.

However, the difference in hierarchy between reading-passage and word-list style may also be due to the token ratio of three to one. In other words, this may not have any social meaning. The logistic regression outcome for style, which replicates the findings of the VARBRUL programs, is in line with Clarke’s statement that innovative features enter the speech community of Newfoundland and thus St. John’s via the formal styles (e.g. Clarke 2012: 514 for *trap*). Reading-passage style favors lowering of *dress* almost as much as interview style disfavors the application value. If this indeed has social meaning and is thus more than an artifact of the logistic regression analysis, it supports the interpretation of the results of the linear regression outlined above: *dress* lowering has only very recently begun for the young speakers and is hence more than twice as often applied in the formal styles than in spontaneous speech (application rates of 61.4% versus 27.3%). This would be in line with Clarke’s findings for *trap* in the 1980s (1991: 116).

The local-ness index does not play a role with regard to lowering of *dress*, but the tendency points in the expected direction: the lower the local-ness score, the higher are the first formant values, i.e. *dress* is lowered. The same is true for number of following syllables. *dress* in environments with syllables following the stressed vowel’s syllable favor the application value. The tendency is, however, due to chance. Likewise, the influence of following phonological context is due to chance. Tendentiously, velar (and labial) stops favor the application value, which is in line with the result of the linear regression; labiodental fricatives minimally disfavor a *dress* vowel lower than (i.e. a first formant value higher than) 669 Hz, and the last two factors, apical stops and apical fricatives, disfavor the application value even more than that.
5.5 The Shift of Kit

As mentioned for the literature on dress in St. John’s, Newfoundland, the literature on kit is sparse. With regard to the Canadian Shift in other regions of Canada, the participation of kit in the shift is contested. Whereas most studies find a shift of kit in apparent time,\textsuperscript{121} Labov, Ash and Boberg (2006b: 220) maintain that there is “no shift of /i/ [...] indicated in the [Atlas of North American English] data for Canada” as a whole, including St. John’s, Newfoundland. Having stated that, Labov, Ash and Boberg (2006b: 221) show a table with a significant positive age coefficient for formant one and thus in a direction opposite to what is expected for kit in the Canadian Shift: The older speakers are, the higher are their first formant values (lowering). Unfortunately, the authors do not elaborate on this apparent contradiction, most likely due to the fact that the atlas is not solely concerned with Canada. In light of Labov, Ash and Boberg’s (2006b) internal inconsistency and the results of most other studies, I do not see any a priori doubt that kit is participating in the shift (cf. Hypothesis 6 in Section 3.4).

As for dress, Lawrance (2002) and Boberg (2005) find retraction to be the significant movement of kit, so that the Lawrance suggests parallel retraction with dress and Boberg suggests parallel retraction with dress and trap, instead of a classic chain shift of the three lax vowels. In St. John’s, Newfoundland, Hollett (2006: 156, 2007: 58) finds significant lowering of kit for her young speakers in real time and retraction of kit for her older speakers in apparent time. Again, the results outlined below will be discussed in light of those studies that used a methodology similar to the one I employed in this study.

Following the sequence in the previous sections, the first subsection outlines the analysis of the movement of kit in apparent time in my data via an ANOVA of age group and the Euclidean Distances (EDs) between kit and strut. Data exploration has shown that age is the most important predictor for the change of formant one (cf. Appendix E.4.1). As outlined in Subsection 5.1.1, strut is stable in apparent time, so that the difference in EDs between kit and strut per age group reflects the shift of kit. Correlation coefficients will indicate which formant correlates more strongly with age in order to establish a trajectory of the change (cf. Boberg 2005: 135) and thus to assess the relative contribution of first and second formants. In order to statistically evaluate the correlation between the two formants and age, I test it for significance. The second subsection provides the results of multiple regression analyses. The dependent variable for these analyses is chosen on the basis of the correlation’s significance tests and the literature. The logistic regression is based on the categorization of the dependent variable

5. Analysis and Discussion

via Optimal Binning and Decision Trees (cf. Appendix E.4). The analysis triangulates the results of the shift of KIT in apparent time and extends the findings in terms of the relative contributions of the other (social) predictors to the shift of KIT.

5.5.1 KIT Movement across Age Groups

Parallel to the analysis of DRESS, KIT tokens before nasals are excluded, in addition to the other categorical exclusions, yielding a total number of 1369 observations. This number of KIT tokens is more than twice as large as that analyzed by Labov, Ash and Boberg (2006b: 221) acoustically (n = 605) in order to draw a valid picture for the speech community of St. John’s. In the data for the atlas, Labov, Ash and Boberg (2006b: 220) find that three respondents from the Maritimes perceive KIT and DRESS before nasals to sound very similar. Although there is no general attestation of the presence of the pin-pen merger as far north as St. John’s, Newfoundland, I exclude nasals from the analysis of KIT vowels, primarily because they should have the strongest lowering and retraction effect on the vowel (and due to effects of coarticulation; cf. Labov 1994: 197 and Thomas 2001: 52). A result concerning the movement of KIT in apparent time could then be

Figure 5.17: Boxplots of the Euclidean Distances between KIT and STRUT per age group (n = 34).
5.5. The Shift of *Kit*

<table>
<thead>
<tr>
<th>Test</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global stat</td>
<td>0.57</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.08</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.95</td>
</tr>
<tr>
<td>Link function</td>
<td>1.00</td>
</tr>
<tr>
<td>Heteroscedasticity</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 5.14: Assumptions of an ANOVA regarding the distribution of age groups and EDs between *kit* and *strut*: results of the global test on four degrees of freedom ($n = 34$). Level of significance is set to 0.05.

biased by the conditioned merger. Recall that Boberg (2005) and Hollett (2006, 2007) included nasals in their studies, so that comparability to my study is very limited.

Figure 5.17 shows the boxplots of the Euclidean Distances (EDs) between *kit* and *strut* per age group. The youngest and the middle-aged speakers both show one conservative outlier with a larger distance than the speakers within 1.5 times the interquartile range. It is also apparent that the oldest speakers have the largest mean distance (963 Hz) between the two vowels. The mean distances for the middle-aged speakers (792 Hz) and the youngest speakers (707 Hz) is quite similar, but the notches between all three age groups do not overlap. The size of the box (interquartile range) for the oldest speakers is notably larger than those of the other two age groups, suggesting that their variance may not be homogenous in comparison to that of each of the other two age groups. The size of the whiskers for the old speakers is unequal, indicating that some of them are very conservative, but the majority tends to cluster around an ED measure of 900 Hz. The space between the mean (+) and the median (horizontal bold line) in the boxplot of the old speakers suggests additionally that the their EDs may not be normally distributed.

Peña and Slate’s (2003, 2006) global test for linear model assumptions on four degrees of freedom does not support the subjective interpretation of the figure. The results are shown in Table 5.14 and suggest that the assumptions of ANOVAs are not violated. Although outliers are present for middle-aged and young speakers, the skew of their data sets is not significant. Yet, at 0.08, it is close to the threshold of significance ($\alpha = 0.05$), most likely due to the youngest speakers’ outlier, as their boxplot indicates only a few data points to be smaller in distance than those within the interquartile range.

The result of a monofactorial ANOVA supports the interpretation of the boxplots in terms of significant differences: $F_{2, 31} = 12.46, p < 0.001$. The variable age group explains more than 40% of the overall variance: multiple $R^2 = 0.45$, adjusted $R^2 = 0.41$. Pairwise *post-hoc* comparisons of the mean EDs (Tukey’s HSD) show that two of the means differ significantly: The young ($p < 0.001$) and middle ($p = 0.005$) age groups differ from the oldest speakers, but the middle-aged do not differ from the young speakers at a statistically significant level ($p = 0.18$). Unlike outlined by Hollett (2006: 156, 2007: 57-
In order to assess the relative contributions of the two formants, Kendall’s correlation coefficient will serve as a preliminary diagnostic. For per-speaker F1 means and age, $\tau = -0.31$, and for F2 means and age, $\tau = 0.5$. Subjectively, the correlation of age and formant two seems much stronger; however, both correlations are intermediate (cf. Gries 2009: 139). In addition, Boberg’s (2005) and Lawrance’ (2002) interpretation of KIT (and DRESS) retracting, but not lowering, was based on the fact that they did not find any significant lowering of KIT (and DRESS) at all. As for DRESS, I test the significance of the correlation between formant and age. The intermediate negative correlation of formant one and age is statistically very significant: $\tau = -0.31$, $z = -2.547$; $p = 0.011$. Age and formant two correlate highly statistically: $\tau = 0.5$, $z = 4.095$, $p < 0.001$.

These results suggest that retraction and lowering of KIT are simultaneous movements. As Sadlier-Brown and Tamminga (2008: 10) find for Halifax, Nova Scotia, KIT is moving diagonally towards the position of DRESS in St. John’s English in apparent time. The results of Boberg (2005), Labov, Ash and Boberg (2006b) and Lawrance (2002) can thus not be corroborated for urban Newfoundland English. The mean first formant value Boberg (2010: 145) outlines as a national average for Canada (563 Hz, sd 41 Hz) is not approximated by my youngest speakers. They have a mean first formant value of 527 Hz (sd 17 Hz), which is yet substantially lower than that for old speakers (497 Hz, sd 28 Hz). The movement of KIT corresponds to Principle II (cf. Labov 1991: 7, 1994: 176; Labov, Ash and Boberg 2006b: 16), due to the position of KIT on the non-peripheral track. As for DRESS, modeling the formant two correlations with age of KIT, DRESS and TRAP simultaneously will show parallel retraction, as suggested by Boberg (2005: 146). This does, however, not contradict a chain-shift like behavior of the three lax vowels.

### 5.5.2 Generalized Mixed-effects Modeling of Kit Lowering

Due to the diagonal movement of KIT, as outlined in the previous subsection, formant one will serve as the dependent variable in the linear and logistic regression. I treat the vowel in the same way I treated DRESS with regard to participation in the Canadian Shift (cf. Labov, Ash and Boberg 2006b: 130 for DRESS). The analysis consists of 1369 tokens of KIT, excluding pre-nasal environments, partially due to possible lowering effects that may be linguistically conditioned as in, for instance, the pin-pen merger. The amount of tokens is relatively high, which I deemed necessary in light of the dissent in the Atlas of North American English (Labov, Ash and Boberg 2006b: 220-221). Both mixed-effects models are run with the Bonferroni approximation for multiple testing of factors. Linear regression estimates in SPSS are based on the REML method, logistic regression estimates in Rbrul on ML. The former is further based on treatment contrasts and the latter on sum contrasts.
As outlined in Subsection 4.6.3.4, place and manner of articulation of following sounds are combined to one interaction group, which I refer to as ‘PoA MoA’. The factor levels in this group are apical stops, apical fricatives, labials and velar stops. The independent variables taken into consideration in both regression analyses are age, sex, LItotal (continuous), style, glottal state, PoA MoA and number of following syllables. In the linear regression, age is modeled continuously, after it has been centered on its minimum value; the same holds for LItotal (age = 20, LItotal = 1.5). The interactions between the fixed effects are not significant, so that I disregard them in the discussion below. Speaker and lexical item are entered into the model as random effects.

The logistic regression in Rbrul is based on the first formant threshold identified by Optimal Binning (cf. Appendix E.4.2). It suggested three bins for the dependent variable F1: The first bin has an upper end point of 486 Hz and includes the majority of the tokens of the old speakers (approximately 60%). The upper end point of the second bin and thus the lower end point of the third bin is 523 Hz. Although the latter does not include the majority of the young speakers’ tokens, almost none of the old speakers’ tokens (approximately 9%) lie within the third bin. This seems to suggest that 523 Hz is a rather innovative threshold for the realization of kit in my data from St. John’s, Newfoundland. This is the value I use as the threshold for lowering of kit in the logistic regression. The predictors are the same as outlined above for the linear regression. In contrast to the linear regression model, age is modeled in groups, rather than continuously.

### 5.5.2.1 Linear Regression

Without interactions, the model’s deviance is 14,021 (AIC 14,027; $R^2_{\text{GLMM}(m)} = 0.15$). As the linear regression model for dress also suggested, age is the most important predictor ($F_{1,1359} = 16.956, p < 0.001$). In this model, the social predictor age is also instantly followed by a linguistic variable in terms of a hierarchical decrease in p-values, but instead of glottal state, it is number of following syllables ($F_{1,1359} = 13.785, p < 0.001$). The next predictor that reaches significance is style ($F_{1,1359} = 7.273, p = 0.007$). The crossed category of place and manner of articulation has a p-value at the threshold to significance ($F_{3,1359} = 2.431, p = 0.064$), followed by the glottal state of the consonant following kit ($F_{1,1359} = 2.773, p = 0.096$). LItotal and sex are not significant ($F_{1,1359} = 1.428, p = 0.232$ and $F_{1,1359} = 0.29, p = 0.866$, respectively). The random effects are both significant: for speaker, Wald’s Z = 3.100, $p = 0.002$, and for lexical item, Wald’s Z = 4.364, $p < 0.001$ (cf. Appendix F.5.1). This pattern of the predictors suggests that lowering of kit might be even more recent than lowering of dress. This interpretation becomes more accessible when the individual factors per group are inspected, shown in Table 5.15.

As expected, age has a negative coefficient, which shows that the younger speakers are, the higher are their first formant values (i.e. lowering of kit). The age coefficient of dress (−1.498) is greater than that of kit (−0.959), suggesting that kit is lagging behind
### Table 5.15: REML-based results of linear regression analysis of kit in SPSS via GLMM (treatment contrasts), including speaker and word as random effects (n = 1369, deviance = 14,021). The p-values of the insignificant results are provided in square brackets.

<table>
<thead>
<tr>
<th>Model Term</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic t</th>
<th>p-Value</th>
<th>95% Confidence Interval Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>539.943</td>
<td>11.334</td>
<td>47.640</td>
<td>0.000</td>
<td>517.709</td>
<td>562.176</td>
</tr>
<tr>
<td>Age</td>
<td>-0.959</td>
<td>0.233</td>
<td>-4.118</td>
<td>0.000</td>
<td>-1.415</td>
<td>-0.502</td>
</tr>
<tr>
<td>Number of Foll. Syllables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1+</td>
<td>-17.227</td>
<td>4.640</td>
<td>-3.713</td>
<td>0.000</td>
<td>-26.329</td>
<td>-8.125</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>-9.072</td>
<td>3.364</td>
<td>-2.697</td>
<td>0.007</td>
<td>-15.672</td>
<td>-2.473</td>
</tr>
<tr>
<td>PoA MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apical Stops</td>
<td>2.094</td>
<td>6.607</td>
<td>0.317</td>
<td>0.751</td>
<td>-10.867</td>
<td>15.054</td>
</tr>
<tr>
<td>Velar Stops</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apical Fricatives</td>
<td>-10.733</td>
<td>5.726</td>
<td>-1.874</td>
<td>0.061</td>
<td>-21.965</td>
<td>0.499</td>
</tr>
<tr>
<td>Labials</td>
<td>-12.453</td>
<td>6.889</td>
<td>-1.808</td>
<td>0.071</td>
<td>-25.967</td>
<td>1.061</td>
</tr>
<tr>
<td>Glottal State</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td>-6.447</td>
<td>3.871</td>
<td>-1.665</td>
<td>0.096</td>
<td>-14.042</td>
<td>1.147</td>
</tr>
<tr>
<td>LiTotal</td>
<td>-3.141</td>
<td>2.628</td>
<td>-1.195</td>
<td>0.232</td>
<td>-8.297</td>
<td>2.016</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-1.181</td>
<td>6.988</td>
<td>-0.169</td>
<td>0.866</td>
<td>-14.889</td>
<td>12.527</td>
</tr>
</tbody>
</table>

DRESS. The F and t statistics for number of following syllables are smaller than those for age, which suggests that the latter is more significant than the former. The influence of the factors in number of following syllables is opposite to the tendencies reported for the other lax vowels discussed above. Kit vowels in monosyllabic (e.g. stick) and multisyllabic environments with the stressed vowel in the last syllable (e.g. exist)\footnote{Note that out of 645 kit tokens with no following syllable, only six lexical items have a syllable preceding the stressed vowel. That is, more than 99% of my data in this category is monosyllabic. I thus consider the factor ‘0 following syllables’ to consist of monosyllabic lexical items only.} have significantly higher first formant values (i.e. lowering) than those in multisyllabic environments with the stressed vowel in the first syllable (e.g. different). Lowering is thus clearly not a function of vowel length, since the stressed vowels in lexical items with following syllables are usually shorter and thus more centralized than those in monosyllabic lexical items.

In my data, for instance, kit in monosyllabic words is on average 30% longer than in multisyllabic words: In word-list style, kit in monosyllabic words is on average 120 milliseconds long, in words with one following syllable, it is on average 80 ms long and in words with two or more following syllables, it is on average 70 ms long. In interview...
style, KIT in monosyllabic words is on average 100 ms long, in words with one following syllable, it is 75 ms and in words with two or more syllables 70 ms long. With all styles combined, KIT is 100 ms in monosyllabic words and 70 ms in multisyllabic words, i.e. a difference of 30%.

Lowering of KIT in monosyllabic words among young respondents from St. John’s, Newfoundland, is in line with, for instance, Labov’s (1994: 506-507) findings for tensing of TRAP in Philadelphia: He found gradual (“regular”; 1994: 502) sound change and lexical diffusion change to take place at the same time with regard to /æh/ (tensed and ingliding TRAP, which is the innovative variant). While he maintains on one page that he does “not propose to resolve the original confrontation [of regular sound change versus lexical diffusion] into a simple dichotomy – that here words change and there sounds change” (Labov 1994: 542), he maintains on another that “the realms of regular sound change and lexical diffusion would display complementary distribution [...]” (1994: 543). This notion goes back to the postulations of the neogrammarians, as Schuchardt (1972 [1885]: 57) emphasizes: “While the neogrammarians make the unexceptionability of sound laws dependent upon equality of phonetic conditions, which in my opinion does not exist at all, at the same time they treat with indifference the immediately obvious difference between words”.

Furthermore, Labov conceptualizes regular neogrammian sound change as “a gradual transformation of a single phonetic feature of a phoneme” and as “characteristic of the initial stages of a change [...] without lexical or grammatical conditioning or any degree of social awareness (“change from below”)” (1994: 542). For lexical diffusion change, he states that it is “the result of abrupt substitution of one phoneme for another”, and that it is “characteristic of the late stages of an internal change that has been differentiated by lexical and grammatical conditioning, or has developed a high degree of social awareness or of borrowings from other systems (“change from above”)” (Labov 1994: 542). These ‘conceptions’ are supported by two examples: deletion of glides and schwa would be regular sound changes, and deletion of obstruents would show lexical diffusion (1994: 543).

On page 506, Labov outlines that innovative TRAP is phonetically conditioned in Philadelphia, i.e. only those vowels before nasals are tensed (internal change). TRAP before voiceless fricatives is only tensed in monosyllabic words (e.g. ask), i.e. monosyllabic lexical items are the first to exhibit the changed variant (tensed ash). He further states that the effect of tensing of ash in monosyllabic words is confounded with lexical frequency in that high-frequency items are usually the first to be affected by the change (tensing; Labov 1994: 507). The role of frequency in the vowel change in St. John’s, Newfoundland, will be discussed in detail in Hofmann and Wagner (in prep.).

Although phonological environment usually plays a role in changes from below (regular sound change, e.g. cf. Clarke, Elms and Youssef 1995 for the Canadian Shift), it clearly also affects tensing of TRAP in Labov’s (1994: 506) data, which he describes to be part of
5. Analysis and Discussion

lexical diffusion: In earlier chapters, he maintains that “lexical diffusion had occurred in the past history of this split” of TRAP into tense /æh/ and lax /æ/, and that migration of new words into the tensed TRAP category is favored by words containing ash before nasals and /l/ (cf. Labov 1994: chp. 15 and p. 503). In addition, he emphasizes that more recent raising of historically tensed /æh/ was phonetically conditioned without lexical diffusion (cf. Labov 1994: chp. 16 and p. 503) and that “[...] it has been shown that one of these closely associated processes is a prototypical case of lexical diffusion, and the other is an instance of regular sound change [...]” (Labov 1994: 503). Lexical diffusion change consequently preceded regular (neogrammarian) sound change.

Notions of a strict dichotomy between lexical diffusion change and regular neogrammarian sound change such as Labov’s complementary distribution have always been contested by a number of researchers: For instance, Schuchardt points out that he considers “isolating the consideration of the individual sound from that of the word in which it occurs” as “wrong, at least in the absolute form in which it is asserted” (1972 [1885]: 57). Oliveira (1991) has shown that gradual (not abrupt) lexical diffusion is likely to occur in changes that are regular. Likewise, Bybee (2002) presents evidence that regular sound change shows gradual lexical diffusion, which is highly conditioned by lexical frequency. Labov himself suggests both complementary distribution of regular sound change and lexical diffusion and two polar types with many intermediate combinations of the two types of change (1994: 542). From these descriptions (including Labov’s 1994: 503, 542), it seems to be suggested that 1) regular sound change and lexical diffusion are not in complementary distribution, and 2) that the latter can precede the former, i.e. a sound change can begin as lexical diffusion and later develop into a regular sound change. I understand that this is what my data suggest with regard to KIT lowering in St. John’s, Newfoundland. The overall picture of all the Canadian Shift vowels will be detailed in Subsection 5.6.

In terms of KIT, it seems at a first glance that the young speaker’s innovative lowered variant of the vowel is 1) at a late stage of a change that has been differentiated by lexical conditioning, 2) at a high level of social awareness, and 3) borrowed from mainland Canada. With regard to the first point, only the monosyllabic words display significant lowering of KIT, which is typically the result of lexical diffusion. It is at a late stage because it is not only individual common (highly frequent) monosyllabic words anymore in which the innovative variant is produced, but all monosyllabic words – or at least virtually all. However, the innovative lowered KIT variant [i] is “a single phonetic feature of a phoneme” (Labov 1994: 542) that gradually transforms as a function of age (see below). Second, the facts that style is significant and that the formal styles have the significantly higher first formant values (i.e. lowering of KIT) might be taken as an indicator that young speakers are aware of the innovative variant (Clarke 2012: 514 for TRAP). The speakers seem to intentionally use it more often in formal styles, which suggests that the innovative variant has not yet been integrated into their vernaculars
or “do[es] not immediately affect the vernacular patterns” (Labov 1994: 78). However, a Canadian-Shift-like vowel realization significantly favored in (formal) word-list style was also attested for the Canadian mainland by Hall (fc.). Unlike Clarke, she does not attribute this finding to a change from above, but to the underlying phonetics or clarity in articulation (cf. Subsection 2.3.5). Third, the very recent innovative variant is borrowed because it came from the Canadian mainland. However, such borrowings are usually introduced by “the dominant social class” who considers “other speech communities” to “have higher prestige” (Labov 1994: 78), but my respondents are from the interior of the social hierarchy and thus unlikely to have introduced the feature. Furthermore, such borrowings “appear primarily in careful speech, reflecting a superposed dialect learned after the vernacular is acquired” (Labov 1994: 78). Consequently, if KIT lowering was borrowed, it does not have to be primarily a function of age, as anyone who desires mainland-like KIT realizations can simply learn them (abrupt change). If we add all of these indicators together, KIT lowering seems to show many intermediate combinations of lexical diffusion change (from above) and regular sound change (from below).

Assuming that KIT lowering is a lexical diffusion change, the $p$-values close to significance of the linguistic variables glottal state and PoA MoA seem to suggest that it is on the verge of leaving the lexical diffusion phase and of entering the subsequent one in which the three linguistic variables (voice, place and manner) condition the shift of the vowel more significantly than number of following syllables (cf. Labov 1994: 503). The interpretation that KIT has just begun to shift as a change from above and that it will eventually become a change from below is supported by two facts:

1) The lax vowels in the Canadian Shift are defined to move in a subsequent manner, so that TRAP and DRESS should be more advanced than KIT. This is particularly evident with regard to TRAP. As I outlined in Section 5.3.2.1, young females lead retraction of TRAP with no significant effect of style, i.e. the movement of this vowel is clearly a change from below – at least at the present stage. Clarke (2012: 514) suggests that this change of TRAP was originally brought to the linguistic system of respondents from St. John’s via the formal styles. The movement of TRAP in St. John’s English was first identified as early as in the 1980s (Clarke 1991), which had been almost exactly one generation ago (25 years) by the time I interviewed my respondents (2011). With regard to DRESS, number of following syllables does not play a role in its movement, but age and glottal state do. Depending on the fashion of statistical assessment, style is variably significant, which may either be understood as a remnant of the original manner in which the shift entered the linguistic system of respondents from St. John’s (lexical diffusion/change from above), or as an artifact of the method (e.g. categorization of the two continuous variables age and F1). If it was the latter, DRESS lowering would also be affecting the vernacular of the speakers. This is exactly what the linear regression of DRESS suggests, since style is not significant and DRESS tends to be more lowered in spontaneous speech
– characteristic of a regular sound change. If KIT is following the trajectory of DRESS, it will enter the vernacular of speakers from St. John’s eventually.

2) Within the group of monosyllabic words, we can already find a regular sound change pattern that is almost identical to that of DRESS (cf. Subsection 5.4.2.1): The significances of all of the predictors change their value beyond the alpha threshold of 0.05, except for age, which remains the most significant factor ($F_{1, 627} = 15.687$, $p < 0.001$), and glottal state ($F_{1, 627} = 3.963$, $p = 0.047$), which is close to the threshold of insignificance. The pattern is remarkably similar to that of DRESS, because age is directly followed by glottal state. The fact that the $p$-value of the latter is rather high may indicate that it is about to become more influential – as influential as it already is for DRESS. The crossed category of place and manner of articulation is the next predictor in decreasing $p$-value hierarchy ($F_{3, 627} = 2.308$, $p = 0.075$), but it is insignificant. In fact, its $p$-value is even greater than in the full model shown above (0.064). Style has no significant effect on KIT lowering in monosyllabic words ($F_{2, 627} = 2.237$, $p = 0.108$) at all, which supports the interpretation that KIT lowering is a change from below and that it is following the trajectory of DRESS. LI, total and sex are both insignificant as well ($F_{1, 627} = 1.589$, $p = 0.208$ and $F_{1, 627} = 1.019$, $p = 0.313$, respectively), the latter is in fact the factor group with the highest $p$-value.

Table 5.16 shows the individual factors of the linear regression of only those 637 tokens that occur in monosyllabic words. The pattern of the results is similar to that of DRESS: voiceless environments have higher first formant values than voiced environments and age is the most significant predictor. The only exception is the crossed PoA MoA category: Although velar stops cause the first formant to be closer in value to the apical sounds than to the labials, the former tend to cause higher first formant values than velar stops. It would be interesting to see how these factors influence formant one within the young speakers and the monosyllabic words only for both vowels, KIT and DRESS. However, this is not possible due to an insufficient number of tokens for such models.

The result is generally in line with those reported by the studies that investigated the shift on the Canadian mainland (cf. Footnote 121). The statement that KIT is

---

123 The model’s deviance is 6518 (AIC = 6524; $R^2_{GLMM(m)} = 0.175$). Out of 645 KIT tokens, 8 had to be deleted due to an interaction between apical fricatives and style ($n = 637$).

124 The tabular overview of the results discussed here and those for the random effects can be found in Appendix F.5.2. Note that lexical item as a random effect is not significant ($p = 0.094$), but excluding it from the model results in a change in significance of the linguistic fixed effects. Hence, it has to remain in the model (cf. Subsection 4.6.3.3 and Jaeger 2008: 444).

125 Recall that velar stops favored lowering of DRESS the most, and apical stops the least (cf. Table 5.12 in Subsection 5.4.2.1).

126 Even if I did have sufficient tokens, a direct comparison of the models for the individual lax vowels would still be impossible, since my data are largely based on interview tokens. That is, the phonological environments in which /æ/ and /ɛ/ occur are not exactly the same, and neither is the ratio between these environments across the vowels. Likewise, the lexical frequencies between words that contain the vowels differ, so that /ɛ/ may be uttered in the highly frequent lexical item get, but the same words do not exist for /æ/ and /ɛ/. The imbalance could be controlled via word lists, but without reading-passage and interview data a direct stylistic comparison is not possible and the suitability of the word list is not contextualized (cf. Subsection 4.2.1.1).
### Table 5.16: REML-based Results of Linear Regression Analysis of *Kit* in Monosyllabic Lexical Items

The regression model was run in SPSS via GLMM (treatment contrasts), including speaker and word as random effects (n = 637, deviance = 6518). The p-values of the insignificant results are provided in square brackets.

<table>
<thead>
<tr>
<th>Model Term</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic t</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>536.212</td>
<td>13.679</td>
<td>39.199</td>
<td>0.000</td>
<td>[509.349, 563.075]</td>
</tr>
<tr>
<td>Age</td>
<td>-1.113</td>
<td>0.281</td>
<td>-3.961</td>
<td>0.000</td>
<td>[-1.664, -0.561]</td>
</tr>
<tr>
<td>Glottal State</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced</td>
<td>-9.895</td>
<td>4.971</td>
<td>-1.991</td>
<td>0.047</td>
<td>[-19.656, -0.134]</td>
</tr>
<tr>
<td>PoA MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apical Stops</td>
<td>7.738</td>
<td>7.170</td>
<td>1.079</td>
<td>[0.281]</td>
<td>[-6.342, 21.818]</td>
</tr>
<tr>
<td>Velar Stops</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apical Fricatives</td>
<td>-4.441</td>
<td>8.121</td>
<td>-0.547</td>
<td>[0.585]</td>
<td>[-20.389, 11.507]</td>
</tr>
<tr>
<td>Labials</td>
<td>-13.704</td>
<td>9.058</td>
<td>-1.513</td>
<td>[0.131]</td>
<td>[-31.492, 4.084]</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Passage</td>
<td>0.100</td>
<td>6.000</td>
<td>0.017</td>
<td>[0.987]</td>
<td>[-11.683, 11.883]</td>
</tr>
<tr>
<td>Word List</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>-9.895</td>
<td>5.768</td>
<td>-1.597</td>
<td>[0.111]</td>
<td>[-20.540, 2.115]</td>
</tr>
<tr>
<td>Litotal</td>
<td>-4.019</td>
<td>3.189</td>
<td>-1.260</td>
<td>[0.208]</td>
<td>[-10.281, 2.243]</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-8.503</td>
<td>8.425</td>
<td>-1.009</td>
<td>[0.313]</td>
<td>[-25.046, 8.041]</td>
</tr>
</tbody>
</table>

### 5.5. *The Shift of Kit*

Not participating in the Canadian Shift (Labov, Ash and Boberg 2006b: 220) cannot be supported with my data. Unlike suggested by Clarke, Elms and Youssef (1995: 215), following voicing is not only significant for the lowering of *dress*, but also for the lowering of *kit* in monosyllabic lexical items. However, for all *kit* tokens in my data set, none of the three factor groups, glottal state, place or manner of articulation, are significantly affecting the quality of the vowel. This result is in line with that of Clarke, Elms and Youssef (1995: 215) for Ontario – on the Canadian mainland.

In terms of style, the interpretation offered for the fact that reading-passage style seems to cause a more formal realization than word lists (cf. Subsection 5.4.2.1) seems to hold for *kit* like it did for *dress*. In the current regression analysis, the token number between reading passage and word list is not as skewed for *kit* as it is in the one of *dress*, but reading passage still causes higher first formant values than word list does. The effect is, however, so marginal that the combination of the two formal styles provides the same accuracy in terms of causing higher first formant values as the separate modeling of the two factors. In addition, style, the only indicator of social awareness of the innovation in my data, is more insignificant than glottal state of the following sound, so that it...
seems plausible that KIT lowering within monosyllabic words is rather a regular sound change. The fact that style is also not significant in the linear regression of DRESS seems to indicate that lowering of KIT is just becoming phonologically constrained, with social awareness just losing its significant role. The pattern that is (yet) inconsistent for style across the two vowels is that DRESS is lowered more often in spontaneous speech, whereas KIT is in the formal styles. This inconsistent pattern of style is, however, not significant for either vowel, while the consistent pattern of age and glottal state is significant for both vowels. LITotal and sex also behave similarly for KIT in monosyllabic lexical items as they do for DRESS in all lexical items (cf. Subsection 5.4.2.1). With regard to sex being insignificant, my result corroborates that of Clarke, Elms and Youssef (1995: 217).

I interpret the similarities between KIT in monosyllabic lexical items and DRESS as evidence for the subsequent change of the two lax vowels as part of the Canadian Shift. My data suggest consequently that a regular sound change is embedded in the last state(s) of lexical diffusion, insofar as the innovative variant from the Canadian mainland is losing social awareness within monosyllabic lexical items. As maintained by Labov (1994: 503), phonological constraints become more prominent in influencing the innovative variant (lowering of KIT) in the course of a sound change. Given the subsequent movement of TRAP, DRESS and KIT in the Canadian Shift, KIT seemingly becomes similarly constrained as DRESS, which in turn is likely to become constrained as is typical of a regular sound shift and in the present case already indicated by the behavior of TRAP. The exclusive character of either regular sound change or lexical diffusion that Labov (1994: 543) states seems thus not replicable with my data, which is in line, for instance, with studies conducted by Bybee (2002) and Oliveira (1991). Having stated that, the two indicators, style and number of following syllables, are not enough to identify and define lexical diffusion change and regular sound change sufficiently. Likewise, the differences between the two regression models for KIT lowering in multisyllabic and monosyllabic lexical items versus monosyllabic lexical items only are quite small: $R^2$ improves only from 15% to 17.5%, and the age coefficient increases only slightly from –0.959 to –1.113.

### 5.5.2.2 Logistic Regression

The model’s deviance is 1336 (AIC = 1366; $R^2_{\text{GLMM}(m)} = 0.205$ and $R^2_{\text{GLMM}(c)} = 0.462$) with 15 degrees of freedom and an intercept of –1.964. The grand mean of 0.299 indicates that 29.9% of all KIT tokens have the application value of a first formant value greater than 523 Hz (cf. Appendix E.4.2). This grand mean for KIT is smaller than that of DRESS by approximately 10% (cf. Subsection 5.4.2.2), which seems to support the interpretation that KIT is behind DRESS in its advancement within the shift.

The overall results of the logistic regression are surprisingly different from those of the linear regression discussed above. Number of following syllables is much more significant ($p < 0.001$) than age group ($p = 0.001$). All of the other factor groups are not significant: style ($p = 0.169$), LITotal ($p = 0.173$), sex ($p = 0.177$), glottal state ($p = 0.213$) and
5.5. The Shift of Kit

<table>
<thead>
<tr>
<th>Factor</th>
<th>Log Odds</th>
<th>Tokens</th>
<th>Application Rates</th>
<th>Centered Factor Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Foll. Syllables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.757</td>
<td>645</td>
<td>0.381</td>
<td>0.681</td>
</tr>
<tr>
<td>1</td>
<td>-0.070</td>
<td>603</td>
<td>0.240</td>
<td>0.483</td>
</tr>
<tr>
<td>2+</td>
<td>-0.687</td>
<td>121</td>
<td>0.157</td>
<td>0.335</td>
</tr>
<tr>
<td>Age Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>1.099</td>
<td>493</td>
<td>0.477</td>
<td>0.724</td>
</tr>
<tr>
<td>Middle</td>
<td>-0.136</td>
<td>521</td>
<td>0.265</td>
<td>0.476</td>
</tr>
<tr>
<td>Old</td>
<td>-0.963</td>
<td>355</td>
<td>0.104</td>
<td>0.25</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word List</td>
<td>0.181</td>
<td>237</td>
<td>0.426</td>
<td>[0.545]</td>
</tr>
<tr>
<td>Reading Passage</td>
<td>0.091</td>
<td>249</td>
<td>0.349</td>
<td>[0.523]</td>
</tr>
<tr>
<td>Interview</td>
<td>-0.272</td>
<td>883</td>
<td>0.251</td>
<td>[0.432]</td>
</tr>
<tr>
<td>L1total +1</td>
<td>-0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.26</td>
<td>738</td>
<td>0.337</td>
<td>[0.565]</td>
</tr>
<tr>
<td>Male</td>
<td>-0.26</td>
<td>631</td>
<td>0.255</td>
<td>[0.435]</td>
</tr>
<tr>
<td>Glottal State</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>0.157</td>
<td>906</td>
<td>0.326</td>
<td>[0.539]</td>
</tr>
<tr>
<td>Voiced</td>
<td>-0.157</td>
<td>463</td>
<td>0.248</td>
<td>[0.461]</td>
</tr>
<tr>
<td>PoA MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apical Stop</td>
<td>0.307</td>
<td>323</td>
<td>0.378</td>
<td>[0.576]</td>
</tr>
<tr>
<td>Velar Stop</td>
<td>0.247</td>
<td>273</td>
<td>0.308</td>
<td>[0.561]</td>
</tr>
<tr>
<td>Apical Fricative</td>
<td>-0.178</td>
<td>424</td>
<td>0.285</td>
<td>[0.456]</td>
</tr>
<tr>
<td>Labials</td>
<td>-0.376</td>
<td>349</td>
<td>0.238</td>
<td>[0.407]</td>
</tr>
</tbody>
</table>

Table 5.17: ML-based results of logistic regression analysis of kit in Rbrul via GLMM (sum contrasts), including speaker and word as random effects (n = 1369, deviance = 1336). The factor weights within the insignificant factor groups are provided in square brackets (n.s. = not significant).

place/manner of articulation (p = 0.232). The standard deviation of speaker as a random effect is 0.931, and that of word label is 0.828. In the linear regression above, the significances of age and number of following syllables were quite high, but the former seemed to be more significant than the latter. At a first glance, the results seem to suggest that kit is not participating in the shift when motivated predominantly by age. Instead, it seems to be primarily a function of number of following syllables. The individual factors are shown in Table 5.17.

Within the predictor number of following syllables, we find the exact same pattern that I showed above in the linear regression: It is the monosyllabic lexical items in which kit lowering is strongly favored, whereas in the other two factors it is slightly to strongly disfavored. This supports the interpretation above insofar as kit lowering is not a function of vowel length, but rather of a lexical-diffusion component in the change since not all lexical items are affected. It also affirms that the binary coding of this factor group,
which was necessary in the linear regression, did not distort the results: stressed KIT vowels disfavor lowering with one following syllable as well as with two or more following syllables. This pattern is linear and suggests that stressed /ɪ/ followed by one syllable is at the threshold to favor lowering of the vowel. This may mean that the lexical-diffusion change/component is in a rather late state in which predominantly longer multisyllabic lexical items disfavor lowering of /ɪ/.

In terms of age, the young speakers clearly and strongly favor lowering of KIT with an application rate of almost 48%. In stark contrast to them, middle-aged and old respondents from St. John’s disfavor lowering with application rates of 26.5% and 10.4%, respectively. The age group pattern supports the interpretation that KIT lowering is at a very early stage when compared to DRESS and TRAP participation in the shift in St. John’s, Newfoundland.

The apparent contradiction of the results of the linear regression outlined above and the logistic regression shown here in terms of significance of the two factor groups, number of following syllables and age, is resolved when both factor groups are plotted against one another. Figure 5.18 shows a pattern which compellingly supports the interpretation offered for the results of the linear regression above. Within all three groups of possible
syllables following the stressed vowel, there is a clear pattern of increased lowering of KIT, as the application rates of first formant values greater than 523 Hz increase. The figure suggests that the effect is not only consistent across, but also within age groups. It is, however, only the youngest speakers who show a significant increase in the application rate in monosyllabic words (zero syllables following the stressed vowel). This pattern suggests that KIT lowering is predominantly a function of age rather than a function of following syllables, as the linear regression above suggested. The logistic regression model may have yielded higher \( p \)-values for following syllables because the young speakers already show higher rates of KIT lowering when the vowel is followed by two and more syllables than the other two age groups. The figure further suggests that the sharp increase for the young speakers in monosyllabic words may mark the begin of an S-shaped curve of a regular change in progress (cf. Labov 1994: 65).

The fact that style is not significant in the model discussed here seems to support the idea that KIT lowering is a regular sound change, although it is one which started via lexical diffusion. The role of style is ambiguous throughout all the tested vowels. In the majority of cases it is not significant whatsoever, and in the few cases style is significant, it is not consistently so across different statistical modeling techniques. The inconsistent behavior of style makes it admittedly difficult to include its contribution in interpreting the data. One tendency of style in this model seems to be that word list strongly favors lowering, reading passage slightly favors, too, but interview strongly disfavors. This supports the general pattern for St. John’s, Newfoundland, that was suggested by Clarke (2012: 514) for TRAP, but it does not support the interpretation offered in Subsection 5.4.2.2 for DRESS.

In this interpretation, I outlined that respondents from St. John’s, Newfoundland, are particularly familiar with reading tasks as an instantiation of formality, so that this familiarity has them read texts more formally than word lists. The logistic regression model for KIT suggests a classic pattern of decreasing formality from word list via reading passage to spontaneous speech, but it is not significant. In contrast, the linear regression model for KIT suggests that reading-passage style causes even higher first formant values than word-list style, which is consistent with the interpretation of familiarity. I consider the same two options possible that I offered for DRESS to explain the differences in formality: 1) With regard to DRESS, the imbalance between token number for reading passage to word list might have been responsible for the strong favoring effect of the former, and 2) this favoring effect is stronger in the reading passage because of familiarity with reading so that respondents from St. John’s, Newfoundland, lose this formality first in word-list productions of DRESS and only much later in reading-passage realizations of DRESS. The second explanation is plausible, because KIT is at a earlier stage of the shift than DRESS, and because KIT seems to show the same pattern as DRESS when monosyllabic words are modeled alone, as shown above.
The general pattern of style seems to be that formal styles slightly tend to favor the innovative variants of DRESS and KIT, although this is most often insignificant. The difference in effect of word list and reading passage may not necessarily have social meaning, i.e. it could be an artifact of the data, although the effect of the two seems reversed when style is significant. The tendencies of the other insignificant factor groups, LItotal, sex, glottal state and place/manner of articulation, seem to be consistent with the linear regression shown above: non-local respondents have higher application rates for the innovative variant; females and voiceless sounds, velar stops and apical stops favor the application value, and males and voiced sounds, apical fricatives and labials disfavor a first formant value greater than 523 Hz.

The assumption that lowering of KIT is a regular sound change is given additional support by the logistic regression of those KIT tokens that occur in monosyllabic lexical items only. In this model, 8 out of 645 tokens had to be deleted due to an interaction between apical fricatives and style, leaving a total number of 637 tokens for the logistic regression of KIT in monosyllabic lexical items only. The model’s deviance is 657 (AIC 683) with 13 degrees of freedom and an intercept of −1.672. The estimates are based on REML and the grand mean is 0.374, which is almost as high as that for the logistic regression of DRESS (0.393).

The fixed predictors change their significances in the same fashion as they did in the linear regression of KIT in monosyllabic lexical items above and in the logistic regression of DRESS: age group is the most important predictor ($p < 0.001$), and it is closely followed by glottal state ($p = 0.009$). Unlike for DRESS, sex is also significantly affecting the lowering of KIT in monosyllabic lexical items ($p = 0.024$). The crossed category of place and manner of articulation is not significant ($p = 0.114$), and neither are the social categories local-ness index ($p = 0.119$) and style ($p = 0.295$). In fact, style is the least significant predictor in the model, which suggests that lowering of KIT is not a function of social awareness (anymore) within the monosyllabic lexical items. This is supported by the fact that age group and sex are the only significant predictors in this subordinate model together with glottal state. The random predictors speaker and lexical item have standard deviations of 0.829 and zero, respectively. Although a standard deviation of zero for lexical item should lead to excluding that random effect, in so doing, the significances of the predictors change, so that such an exclusion most likely leads to false positive results (cf. Subsection 4.6.3.3 and Jaeger 2008: 444).

Since age group and sex are both significant by themselves, I crossed the categories of age and sex for the model I outline below (cf. Table 5.18), as I did for TRAP (cf. Subsection 5.3.2.2). Although the amount of tokens for KIT in monosyllabic lexical items (637) is similar to the amount of TRAP tokens in total (646), the model for KIT did not show any interactions. Unlike in the model for TRAP, some of the factors in place/manner of articulations had to be recoded, after Bonferroni-corrected multiple comparisons of the first formant means per factor/variant (e.g. cf. Abdi 2007: 106; Crawley 2007: 374-
5.5. The Shift of Kit

### Table 5.18: REML-based results of logistic regression analysis of kit in monosyllabic lexical items (in Rbrul via GLMM; sum contrasts), including speaker and word as random effects (n = 637, deviance = 657). The factor weights within the insignificant factor groups are provided in square brackets (n.s. = not significant).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Log Odds</th>
<th>Tokens</th>
<th>Application Rates</th>
<th>Centered Factor Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age and Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young Females</td>
<td>1.589</td>
<td>152</td>
<td>0.658</td>
<td>0.831</td>
</tr>
<tr>
<td>Young Males</td>
<td>0.871</td>
<td>92</td>
<td>0.511</td>
<td>0.705</td>
</tr>
<tr>
<td>Middle Females</td>
<td>0.378</td>
<td>109</td>
<td>0.387</td>
<td>0.593</td>
</tr>
<tr>
<td>Middle Males</td>
<td>-0.588</td>
<td>111</td>
<td>0.252</td>
<td>0.357</td>
</tr>
<tr>
<td>Old Females</td>
<td>-0.605</td>
<td>72</td>
<td>0.194</td>
<td>0.353</td>
</tr>
<tr>
<td>Old Males</td>
<td>-1.646</td>
<td>101</td>
<td>0.089</td>
<td>0.162</td>
</tr>
<tr>
<td><strong>Glottal State</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless</td>
<td>0.382</td>
<td>397</td>
<td>0.421</td>
<td>0.594</td>
</tr>
<tr>
<td>Voiced</td>
<td>-0.382</td>
<td>240</td>
<td>0.296</td>
<td>0.406</td>
</tr>
<tr>
<td><strong>PoA MoA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apical Stop</td>
<td>0.416</td>
<td>225</td>
<td>0.449</td>
<td>[0.603]</td>
</tr>
<tr>
<td>Velar Stop</td>
<td>0.105</td>
<td>156</td>
<td>0.380</td>
<td>[0.526]</td>
</tr>
<tr>
<td>Apical Fricative</td>
<td>-0.143</td>
<td>100</td>
<td>0.330</td>
<td>[0.464]</td>
</tr>
<tr>
<td>Labials</td>
<td>-0.379</td>
<td>156</td>
<td>0.327</td>
<td>[0.406]</td>
</tr>
<tr>
<td><strong>LItotal +1</strong></td>
<td>-0.235</td>
<td></td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td><strong>Style</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word List</td>
<td>0.220</td>
<td>131</td>
<td>0.481</td>
<td>[0.565]</td>
</tr>
<tr>
<td>Reading Passage</td>
<td>0.042</td>
<td>138</td>
<td>0.391</td>
<td>[0.489]</td>
</tr>
<tr>
<td>Interview</td>
<td>-0.263</td>
<td>368</td>
<td>0.329</td>
<td>[0.445]</td>
</tr>
</tbody>
</table>

The individual factors are shown in Table 5.18. The results of the individual factors of the crossed age and sex group show a classic regular sound change pattern: young females lead the change with lowering rates of 66%, i.e. two thirds of their kit tokens have a first formant value greater than 523 Hz (cf. Figure 5.19). These lowered kit tokens occur most likely before voiceless consonants. The young females are followed by their male counterparts with an application rate of 51%, i.e. half of their tokens have first

377), which is most likely responsible for the absence of interaction. The relatively large number of factors within the crossed age and sex group did not seem to be problematic.

The model’s characteristics change slightly compared to the one with age and sex as individual factor groups. The deviance is the same (657, AIC = 687; \( R^2_{\text{GLMM}(m)} = 0.274 \) and \( R^2_{\text{GLMM}(c)} = 0.401 \)), but the number of degrees of freedom is higher (15). The intercept is −1.632 and the grand mean remains unchanged at 0.374. The predictors’ \( p \)-values changed accordingly: As expected, the crossed age and sex group is highly significant \( (p < 0.001) \) and is directly followed by glottal state \( (p = 0.009) \). The other predictors are slightly less significant than in the model with age and sex run as separate factor groups (PoA MoA, \( p = 0.116 \); LItotal, \( p = 0.151 \); style, \( p = 0.307 \)).

The individual factors are shown in Table 5.18.
formant values greater than 523 Hz. The use of the innovative KIT variant seems to have emanated from the middle-aged females, as their first formant values also already favor the application value with 39% of their tokens being lowered. In contrast, middle-aged males still disfavor the application value, as only 25% of their KIT tokens are lowered when 523 Hz is the threshold. The difference of 14% between the middle-aged females (39%) and the middle-aged males (25%) is almost as high as the one between young females and males (15%). Similarly, the old females disfavor KIT lowering with an application rate of 19%, which is about 10% more than the application rate of the old males (9%). The differences between the conservative males’ realization and the innovative females’ realization of KIT increases slowly, but linearly and inversely with age groups.

Lowering of KIT within monosyllabic environments is a realization option for both sexes alike. Since the difference between the sexes is not very substantial, it is surprising that sex is significant in the model in the first place. No such significant effect was found for DRESS, although the behavior of DRESS should be indicative of what innovative KIT will most likely be constrained by in time. One possible explanation is that on the verge of a lexical diffusion change to a sound change that encompasses all lexical items regardless of their number of following syllables, men may have caught up with women in terms of
realization rates of the innovative variant. That is, the lexical diffusion phase may need to be social reality for all speakers before another phase of sound change can be entered. Thus, in the early stages of a regular sound change, i.e. stages as the one dress is in at present, females lose their lead temporarily and may speed up again in time. This interpretation is, however, not supported by the linear regression of kit in monosyllabic lexical items mentioned above.

Hence, sex’s significance could also have no social meaning. It may be that the finding is due to the number of tokens: The analyses of kit are based on one-third as many tokens as these of dress. Another possible explanation is that the significance of sex is due to logistic regression modeling, as it was not significant in the linear model of kit in monosyllabic lexical environments (cf. Table 5.16 in Subsection 5.5.2.1). This interpretation is supported by the fact that both models are quite similar in terms of the remaining factor groups and even the individual factors. In addition, similarly to dress, age and glottal state are the only two significant factor groups.

In any case, the logistic regression of kit in monosyllabic lexical items has shown that within this environment, a regular sound change pattern is visible, which is embedded in a lexical diffusion change. This finding is in line with those of Bybee (2002) and Oliveira (1991), challenges Labov’s (1994) complementary-distribution classification of the two changes, but is in line with his classification of the two changes being polar types with many intermediate combinations. The regular sound-change pattern is remarkably similar to the one shown by dress, which underlines the subsequent nature of the shifting behavior of the participating lax vowels. The only linguistic factor group to be significant is glottal state. Following voiceless consonants favor lowering of kit with an application rate of 42%. The crossed category of place and manner of articulation is insignificant, but shows tendencies similar to the one suggested by the linear regression of kit in monosyllabic lexical items: apical and velar stops favor the application value (application rates of 45% and 38%, respectively); labials and apical fricatives disfavor the application value (both have an application rate of approximately 33%). Lltotal and style do also not play a significant role in the innovative realization of kit: The former suggests that non-local respondents are more likely to utter kit with first formants higher than 523 Hz, and the latter displays the a priori pattern for Newfoundland (cf. e.g. Clarke 2012: 514 for innovative trap): Word list and reading passage favor the innovative variant and interview style disfavors lowering of kit.

5.6 The Canadian Shift in St. John’s

Figure 5.20 visualizes the results of the analyses outlined in the previous subsections. All of the lax vowel movements per age group have social meaning. That is, statistical assessment confirms that the young speakers differ significantly from the old (and variably middle-aged) speakers in their realization of the three lax vowels. Their retraction of
merged LOT-THOUGHT is not significantly different from the position of the middle-aged and old speakers, which suggests that the position as far back as 1275 Hz Labov, Ash and Boberg (2006b: 219) have shown for mainland Canadian LOT-THOUGHT is not necessarily a prerequisite for the chain shift to commence. The acoustic position of the merged vowel phoneme for the young speakers is at 1315 Hz rather close to the atlas’ of 1275 Hz, unlike the acoustic position they suggested for a 35-year-old case study from St. John’s, Newfoundland, of more than 1500 Hz (2006b: 221). A movement of STRUT as part of the Canadian Shift, which Clarke, Elms and Youssef (1995) found in Ontario, cannot be replicated with my data. This result was also reported by, for instance, Labov, Ash and Boberg (2006b: 220). Wedge /ʌ/ is in an acoustically stable position in apparent time and thus served as a point of reference for the Euclidean Distance (ED) metric.

5.6.1 The Shift across Age Groups

The lax vowel movements that are agreed upon to participate in the Canadian Shift, TRAP, DRESS and KIT, were first analyzed only with regard to age groups. At least for the young speakers, a significant difference in their realizations to those of the old speakers could be attested. I conducted these analyses via the EDs between each lax vowel and STRUT. The EDs were based on the first and second formant means per speaker (cf.
Table 5.19: Formant means of the three lax vowels and merged LOT-THOUGHT per age group (n = 34; ED = Euclidean Distance; Sd = standard deviation; *Euclidean Distances between each lax vowel and STRUT and between LOT and THOUGHT, respectively).

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Age Group</th>
<th>F1 Mean</th>
<th>F1 Sd</th>
<th>F2 Mean</th>
<th>F2 Sd</th>
<th>ED* Mean</th>
<th>ED* Sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOT-THOUGHT</td>
<td>Old</td>
<td>796</td>
<td>35</td>
<td>1360</td>
<td>140</td>
<td>111</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>803</td>
<td>30</td>
<td>1358</td>
<td>89</td>
<td>49</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td>812</td>
<td>34</td>
<td>1315</td>
<td>63</td>
<td>36</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>805</td>
<td>33</td>
<td>1343</td>
<td>95</td>
<td>61</td>
<td>51</td>
</tr>
<tr>
<td>TRAP</td>
<td>Old</td>
<td>774</td>
<td>55</td>
<td>1952</td>
<td>65</td>
<td>611</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>766</td>
<td>41</td>
<td>1849</td>
<td>120</td>
<td>518</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td>803</td>
<td>62</td>
<td>1725</td>
<td>45</td>
<td>359</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>782</td>
<td>56</td>
<td>1828</td>
<td>123</td>
<td>481</td>
<td>151</td>
</tr>
<tr>
<td>DRESS</td>
<td>Old</td>
<td>631</td>
<td>43</td>
<td>2176</td>
<td>101</td>
<td>839</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>660</td>
<td>55</td>
<td>2028</td>
<td>137</td>
<td>642</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td>684</td>
<td>33</td>
<td>1970</td>
<td>133</td>
<td>595</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>660</td>
<td>50</td>
<td>2048</td>
<td>151</td>
<td>677</td>
<td>177</td>
</tr>
<tr>
<td>KIT</td>
<td>Old</td>
<td>497</td>
<td>28</td>
<td>2279</td>
<td>127</td>
<td>963</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>508</td>
<td>28</td>
<td>2141</td>
<td>75</td>
<td>792</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td>527</td>
<td>17</td>
<td>2059</td>
<td>83</td>
<td>707</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>511</td>
<td>27</td>
<td>2151</td>
<td>126</td>
<td>808</td>
<td>150</td>
</tr>
</tbody>
</table>

Boberg 2010: 146, 201, 2011: 22). In order to assess the relative contributions of first and second formant to the difference in realizations, I employed Kendall’s correlation coefficient and tested the correlation between age and formant values statistically. The mean formant values of the individual speaker means per age group, the EDs and the standard deviations are shown in Table 5.19.

Statistical assessment of LOT and THOUGHT suggested the merged status of the two vowel classes; their merger is thus no longer an ongoing process in St. John’s, Newfoundland. Merged LOT-THOUGHT did not turn out to be significantly moving in second formant space: as Table 5.19 shows, the merged vowel is slightly retracted to an acoustic position of 1315 Hz in apparent time, which is by no means a front position as maintained by Labov, Ash and Boberg (2006b: 221).

Although Kendall’s \( \tau \) suggested no significant difference in the first formant values of TRAP in apparent time, it did so in the second formant values. Young speakers realize TRAP in a significantly more retracted position than both the middle-aged and old speakers. Their mean second formant value of 1725 Hz is similar to that of mainland Canada, as stated by Boberg (2010: 145) and Labov, Ash and Boberg (2006b: 220).

Despite the lack of significant differences in mean EDs between old and middle-aged speakers with regard to TRAP, in the realization of DRESS and KIT, middle-aged and young speakers pattern together. This result is due to the use of EDs in order to statistically
assess the movement of a vowel. It is a means to establish movement in apparent time, but not to analyze fine-grained differences in terms of age. As I will summarize below, the latter two lax vowels are at a much less advanced stage in the Canadian Shift than TRAP. Kendall’s correlation coefficients showed that both formants of the two lax vowels significantly correlate with age.

Although the logistic regression modeling of KIT suggested it is not age that has the strongest effect on its movement in first formant space, all of the other statistical tests suggested the opposite. It has become clear that KIT is participating in the shift, but to a lesser degree than DRESS. Both vowels’ first formant means are not as high as those reported by Boberg (2010: 145) for mainland Canada, but the mean for DRESS is within the threshold of 650 Hz established by Labov, Ash and Boberg (2006b: 219).

5.6.2 Generalized Mixed-effects Models Compared

Mixed-effects modeling of LOT and THOUGHT supported the notion that these vowels are merged in St. John’s, Newfoundland. In line with Hypothesis 2 (cf. Section 3.4), the merger is not a change in progress and thus, with the exception of some older speakers, completed. The results of the linear and logistic regression models of the lax vowels involved, i.e. TRAP, DRESS and KIT (cf. research question a), provide mixed support for those reported in the literature. The significant predictors responsible for the change in formant values per lax vowel are summarized in Table 5.20.

Research questions b) and c) and the respective hypotheses (cf. Section 3.4) will be addressed in the remainder of this subsection. For all three lax vowels, age is the most important predictor independently of any of the other social predictors, as none of their interactions with age is significant (cf. Hypothesis 7). This finding is in line with that of Clarke (1991: 112) for other innovations in St. John’s, Boberg (2005: 141) for the Canadian Shift in Montreal, and Boberg (2010: 147, 2011: 24) and Labov, Ash and Boberg (2006b: 221) for the Canadian Shift from coast to coast.

The dependent variable for TRAP retraction is formant two. The difference in the results of the two regression models for that lax vowel is minimal. Both suggest significant effects for age and sex only. The pattern shows a regular sound change with females so far in the lead that middle-aged females are even more advanced than young males, as the logistic regression model results suggest (cf. Hypotheses 4, 7 and 8). Style is insignificant, but showed a pattern as expected for regular changes: retraction is most advanced in interview style (cf. Subsection 5.3.2). This indicates that the innovative retracted variant is present in the vernacular of young respondents from St. John’s, Newfoundland (cf. Hypothesis 9).

For the remaining two lax vowels, the dependent variable is formant one. For DRESS, the results between the models differ insofar as style is significant in the logistic regression model, but not in the linear regression model. The status of the more significant
5.6. The Canadian Shift in St. John’s

Table 5.20: Tabular comparison of the results of linear and logistic regression for the lax vowels, TRAP, DRESS and KIT.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Linear Regression</th>
<th>Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predictor</td>
<td>p-Value</td>
</tr>
<tr>
<td>TRAP</td>
<td>Age (cont.)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRESS</td>
<td>Age (cont.)</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Glottal State</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>Voiceless</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Voiced</td>
<td>-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIT</td>
<td>Age (cont.)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td># of Foll. Syll.</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Style</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>Formal</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Interview</td>
<td>-9.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

predictors, age and glottal state, is less uncertain: undoubtedly, both play the most significant role in favoring the innovative lowered variant. Judging from the different p-values between the two regression models, the fact that style is significant only in the logistic regression model seems to be a result of categorizing age into three groups and formant one into values greater and smaller than 669 Hz. In this regression model, the p-values are much more similar for all three predictors and greater than in the linear regression model (cf. Hypotheses 5, 7 and 9).

In terms of KIT, the difference is less easily attributable to the different categorizations of age and formant one than for DRESS. Assuming that categorization may play a similar role, it seems that it reduced the significance of age group and style dramatically. In this regard the pattern is similar to that of DRESS, as the p-values for age and style decrease,
while that of number of following syllables seems to remain at its high significance. From these findings, I concluded that lowering of KIT is at a less advanced stage than that of DRESS. An investigation of KIT vowels only in monosyllabic lexical items revealed that within these, the pattern suggested by DRESS seems to hold also for KIT: Age and glottal state are the significant predictors, with a variably significant contribution of sex. Only the logistic regression model suggested a significance of sex, while at the same time the model confirmed the results of the linear regression that age and glottal state are the most significant predictors (cf. Hypotheses 6, 7, 8 and 9).

Recently, Clarke (2012: 514) has stated that TRAP retraction is strongly favored in the formal styles of her female respondents recorded in the 1980s. She has further emphasized that the few innovative realizations she could find (application rate 29%; also cf. Clarke 1991: 118) were predominantly produced in the upper social class. In the more than 25 years that have passed since the 1980s, the innovative realization of this variant is particularly evident in interview style (spontaneous speech) of the speech of young middle-class women. This may suggest that after one generation, the innovation has entered the vernacular speech of young speakers of St. John’s English (as witnessed by a very high application rate of 79.4%).

In terms of DRESS, the innovative variant is significantly more often uttered in formal styles than in interview style – at least in the logistic regression model. This pattern is quite similar to that reported by Clarke (2012: 514) for the realization of TRAP in St. John’s, so that it can be understood as the next vowel to be shifting in this fashion. That is, in some 20 years from now, DRESS should display the same pattern of variation in the informal style of speakers of St. John’s English as TRAP does in my data set recorded in 2011. The fact that style is not significant in the linear regression of DRESS may mean that its innovative lowered realization is on the verge of entering the vernacular system of the St. John’s English speakers. It may, however, also mean that the logistic model returns p-values that are due to the make-up of the categories (e.g. categorized dependent variable, categorized age groups), rather than that they are a reflection of the social reality of the speakers: Continuous modeling of age results in much smaller p-values than modeling it in groups ($p = 0.002$ versus $p = 0.014$), which seems to have caused glottal state and style to become similarly significant ($p = 0.016$ and $p = 0.019$, respectively).

With regard to KIT, the transitional stage of DRESS in my data in the course of the Canadian Shift seems to be supported. The linear regression results suggest that age and number of following syllables are the most significant predictors for the innovative lowered realization of KIT in St. John’s English, closely followed by style. Again, the innovative realization is significantly more often uttered in the formal styles, which supports the interpretation that the current realization of KIT indicates the position where DRESS used to be, and that the realization of TRAP indicates the position that DRESS is moving towards. In addition to the innovative realization of KIT occurring primarily in formal
styles, it also occurs primarily in monosyllabic lexical items. This may suggest that the innovative feature (kit lowering) has entered the linguistic system of speakers from St. John’s via lexical diffusion. In such changes, the innovative variant occurs (not necessarily abruptly) in some (monosyllabic) lexical items, but not in others (e.g. in bit, but not in bid), and then (gradually) spreads to all lexical items containing the kit phoneme, regardless of whether these items are mono- or multisyllabic.

Although Labov (1994: 543) variably argues for the mutually exclusive character of regular sound change and lexical diffusion change (complementary distribution or alternatively polar types with intermediate combinations; 1994: 542), Bybee (2002) and Oliveira (1991) – among others – present evidence that both may occur at the same time, that one may well precede the other and that they may gradually change into one another. The innovative variant of kit in St. John’s English is primarily conditioned by age, following voicing and sex within monosyllabic lexical items (cf. Subsection 5.5.2) and thus shows a regular sound change pattern within lexical diffusion change, as proposed by Bybee (2002) and Oliveira (1991).

The innovative kit variant may eventually be traceable as a ‘pure’ regular sound change, similarly to dress, meaning that multisyllabic lexical items will also contain more of the innovative realization (application rate of 25% in my data; cf. Subsection 5.5.2.2). The behavior of kit is thus remarkably similar to the behavior of trap when the latter was uttered by the young females recorded in the 1980s (application rate of 29%; cf. Clarke 1991: 118, 2012: 514). In terms of style, it is far from clear why it seems to be insignificant in the logistic regression model, although it is significant in the linear regression. Style in the regression modeling of kit behaves opposite to what was observed in the regression modeling of dress. As offered above, I understand this difference between the findings of the two regression models most likely to be an artifact of the categorization process of the dependent variable and age, rather than to be socially meaningful.

In a brief pre-conclusion, the answer to the main research question of whether the innovative, mainland Canadian English vowel-shift pattern is present in St. John’s (cf. Section 3.4) is affirmative. I propose that this pattern shows intermediate combinations of lexical diffusion change characteristics and of regular sound change characteristics. Whether the Canadian Shift is one or the other, which is clearly not part of my research question and thus beyond the scope of this dissertation, cannot be answered with my data for St. John’s, and it cannot be answered with the data available for mainland Canada investigated in other studies, which I presented in Section 2.3. The characteristics of the Canadian Shift that are known and that have been replicated across studies for the Canadian mainland are: young urban middle-class speakers lead in the retraction of trap, lowering of dress and lowering of kit; for trap, females lead; for dress, sex plays a variably significant role; and for kit, either males or females lead. These
characteristics are predominantly based on word-list style data. A similar, if not in large part the same, pattern is true for St. John’s Newfoundland.

An extension to this answer can be derived when Clarke’s (1991, 2012) findings for St. John’s are added, when the behavior of style in my data is taken to be socially meaningful and when the difference of kit lowering in mono-/multisyllabic versus monosyllabic words is considered as a strong indicator of lexical diffusion: In this case, the Canadian Shift seems to have entered the linguistic system of speakers from St. John’s, Newfoundland, as a lexical diffusion change above the level of social awareness via the formal styles, beginning in the upper classes (cf. Clarke 2012: 514) and changing into a regular sound change pattern after one generation or two (cf. Hypothesis 1). The fact that the middle-class speakers included in this analysis already display application rates of the innovative variant up to 40% in monosyllabic lexical items seems to show that even lowered kit has spread to the middle of the social hierarchy, which will continue the change in progress in a regular fashion, as suggested by dress and trap, i.e. young middle-class females will eventually lead the change independently of lexical item (and frequency; cf. Hofmann and Wagner in prep.).

This interpretation also seems to be supported by the marginal $R^2_{GLMM}$ values of the individual mixed models. Table 5.21 summarizes and juxtaposes each of Nakagawa and Schielzeth’s (2013) $R^2_{GLMM(m)}$ with $R^2$ obtained from the respective fixed-effects models without the random effects. I modeled each of the lax vowels with Rbrul (Johnson 2009) traditionally, in a fashion similar to the VARBRUL programs, in order to compare their $R^2$ values with those outlined by Labov, Ash and Boberg (2006b). They state that the factors they included in their analysis of the Canadian Shift “accounted for 30 to 50 percent of the variance for the sound changes in progress” (2006b: 220). Although they do not explicitly state how they calculated $R^2$ values for their analyses, it is apparent nowhere in the atlas that any random effects were taken into consideration in their regression analyses. This is not surprising given that: First, traditional VARBRUL software does not offer any possibility to include random effects (cf. Johnson 2009; Saito 1999), second, the relatively recent proliferation of mixed-effects modeling in variationist sociolinguistics (e.g. cf. Gorman and Johnson 2013; Jaeger 2008) would very likely have led the authors to place special emphasis on employing such modeling, and third and most importantly, there is (still) no consensus on how to define and consequently calculate $R^2$ in mixed-effects models (Nakagawa and Schielzeth 2013: 134; also c.f. Kramer 2005: 150).

According to Labov, Ash and Boberg’s Table 15.1 (2006b: 221), only the vowels /æ/, /e/ and /i/ show a significant age gradient for formant one (/æ/ and /i/) or formants

---

127 Marginal $R^2_{GLMM}$ ($R^2_{GLMM(m)}$) refers to the variation explained by just the fixed effects in a mixed-effects model, and conditional $R^2_{GLMM}$ ($R^2_{GLMM(c)}$) refers to the variation explained by fixed and random effects in a mixed-effects model (Nakagawa and Schielzeth 2013: 137).
one and two (/e/).\footnote{Labov, Ash and Boberg’s vowels /æ/, /e/ and /i/ correspond to the vowels /æ/, /e/ and /i/ and Well’s (1982c) lexical sets TRAP, DRESS and KIT, respectively. The age gradient for formants one and two of /o/ (i.e. merged LOT-THOUGHT) is not significant.} According to their text, only /æ/ and /e/ are part of the sound change in progress, the Canadian Shift, after it has been triggered by the low-back merger (Labov, Ash and Boberg 2006b: 220), so that the 30\% to 50\% of the variance accounted for most likely refer to the vowels /æ/ and /e/ only.

Since $R^2$ is understood to be a standardized effect statistic (Nakagawa and Schielzeth 2013: 134), it can be used to compare explanatory values of regression models across different data sets. As Table 5.21 shows, $R^2$ of DRESS (/e/) in the fixed-effects linear regression model is 0.278, and Nagelkerke’s $R^2$ of DRESS in the fixed-effects logistic regression model is 0.299. 27.8\% and 29.9\%, respectively, of variance explained by the

\begin{table}[ht]
\centering
\begin{tabular}{llcccccc}
\hline
\textbf{Vowel} & \textbf{Model Details} & \multicolumn{2}{c}{\textbf{Linear Regression}} & \multicolumn{2}{c}{\textbf{Logistic Regression}} \\
& & \textbf{Mixed-effects} & \textbf{Fixed-effects} & \textbf{Mixed-effects} & \textbf{Fixed-effects} \\
\hline
\text{TRAP} & Deviance & 7755 & 8,259,389 & 519 & 576 \\
& AIC & 7761 & 7968 & 551 & 604 \\
& \textit{Degrees of Freedom} & 11 & 12 & 16 & 14 \\
& Intercept & 1720 & 1888 & -1.166 & -1.032 \\
& Grand Mean & 1860 & 1835 & 0.395 & 0.395 \\
& $R^2_{\text{GLMM(m)}}/R^2$ & 0.394 & 0.482 & 0.434 & 0.491* \\
\hline
\text{DRESS} & Deviance & 11,221 & 4,282,845 & 990 & 1141 \\
& AIC & 11,227 & 11,692 & 1020 & 1167 \\
& \textit{Degrees of Freedom} & 9 & 11 & 15 & 13 \\
& Intercept & 731 & 685 & 0.459 & 0.187 \\
& Grand Mean & 659 & 649 & 0.393 & 0.393 \\
& $R^2_{\text{GLMM(m)}}/R^2$ & 0.147 & 0.278 & 0.233 & 0.299* \\
\hline
\text{KIT} & Deviance & 14,021 & 2,722,631 & 1336 & 1426 \\
& AIC & 14,027 & 14,305 & 1366 & 1452 \\
& \textit{Degrees of Freedom} & 9 & 10 & 15 & 13 \\
& Intercept & 540 & 516 & -1.964 & -1.556 \\
& Grand Mean & 510 & 502 & 0.299 & 0.299 \\
& $R^2_{\text{GLMM(m)}}/R^2$ & 0.15 & 0.191 & 0.205 & 0.232* \\
\hline
\text{KIT (monosyll.)} & Deviance & 6518 & 1,255,976 & 657 & 686 \\
& AIC & 6524 & 6662 & 687 & 712 \\
& \textit{Degrees of Freedom} & 9 & 10 & 15 & 13 \\
& Intercept & 536 & 522 & -1.632 & -1.499 \\
& Grand Mean & 513 & 510 & 0.374 & 0.374 \\
& $R^2_{\text{GLMM(m)}}/R^2$ & 0.175 & 0.255 & 0.274 & 0.295* \\
\hline
\end{tabular}
\caption{Summary of Nakagawa and Schielzeth’s (2013) marginal $R^2_{\text{GLMM}}$ values of each of the lax vowels’ mixed-effects models and juxtaposition with $R^2$ values of the lax vowels’ fixed-effects models (*Nagelkerke’s $R^2$).}
\end{table}

5.6. The Canadian Shift in St. John’s
models for **dress** are almost identical to the $R^2$ value of 30% Labov, Ash and Boberg (2006b: 220) outline. The same is true for **trap**: $R^2$ in the fixed-effects linear regression model is 0.482 and Nagelkerke’s $R^2$ in the fixed-effects logistic regression model is 0.491, i.e. 48.2% and 49.9% of variance explained by the models for **trap** is almost identical to the $R^2$ value of 50% outlined by Labov, Ash and Boberg (2006b: 220).

In my data set, **kit** is also part of the Canadian Shift and shows $R^2$ values of 0.191 in the fixed-effects linear regression model and of 0.232 in the fixed effects logistic regression model, i.e. roughly 20% of the variance is explained by the models. This is rather low compared to $R^2$ for **trap** and **dress**, but the regression models of **kit** in monosyllabic lexical items yield $R^2$ values of roughly 26% and 30%, respectively. These values are close to the ones for **dress**, indicating that within monosyllabic lexical items, lowering of **kit** can be explained by the same social and linguistic factors as lowering of **dress**, i.e. those that are typical of regular sound change. In that fashion, $R^2$ values support the interpretation of **kit** lagging behind **dress**, which in turn lags behind **trap** in terms of advancement in the change in progress.

In contrast to the $R^2$ values derived from the fixed-effects models, $R^2_{GLMM(m)}$ for mixed-effects models are 2% to 13% smaller in value.\(^\text{129}\) However, the relation between the $R^2_{GLMM(m)}$ values of the three lax vowels is identical to that of the fixed-effects $R^2$ values. The reasons for the difference between $R^2_{GLMM(m)}$ and $R^2$ values may be due to several facts, e.g. inclusion of additional (random) effects, the definition of $R^2$ in mixed-effects models, the manner in which $R^2_{GLMM(m)}$ is calculated, etc. The difference can thus hardly be interpreted as that the lower $R^2$ values indicate less suitable predictors in my mixed-effects models than in the fixed-effects models of Labov, Ash and Boberg (2006b). It has become evident that the standardized effect statistic $R^2$ can only be compared to other $R^2$ values that have been obtained from the same statistical modeling, e.g. within mixed-effects models, fixed-effects models and ANOVAs (cf. $R^2$ values obtained from ANOVAs with Euclidean Distances outlined in Subsections 5.3.1, 5.4.1 and 5.5.1). The comparability of $R^2$ values between different data sets has thus to be ascertained by similar statistical modeling and – more importantly – by the manner of calculating $R^2$, e.g. McFadden’s $R^2$, Nagelkerke’s $R^2$, Nakagawa and Schielzeth’s $R^2_{GLMM}^*$, etc.

The interpretation that **dress** and **kit** have begun to lower in St. John’s as lexical diffusion changes and now gradually turn into regular changes is based on four indicators: 1) Clarke (1991: 116) stated that **trap** retraction is most often found in the formal styles in the 1980s, which most likely led her to propose that incipient changes in St. John’s or Newfoundland are changes from above (2012: 514); this would suggest that retracted

\(^{129}\) For **trap**, the regression models yield $R^2_{GLMM(m)}$ values of 0.394 ($R^2_{GLMM(c)} = 0.655$) for the linear regression model and of 0.434 ($R^2_{GLMM(c)} = 0.659$) for the logistic regression model; for **dress**, $R^2_{GLMM(m)} = 0.147$ ($R^2_{GLMM(c)} = 0.615$) and $R^2_{GLMM(m)} = 0.233$ ($R^2_{GLMM(c)} = 0.566$), respectively; for **kit**, $R^2_{GLMM(m)} = 0.15$ ($R^2_{GLMM(c)} = 0.424$) and $R^2_{GLMM(m)} = 0.205$ ($R^2_{GLMM(c)} = 0.462$), respectively; for monosyllabic **kit**, $R^2_{GLMM(m)} = 0.175$ ($R^2_{GLMM(c)} = 0.392$) and $R^2_{GLMM(m)} = 0.274$ ($R^2_{GLMM(c)} = 0.401$), respectively.
TRAP has entered the linguistic system via the formal styles in the 1980s because TRAP is most often retracted in the vernacular in my data from 2011; 2) lowering of DRESS (only in the logistic regression) and lowering of KIT (in both the linear and logistic regression) are favored in the formal styles in my data; 3) KIT lowering is favored in monosyllabic lexical items, but not all lexical items; and 4) the regression models and the $R^2$ values suggest a stage-like pattern for the shift of the three vowels. The two strongest indicators of the four are style as an indicator of social awareness (change from above = lexical diffusion change) and that monosyllabic lexical items show lowered KITs that are constrained by factors typical of regular sound change.

In terms of the style, more perceptual data is needed to substantiate the presence of social awareness in the Canadian Shift vowels for middle-class St. John’s residents than merely style as attention to speech, particularly since the role of style cannot be maintained across the triangulation of the results via linear and logistic regression: lowering of DRESS is favored in the informal style in the linear regression. Moreover, other perceptual data might reveal an inconsistent pattern in the results for style: My respondents, for instance, answered the question of whether they feel to be Newfoundlanders or Canadians with a uniform “Newfoundlander first and Canadian second”. In terms of style, not a single study investigated the stylistic profile of the Canadian Shift on the mainland for a large number of speakers. Although Boberg (2008b, 2010) has such data available in his PCE study of 108 young middle class respondents, he does not analyze spontaneous speech (cf. Section 2.3). Further, change from above refers to above the level of social awareness and the highest social class simultaneously (cf. Section 2.2). Not a single study investigated the social-hierarchy constraints in the Canadian Shift on the Canadian mainland in order to provide empirical data as to whether the interior social groups are leading this change. Labov, Ash and Boberg (2006b) have such data for 33 Canadian respondents, but only report on the role of age (cf. Section 2.2), the only factor that is uncontested in virtually all studies on the Canadian Shift, including mine for St. John’s, Newfoundland.

In terms of number of following syllables, the differences between the two regression models for KIT lowering in multi-/monosyllabic versus monosyllabic lexical items only are quite small: both the $R^2$ value and the age coefficient improve only slightly. Furthermore, age is virtually always the strongest indicator for the innovative variant, which does not necessarily have to be very strong in changes from above, as anyone can learn the prestigious variants (abrupt change). Likewise, following and preceding place and manner of articulation as well as following glottal state have been variables in the atlas (Labov, Ash and Boberg 2006b), of which usually manner constrains the innovation in change from below (cf. Labov, Ash and Boberg 2006b: 197), but the authors do not report on the constraints of the linguistic variables for the Canadian Shift on the mainland.

In essence, some indicators for lexical diffusion change are present in my data. They co-occur with some indicators of regular sound change in my data. In sum, there is
neither enough evidence available for the behavior of the Canadian Shift on the Canadian mainland in this regard, nor are style and number of following syllables sufficient to undoubtedly ascribe the label ‘lexical diffusion change only’ or ‘regular sound change only’ to the Canadian Shift in St. John’s. The pattern found and described here thus is a sound change that rather shows intermediate combinations of characteristics of lexical diffusion and regular sound change. Thus, Clarke’s claim that incipient changes in St. John’s are changes from above (2012: 514) is too strong a claim to make; and if this claim was valid, there is no data available that would help to refute it for the Canadian Shift on the mainland.

5.6.3 Examples of St. John’s English from Case Studies

The following case studies provide a complement to the quantitative analyses discussed and summarized above. I chose one representative for each age group, young, middle and old, in order to compare the vowel systems of individual speakers, instead of merely relying on groups of speakers categorized by age and sex. The individual vowel plots cannot provide the generalizable evidence of systematic variation that comparisons of speaker groups can (cf. Thomas 2001: 4), but their exemplary character can support the latter. For the sake of illustrating the rather extraordinary differences between the vowel space of innovative speakers and conservative speakers of the variety of St. John’s English, I selected three rather extreme respondents from my data set for individual analysis. ‘Extreme’ has to be understood in the sense that the youngest and oldest speakers outlined below are so innovative and so conservative, respectively, in their realization of the Canadian Shift vowels that the continuum created between these two extremes covers a broad range of possible linguistic behaviors, from traditional St. John’s English to participation in the shift.

The middle-aged speaker, discussed in Subsection 5.6.3.2, cannot be positioned in the middle of the continuum. His lax front vowels are realized in a quite conservative fashion typical of St. John’s rather than in a neutral fashion typical of North American English vowels in their initial position (cf. Labov, Ash and Boberg 2006b: 12), i.e. before any systematic vowel variation commences. Thus, he has to be positioned much closer to the extreme pole symbolized by the old speaker within the continuum of (non-)participation in the Canadian Shift in St. John’s. Having stated that, the middle-aged speaker is nevertheless quite distinct from the old speaker outlined in Subsection 5.6.3.3.

Predominantly the female middle-aged speakers show individual vowel systems that are in stark contrast to the middle-aged male presented here concerning participation in the Canadian Shift. Since the quantitative analysis generally suggested that the participation in the Canadian Shift as a whole is still a domain of the youngest speakers, including an innovative middle-aged female in this discussion of case studies would have provided an inaccurately skewed picture of the progress of the shift in the speech com-
munity of St. John’s. Statistically, the advancement of middle-aged females is only significant with regard to the retraction of TRAP, but not with regard to the Canadian Shift as a whole. The picture the case studies below provide, namely that of profound differences between innovative young females and conservative middle-aged to old males, conforms more appropriately to the picture drawn by the quantitative analysis.

Unlike Boberg, I decided to include all of the vowels of the three speakers that I measured acoustically in the plots. As he rightly observes, such a scatterplot of a vowel system tends to lose focus (2010: 232). Boberg’s presentation of the individual vowel plots is a result of a procedure in which “[...] only the most notable features of each vowel system have been selected for display [...]” (2010: 232). That is, he provides some selected individual vowels for the features he discusses in more detail (e.g. the Canadian Shift or GOOSE fronting) and only formant means for the remaining vowels. The additional advantage of such a presentation is that only those “most notable features” are shown that fit the desired interpretation. This is definitely not to imply or insinuate that he manipulated his data, but to outline the reason why I decided against pre-selecting the “most notable features” for the (innovative) vowel systems of speakers from St. John’s. I consider this decision additionally of particular importance, because the presence of the Canadian Shift in this speech community is contested. It is not contested in the Canadian regions Boberg illustrates with his case studies.

The drawback of my decision is the lost focus in the vowel plots, especially since my graphs include all of the lexical items in which the respective vowels occur. However, the idea of the respective scatterplots is not to represent each and every single lexical item legibly, but to offer a general impression of the fields of dispersion of those vowels that are understood to participate in the Canadian Shift in some exemplary lexical items. For this reason, I chose two different sets of symbols to indicate the position of the vowels in the F1xF2 plots below. The symbols for merged LOT-THOUGHT, TRAP, DRESS and KIT are simple triangles and squares, whereas those for the other vowels in the plots are modified by ‘x’ and cross symbols, filling the triangular and square shapes. This should allow the reader to differentiate between the vowels of the Canadian Shift and the remaining ones, which serve predominantly to illustrate the general outline of the vowel space.

In addition to the fields of dispersion of the vowels in the graphical representations, I provide formant mean values and standard deviations (sds) in the text in order to ease comparison of the average positions of the vowels between the individual speakers. Each of the individual vowel plots is further complemented by the social background of each of the speakers. The background is combined with a detailed account of how the local-ness of each speaker to St. John’s is calculated and quantified (c.f. Subsection 4.2.2.2).

5.6.3.1 Case Study One: 36HJYF, a Young Female

The young middle-class female, born in 1989, I discuss in what follows is most representative of innovative mainland Canadian speech while being the most local to St.
John’s, Newfoundland, among my young speakers at the same time. When I interviewed speaker 36HJYF in 2011, she was a fifth-year Business student at Memorial University of Newfoundland (Bachelor Commerce Cooperative) and came from a middle-class social background: Her father was a civil servant (senior officer) at Canada Revenue Agency, and her mother was a nurse. Both of her parents are Newfoundlanders: Her father was born in St. John’s, moved around the province with his parents in his adolescence (from Port-aux-Basque to Grand Falls and to Hearts Content), and finished high school in St. John’s. His father resided in St. John’s and Hearts Content. His ancestors were from Olney (Buckinghamshire) in England.

Speaker 36HJYF’s mother was also born in St. John’s, but never moved around the province. At the age of 17, she moved to west end St. John’s. Her father lived in downtown St. John’s, and his ancestors originally came from Ireland, Scotland and England. Speaker 36HJYF herself was born in St. John’s and never traveled or spent longer periods of time away from the city. She completed her primary and secondary education in St. John’s as well. Her local-ness index for generations native to St. John’s (LIgen) is three, and that for local schools attended (LIschool) is also three. Both scores are the highest possible ones in these two categories. Since she has never spent longer periods of time away from St. John’s, the local-ness index for being abroad (Llabroad) is zero. Thus, nothing has to be subtracted from LIgen and LIschool, so that she scores a total of six points in terms of local-ness (LItotal). She is consequently one of the most typical representatives of middle-class St. John’s English in my data set.

Figure 5.21 shows the individual vowel space of speaker 36HJYF, including vowels in the lexical sets DRESS, FACE, FLEECE, FOOT, GOOSE, KIT, LOT-THOUGHT, STRUT and TRAP. Other diphthongs and the allophones before /r/ are excluded from the scatterplot. The young female has a complete merger of the low back vowels (LOT and THOUGHT): the lowest and frontest THOUGHT token is the lexical item thought with a first formant value of 840 Hz and a second formant value of 1351 Hz; the highest and backest LOT token is the lexical item dog with formant values of 768 Hz and 1149 Hz, respectively. All of the other tokens are scattered within the field of dispersion created by these two tokens, except for the (possibly outlying) lexical item odd with THOUGHT-vowel formants of 722 Hz and 1090 Hz. 36HJYF’s mean formant values for merged LOT-THOUGHT are 814 Hz (sd 48 Hz) and 1211 Hz (sd 63 Hz).\(^{130}\)

Closely positioned to the low back vowels, STRUT seems to be dispersed in a fronter position than the former two vowels. Unlike Clarke, Elms and Youssef (1995: 212) argued in the first attestation of the Canadian Shift, this movement is not significant. In fact, STRUT and merged LOT-THOUGHT change position in speaker 36HJYF’s vowel space so that a slight fronting of STRUT is accompanied by a backing of LOT-THOUGHT. Although far from significant, this pattern holds also true when the mean vowel formants for those

\(^{130}\) Note that the vowel means are outlined in a comparative plot in the following subsection.
Figure 5.21: F1xF2 plot of the individual vowel measurements of 36HJYF, a middle-class female from St. John’s, Newfoundland, born in 1989 (21 years of age when interviewed; n = 148). The crosshairs indicate the center of the vowel space.
two vowel classes of all young versus old speakers are compared (cf. Section 5.1). Such
an interpretation allows for the subsequent retraction of TRAP.

Despite the merger, 36HJYF shows more of a trapezoidal than an inverted-triangle vowel space which is, according to Boberg (2010: 236), usually the case for Canadian speakers with a completed low-back merger, retracted TRAP and lowered DRESS and KIT vowels (but cf. Subsection 5.6.3.2 below). The trapezoidal shape would become more apparent in the scatterplot if GOOSE tokens before /l/ were included (cf. e.g. Labov, Ash and Boberg 2006b: 18), and if merged LOT-THOUGHT was dispersed higher towards a mid-back position (cf. Boberg 2010: 236). GOOSE tokens before /l/ are typically positioned in the upper-back corner of the vowel space, representing a non-fronted GOOSE realization, as they tend to be last affected by the centralization of the GOOSE vowels. For speakers with a very advanced Canadian Shift configuration of the respective vowels, TRAP tokens will be “in low-central to almost low-back position as the bottom corner of an inverted triangle” (Boberg 2010: 236).

Although such a configuration is not apparent, the three lax vowels that are considered to be part of the Canadian Shift all have formant values that fall within the thresholds outlined by Labov, Ash and Boberg (2006b: 219) and used as a means of comparison by Boberg (2010: 145). In addition, 36HJYF’s configuration of TRAP, DRESS and KIT is similar to that of Boberg’s Jewish respondent from Montreal (2010: 238), for whom “the Canadian Shift is much less advanced [...]”, but nevertheless present. All of 36HJYF’s TRAP tokens have second formant values smaller than 1825 Hz, some of them being as retracted as 1650 Hz. Moreover, all of her TRAP tokens are realized back of the center of her vowel space (average F1 = 881 Hz and F2 = 1724 Hz, respectively; sds 38 Hz and 68 Hz, respectively).

Her DRESS vowels are dispersed around 750 Hz, which is close to the low and backed position of TRAP, and yet both vowels are clearly separated. This separation may serve as an indication of the margin of security that has been maintained after the vowels have shifted. Her DRESS tokens reflect a low-front rather than mid-front position (mean F1 = 744 Hz and mean F2 = 1872 Hz; sds 32 Hz and 77 Hz, respectively), as they are dispersed far below the mid-line. Almost half of the tokens are also realized slightly back of center, supporting the diagonal movement of the DRESS vowel in apparent time that I have established in the quantitative analyses (whatever, section, bed’s, vet, checked, expect, veterinary, bet and effort).

Her KIT vowels are clearly separated from the dispersion of her DRESS and TRAP vowels in first formant space. While some of these tokens are still realized as high-front vowels with first formant frequencies between 430 and 450 Hz, the majority of her KIT tokens are positioned at first formant values of around 550 Hz (mean F1 = 545 Hz and mean

---

131 Ethnicity was one of the most important predictors in Boberg’s (2005) Montreal study. Jewish speakers were generally less advanced in terms of the Canadian Shift than their Irish and English counterparts.
5.6. The Canadian Shift in St. John’s

F2 = 2037 Hz; sds 46 Hz and 114 Hz, respectively), close to the mid-line. These tokens indicate that KIT is a mid-front vowel in the vowel space of speaker 36HJYF, with three tokens realized back of center (different, consider, pick). As for DRESS, the movement of the KIT vowel in apparent time is a diagonal one, as shown in the quantitative analyses above.

The fields of dispersion for her KIT and DRESS vowels are markedly separate from one another, indicating a relatively conservative realization of the former with regard to the Canadian Shift. Although DRESS and TRAP also do not overlap, the margin of security between KIT and DRESS is much more prominent than between DRESS and TRAP. However, in relation to FACE, most of her KIT tokens have shifted towards a mid-front position. The standard deviations for the three lax vowels are much greater in second formant space than in first formant space, which could be interpreted as the dominant movement of the vowels. As speaker 17LEMM’s vowel space will reveal in the subsection to follow, similar fields of dispersion can be found among his second formant values, although he does not display any sound shift in progress concerning these three front lax vowels. It is the position of the vowel dispersions relative to the center of the vowel space (and to other vowels) rather than the shape or size of the dispersions that indicates retracted TRAP realizations as well as lowered DRESS and KIT realizations with accompanying retraction.

While the very low and front position of the vowel in the single lexical item kit is an outlier, the raised and fronted position of the vowel in the lexical item keg lends some confirmation to Clarke’s (2004b: 370, 2010b: 77) observation that before voiced velar stops, the DRESS vowel may be tensed (and diphthongized) in English-settled areas of Newfoundland. In this case, the realization of the DRESS vowel in keg might thus reflect the English ancestry of 36HJYF’s family on her mother’s side. Alternatively, it might also reflect an effect of coarticulation between the vowel and the following consonant (cf. Roeder 2012: 483).

Other noteworthy vowel configurations are her GOOSE and FOOT vowels. Although numerically underrepresented in comparison to the lax vowels and THOUGHT discussed above, the dispersion of her GOOSE vowel seems to indicate a shifting pattern as well. For Boberg’s female case study born in 1981, he suggested that “this Newfoundlander share[d the fronting of /uw/] with her Canadian compatriots” (2010: 240). This innovative mainland Canadian change in the vowel configuration of the young speakers seems to be apparent for speaker 36HJYF as well, although a much more thorough analysis than interpreting a graph is needed before making such claims. The high-back vowel in the lexical items juice, Jews and shoes is realized with an on-glide in General American English (cf. e.g. Kenyon and Knott 1953), so that these items should be realized with a high-back vowel in a much fronter position than for instance the vowels in goose and food. Clarke (2006: 232-239) has shown that the glide-less American realization of the vowel in Duke is an innovative mainland Canadian feature that has spread to the linguistic
5. Analysis and Discussion

systems of the young speakers in the community of St. John’s already. This pattern is true for 36HJYF’s realization of *Duke* as well ([duk]), so that the vowel’s extremely high second formant value seems to reflect GOOSE-fronting or centralization ([duk]) instead of the realization of an on-glide. The innovative pattern is additionally reflected in the positions of the vowels of the four *goose* tokens between the vowels of *juice* and *shoes*, with two of them clearly realized front of center. The extremely back position of the vowel in *Jews* seems to reflect some kind of hypercorrection: The two lexical items *juice* and *Jews* were placed in immediate vicinity in the word list, constituting a minimal pair with regard to the following consonant. Speaker 36HJYF decided to pronounce the lexical item *Jews* with two altered consonants, the preceding and the following one, corresponding to an orthographic realization of *choose*.

In a similarly superficial interpretation, the centralization of FOOT also seems to be apparent in her vowel configuration. According to Boberg, an ANOVA of age and second formant values of FOOT showed that the realization of this vowel is significantly fronter for young speakers from Montreal than for old ones ($p = 0.022$). This effect was not corroborated in an ANCOVA of age with sex, education and ethnicity as covariates, indicating that the significance of age is confounded with one of the latter three predictors (2005: 144). In his (2010) book, he combined the data sets from Halifax, Nova Scotia, and Vancouver, British Columbia, collected by Sadlier-Brown and Tamminga (2008), with his Montreal data, in which the significant effect of age on FOOT centralization could be replicated ($p = 0.009$).

He maintains that the advancement of FOOT is a phonetic characteristic of Standard (mainland) Canadian English (Boberg 2010: 152), but also emphasizes that “[i]ts structural connection with the Canadian Shift is not clear; it may, in fact, be a purely coincidental development” (Boberg 2005: 145). This characteristic of Canadian English has also been stated by Clarke, Elms and Youssef (1995) and Hung, Davison and Chambers (1993), but the former study did not assign this movement of FOOT a role in the Canadian Shift (1995: 213). However, it is part of (Standard mainland) Canadian English and appears to have entered the linguistic system of middle-class St. John’s English – as has the Canadian Shift, which is clearly evident from speaker 36HJYF’s vowel configuration.

5.6.3.2 Case Study Two: 17LEMM, a Middle-aged Male

The middle-class male was born in 1965 in Stephenville on the west coast of Newfoundland. His family moved to a suburb of St. John’s on the east coast of Newfoundland when he was seven months old and to west end St. John’s when he was six years old. By the time I interviewed him in 2011, he was 46 years of age. He spent one year of his childhood in Labrador City when he was in sixth grade and completed his junior and senior high school in St. John’s. After graduation from high school, he went to Memorial University of Newfoundland for three semesters, but did not finish his Bachelor’s degree in anthropology.
After several short-term jobs as a bartender and a legal collector in St. John’s, he worked as a Flight Service Specialist in Ontario for a year and a half from 1989 to 1990. After returning to St. John’s, he started working as a customer technical assistant and has since been promoted to Customer Care Management, where he had spent the past ten years before being interviewed. His spouse is a part-time interviewer for a market research company in St. John’s.

His father was an Acadian born in Shediac, New Brunswick, and worked as a Marine Electronics Technician. At the age of 21, he moved to Newfoundland. His father resided in Springville, Nova Scotia, and moved to Shediac, New Brunswick, when he married. Their ancestors originally came from Normandy, France. 17LEMM’s mother was born in La Manche, Placentia Bay, and was a homemaker. Her father resided in Salmon Coast, Conception Bay North, before she was born. Their ancestors originally came from England (West Country) and Ireland.

17LEMM’s local-ness index for generation native to St. John’s (LIgen) is one: he was born in Newfoundland, but not St. John’s, so that he scores 0.5 points in this category; his father was born neither in St. John’s nor Newfoundland, so that speaker 17LEMM scores zero points for his father’s birthplace and another 0.5 points for his mother being born in Newfoundland, but not in St. John’s. His local-ness index for schools local to St. John’s is 2.5: he scores one point for having completed his elementary school in St. John’s, another 0.5 points for having completed his high school in St. John’s but spending grade six outside of Newfoundland, and another point for having studied at Memorial University of Newfoundland in St. John’s. From those two local-ness index scores, one point is subtracted because speaker 17LEMM lived in Ontario for more than a year when he was younger than 30. His total local-ness index is thus 2.5, representing a typical middle-class speaker from St. John’s, Newfoundland, in his mid-forties. I consider him to be a typical middle-class speaker because he has spent most of his life in St. John’s before graduating from high school. Unlike typical working-class speakers, he has been quite mobile since, spending extended periods of time in the Maritimes and mainland Canada. As speaker 20SCOF, a 59-year-old middle-class female, puts it: “All Newfoundlanders spend time in Calgary [Alberta, Canada]. It’s kind of like time in purgatory – and then you’re allowed home again” (Hofmann 2011: 10:10-10:17 mins). She has spent a year and a half in Calgary together with her husband in order to brush up their résumés. A typical middle-class Newfoundlander does not work in career professions for longer periods of time, but has several consecutive jobs throughout their professional life: “Newfoundlanders with BAs don’t have occupations, they have a series of jobs” (Hofmann 2011: 12:49-12:52 mins).

Before the time a speaker’s vernacular becomes stable at the age of approximately 20 (cf. e.g. Cukor-Avila and Bailey 2013: 246), speaker 17LEMM has spent periods of seven months (before he was able to talk) and a school year outside of St. John’s and Newfoundland, and between the ages of 24 and 25, he has spent another year and a half
outside the province. Put differently, before he turned 25, he had spent 22 years in St. John’s. Despite his relative mobility later in life, speaker 17LEMM spent almost all of his sociolinguistically formative years in St. John’s so that his variety is nevertheless representative of middle-class St. John’s English.

Figure 5.22 shows his vowel configuration. A complete merger of the low back vowels in the LOT and THOUGHT lexical sets is visible, although in a much fronter position than merged LOT-THOUGHT of the young female discussed above. Although the individual lexical items uttered by 17LEMM are difficult to read in the graph, the symbols indicating vowel positions show that, for instance, the vowels in the lexical items *thought* and *talking* (THOUGHT) have second formant values of 1500 Hz. They are positioned as front as the vowels in the lexical items *got*, *gotta* and *Comma* (LOT). Likewise, the LOT vowels in the lexical items *hobby* and *photography* are realized as far back as 1200 Hz, a position usually attributed to THOUGHT vowels. In between those two second formant values, all of the LOT and THOUGHT tokens are dispersed, suggesting a low-back merger in a front position when compared to 36HJYF (cf. Subsection 5.6.3.1 above). 17LEMM’s merged LOT-THOUGHT vowels overlap partially with his STRUT vowels, while 36HJYF’s LOT-THOUGHT vowels are dispersed in a much backer position than her STRUT vowels.

The same relationship between the two vowels is reflected in their second formant means: 17LEMM’s mean F2 is 1320 Hz (sd 177 Hz) and 36HJYF’s mean F2 is 1211 Hz (sd 63 Hz; cf. the figure below). However, the relative positions of merged LOT-THOUGHT between age groups in second formant space were not significantly different (cf. Subsection 5.2.4).

A similar pattern was offered by Labov, Ash and Boberg (2006b: 221) for their case study from St. John’s, Newfoundland, a 35-year-old male interviewed in 1997. He may thus have been born between 1961 and 1963, similarly to 17LEMM (cf. Subsection 5.2.1). The position of merged LOT-THOUGHT for this speaker is at second formant values as front as 1500 Hz, and marks the lower end of an inverted-triangle-shaped vowel space. Such a vowel space is attributed to those speakers who show a very advanced Canadian Shift in Boberg’s (2010: 236) data from Ontario, with TRAP as the bottom corner of the vowel space. According to Boberg, the advancement of the Canadian Shift correlates with a change in the shape of the vowel space from a trapezoidal to an inverted-triangle shape (2010: 236). However, neither the inverted-triangle shape of the case study’s vowel space Labov, Ash and Boberg present (2006b: 221) nor the trapezoidal vowel space of speaker 17LEMM in my data show a Canadian-Shift-like behavior of the front lax vowels.

In the latter case, TRAP is markedly separate from the merged LOT-THOUGHT vowel in second formant space, which could be interpreted as a margin of security. With such an interpretation, a margin of security does not seem to be necessary for the lax vowels TRAP, DRESS and KIT, which seems to support their relative stability: In such a conservative Newfoundland English realization of the front lax vowels the disambiguation between these could probably be achieved on, for instance, a morphosyntactic level such as word-class membership or context. Together with FACE, they show one great field of dispersion
Figure 5.22: F1xF2 plot of the individual vowel measurements of 17LEMM, a middle-class male from St. John’s, Newfoundland, born in 1965 (46 years of age when interviewed; n = 244).
in the low to high front vowel space. The mean formant values for TRAP are 736 Hz for formant one and 1965 Hz for formant two (sds 40 Hz and 67 Hz, respectively), and all the individual tokens are realized front of center. In addition, when compared to the vowel configuration of speaker 36HJYF, 17LEMM’s TRAP vowel is on average realized fronter (and slightly higher) in relation to the mid-line than 36HJYF’s DRESS vowel (mean F1 = 744 Hz and mean F2 = 1872 Hz; cf. Subsection 5.6.3.1; also cf. Figure 5.23), providing strong evidence for a sound change in progress in the speech community of St. John’s, although 36HJYF is less advanced in her participation in the Canadian Shift than speakers from Montreal or Toronto.

17LEMM’s mean formant values for DRESS are 607 Hz for formant one and 2048 Hz for formant two (sds 47 Hz and 123 Hz, respectively), and the mid-line divides the dispersion of his individual DRESS tokens roughly in half. In comparison to 36HJYF, this position in the vowel space is rather occupied by her KIT vowels than her DRESS vowel tokens. Her mean is much closer to the mid-line and much more distant from the mean FLEECE vowel realization than 17LEMM’s. In terms of formant mean values, 36HJYF’s KIT vowel means is yet 62 Hz higher than 17LEMM’s DRESS vowel (mean F1 = 545 Hz and mean F2 = 2037 Hz), providing some qualitative support for the relatively conservative behavior of KIT with regard to the Canadian Shift in St. John’s (cf. Subsection 5.6.2). Remarkably, 36HJYF’s KIT vowel is (already) as retracted as 17LEMM’s DRESS vowel. 17LEMM’s KIT vowel may thus provide the initial position of 36HJYF’s KIT vowel: Formant one has a mean frequency of 478 Hz, and formant two has one of 2126 Hz, respectively (sds 30 Hz and 153 Hz, respectively). 17LEMM’s KIT vowels are much higher positioned in the vowel space than his FACE tokens and at the same time close in dispersion to his FLEECE vowels.

In terms of GOOSE, 17LEMM does not front the vowel, as the lexical items realized in a front position juice, Duke, Jews and shoes are prescriptively realized with an on-glide (cf. Kenyon and Knott 1953). Those lexical items without such an on-glide (goose, food, pursue and moves) are realized in the upper back corner of his trapezoidal vowel space. The vowel in the lexical item zoo is prescriptively also not realized with an on-glide, but voicing of the preceding alveolar fricative may have caused the relative fronting of the vowel (cf. Figure 5.22). The innovative realization of GOOSE that the young female shows cannot be attested for 17LEMM, but his realization of FOOT is centralizing and thus very similar to that of 36HJYF. The similarity seems to lend some qualitative support to Boberg’s (2005: 145) interpretation of FOOT-centralization being structurally most likely not connected to the Canadian Shift but rather a coincidental development.

5.6.3.3 Case Study Three: 21FJOM, an Old Male

The old middle-class male was born in 1953 in St. John’s and is one of the youngest speakers within the old age group. By the time I interviewed him, he was 58 years old. He lived in west end St. John’s until he turned 21 and joined the armed forces in
5.6. The Canadian Shift in St. John’s

(a) 36HJYF’s individual vowel tokens with means (born in 1989; n = 148).

(b) 17LEMM’s individual vowel tokens with means (born in 1965; n = 244).

Figure 5.23: F1xF2 plots of the individual vowel measurements of 36HJYF and 17LEMM compared.
British Columbia, Canada, for three years. After his service, he returned to west end St. John’s. In 1978, at the age of 25, he married and moved to downtown St. John’s. A year later, he went to live and work in Calgary, Alberta, for a year and a half. He completed his primary and secondary education in St. John’s and went to Memorial University of Newfoundland to earn a Bachelor’s degree of Science in Maths and Physics. He was also awarded a diploma in Technical Engineering and Diplomacy in 1979. He works as a research assistant in the department of Physical Oceanography at Memorial University of Newfoundland.

21FJOM’s father was also born in St. John’s and worked as a provincial government employee. His father lived in Notre Dame Bay as a child and moved to St. John’s in his youth. 21FJOM’s mother was also born in St. John’s and was a homemaker throughout her life. Her father lived in a suburb of St. John’s. Both ancestors came originally from England and Ireland.

21FJOM’s local-ness index for generation native to St. John’s (LIgen) is three: both of his parents and he himself were born in St. John’s. His local-ness index for attending local schools (LIschool) is also three as he completed all of his education in St. John’s, Newfoundland. From these two local-ness indices, two points are subtracted for the time he spent abroad (LIabroad): he spent three years in British Columbia, and a year and a half in Calgary, Alberta. His total local-ness index is thus four, representing a middle-class speaker that has spent more consecutive time in St. John’s, Newfoundland, than a typical middle-class speaker.

His vowel configuration is shown in Figure 5.24. Very generally, the vowel space is somewhat similar to that of speaker 17LEMM discussed above, while at the same time it corresponds more closely to the descriptions offered by Clarke (2010a: 26-31) for (typical) Newfoundland English. 21FJOM’s vowel space has a trapezoidal shape, and the three lax vowels TRAP, DRESS and KIT are dispersed quite closely to one another. In terms of the low back vowels, the scatterplot suggests a completed merger of LOT and THOUGHT: While some LOT tokens are realized as front as 1700 Hz (e.g. the lexical items job, got, don) others are realized as back as 1200 Hz (e.g. oven, officers). Within these two extremes, all of the THOUGHT vowels are dispersed, some of them as front as 1500 Hz (e.g. talk, talking, odd). The average first formant value of his merged LOT-THOUGHT is 784 Hz and the average second formant value is 1422 Hz (sds 81 Hz and 145 Hz, respectively).

This position of merged LOT-THOUGHT is roughly 100 Hz fronter than that of the middle-aged male, 17LEMM, which is in turn roughly 100 Hz fronter than that of the young female, 36HJYF, discussed above. Although this comparison suggests the significant retraction of the low back vowel in apparent time, vacating a space in the central mid-low position of the vowel space, the retraction is not significant when all speakers are statistically assessed (cf. Subsection 5.2.4). The position of some of 21FJOM’s LOT-THOUGHT vowels with second formant frequencies of 1700 Hz is as front as the position of some of 36HJYF’s TRAP vowels with second formant frequencies of 1650 Hz, which may
Figure 5.24: F1xF2 plot of the individual vowel measurements of 21FJOM, a middle-class male from St. John’s, Newfoundland, born in 1953 (58 years of age when interviewed; n = 198).
cause misunderstandings, “as in the case of a St. John’s resident whose first name, John, was repeatedly interpreted as Jan when he travelled to Toronto” (Clarke 2010a: 31). As visualized in Figure 5.25, the relative positions of 21FJOM’s merged LOT-THOUGHT and STRUT are similar to those of 17LEMM.

Although three lexical items of the LOT lexical set, job, got and don, are realized with second formant values of more than 1700 Hz, there is a visible margin of security between the merged LOT-THOUGHT and TRAP vowels. Five of his TRAP tokens (pat, bath, jacket and back [2x]) are realized back of center, which seems to be caused by the extreme front realizations of his FLEECE vowels. In stark contrast, the FLEECE vowel dispersions between the two younger speakers outlined in Figure 5.23 are quite similar in position. This interpretation is also supported by the dispersions of 21FJOM’s TRAP, DRESS, KIT and FACE vowels, when compared to those of 17LEMM, and 21FJOM’s mean formant values. Mean F1 of TRAP is 709 Hz, and mean F2 of the same vowel is 2084 Hz (sds 76 Hz and 89 Hz, respectively). Like merged LOT-THOUGHT, 21FJOM’s TRAP vowel is 100 Hz fronter than that of middle-aged 17LEMM, indicating a similar stability of the two low vowels for both speakers (cf. Figure 5.25). On average, TRAP is also slightly higher when compared to 17LEMM’s (mean F1 736 Hz; cf. Subsection 5.6.3.2). The fronted and raised realization of TRAP vowels is “an obvious feature of many of the speakers [from Newfoundland], whether they represent conservative [Newfoundland English of British origin...], conservative [Newfoundland English of Irish origin...] or more innovative speech [...]” (Clarke 2010a: 29). Some of 21FJOM’s TRAP vowels are realized as raised and fronted as “an [e]-like vowel” (Clarke 2010a: 30): For instance, the lexical item imagine has formant frequencies of 629 Hz and 2229 Hz, respectively, and the lexical item afterwards has formant values of 596 Hz and 2162 Hz, respectively. According to Clarke (2010a: 30), fronting and raising of ash occurs in all linguistic contexts in Newfoundland English, but it is “perhaps most apparent before a nasal consonant”.

Similar to 17LEMM’s vowel space, the dispersions of the DRESS and KIT vowels of 21FJOM are not clearly separable and overlap with that of FACE. 21FJOM’s DRESS vowel is on average realized just above the mid-line. It has a mean first formant value of 560 Hz and a mean second formant value of 2270 Hz (sds 48 Hz and 192 Hz, respectively). It is thus roughly 50 Hz higher and more than 200 Hz fronter than the DRESS vowel of middle-aged 17LEMM (cf. Subsection 5.6.3.2), which may be understood as “[...] some degree of raising in [Standard Newfoundland English...]” (Clarke 2010a: 26), “or tensing” (Clarke 2010a: 28). His fronted and raised realization of the vowel in the lexical item keg might be indicative of his ancestors’ English heritage (cf. Clarke 2004b: 370, 2010b: 77).

21FJOM’s realization of KIT seems to be fairly similar to that of 17LEMM, with FLEECE being realized in an extremely high front position and FACE as well as DRESS being raised and fronted. 21FJOM’s mean formant frequencies for KIT are 476 Hz for formant one and 2408 Hz for formant two (sds 32 Hz and 209 Hz, respectively). The mean first formant value is almost identical to that of 17LEMM (478 Hz; cf. Subsection
5.6. The Canadian Shift in St. John’s

(a) 17LEMM’s individual vowel tokens with means (born in 1965; n = 244).

(b) 21FJOM’s individual vowel tokens with means (born in 1953; n = 198).

Figure 5.25: F1xF2 plots of the individual vowel measurements of 17LEMM and 21FJOM compared.
5. Analysis and Discussion

5.6.3.2), but 21FJOM’s second formant value is more than 250 Hz greater than that of 17LEMM (2126 Hz). While this could be interpreted as, for instance, an indication that KIT retraction is the dominant movement for this vowel in the Canadian Shift, such an interpretation is problematic for several reasons: 1) 17LEMM’s vowel dispersions do not show the positions typically attributed to the vowels in the Canadian Shift (cf. Subsection 5.6.3.2); 2) the quantitative analysis of the KIT vowels of all speakers does not suggest that middle-aged males favor a (diagonal) shift (cf. Subsection 5.5.2); and 3) if the difference in the second formant space of KIT between speakers 17LEMM and 21FJOM is to be attributed to the Canadian Shift, then “retraction” of FLEECE has to be attributed to it as well: 17LEMM’s formant values for FLEECE are 402 Hz and 2611 Hz, respectively (sds 21 Hz and 223 Hz, respectively); 21FJOM’s formant values for FLEECE are 382 Hz and 3111 Hz (sds 26 Hz and 126 Hz, respectively).

If the difference of more than 250 Hz in the second formants of KIT between the two speakers is meaningful in the context of the Canadian Shift, then it seems implausible to disregard the parallel difference of 500 Hz in the second formant values of FLEECE between the two speakers. However, since FLEECE is generally not regarded to be part of the Canadian Shift (cf. Section 2.3), I consider this to be an atypical pattern for a speaker from St. John’s, Newfoundland, with regard to those two vowels, due to speaker idiosyncrasies. In fact, 21FJOM’s vowel space resembles that of old females more closely than that of old males with regard to the high front vowels, which is most likely due to his relatively small vocal tract.

In terms of the GOOSE and FOOT lexical sets, 21FJOM does not show a pattern of centralization as clearly as speaker 17LEMM does. While the lexical item Duke shows a vowel in a centralized position, it could well be an outlier (cf. e.g. shoes and new in Figure 5.24). The dispersion of the two vowels rather suggests that the FOOT vowels are raised and tensed towards GOOSE, while the latter generally maintains a back position. “[T]his trend is associated more with Irish-settled areas of the province than with southwest English-settled regions” (Clarke 2010a: 31). In the latter regions, FOOT raising and tensing occurs less frequently and is most likely phonetically and lexically conditioned, as it usually occurs before voiced alveolar /d/ (and voiceless velar /k/; zero occurrences in this analysis; cf. Clarke 2010a: 31). According to Clarke, such a realization of FOOT may be a historical retention of the long vowel that some of the words containing such a vowel-consonant cluster had in Middle-English (2010a: 31).

In summary, the scatterplots and the discussion of the case studies above provide qualitative support for the quantitative analyses outlined in Sections 5.3, 5.4 and 5.5. In comparison to the latter two case studies, the first clearly shows that the young female, 36HJYF, is participating in the innovative vowel changes referred to as the Canadian Shift, while the middle-aged and old males, 17LEMM and 21FJOM, show no indication of any such participation.
I divided the conclusion into three parts: In Section 6.1, I will first summarize the main results of the previous chapter and, second, attempt to identify the reasons for why young urban Newfoundlanders participate in the Canadian Shift. These are, of course, tightly interwoven with the recent social and economic changes in the province as well as the linguistic identity of the young St. John’s residents. In Section 6.2, I present and discuss the limitations of this study in a bottom-up fashion, starting with the point of vowel measurement as the smallest unit and concluding with the theoretical and methodological framework as the largest unit of this study. Some of these limitations constitute possibly fruitful areas of future investigation, which I will summarize in Section 6.3.

6.1 Socioeconomic Change and Sociolinguistic Identity

The triangulation of my results has produced quite consistent patterns of similarities across all methods. Not unexpectedly, some methods are less accurate or reliable, such as using Euclidean Distances as a metric to determine a shift of vowels in apparent time, but in combination with other methods such as correlation coefficients, the reliability and the validity of the results is generally increased.

As outlined in Subsection 5.6.2, the innovative, mainland Canadian English (CE) vowel shift pattern that is commonly referred to as the Canadian Shift is present in the speech community of St. John’s, Newfoundland. The investigation, description and explanation of this innovation in the historically isolated community was the main goal of this dissertation (cf. Hypothesis 1 in Section 3.4). The results for the lax vowels TRAP, DRESS and KIT even indicate a stage-like pattern as it has been suggested by Clarke, Elms and Youssef (1995: 212) for Ontario and as I have hypothesized for St. John’s, Newfoundland (cf. Hypotheses 4, 5 and 6): with the low-back merger in place (cf.
Hypothesis 2), TRAP retracts, and consequently DRESS and KIT lower. Unlike Clarke, Elms and Youssef (1995: 212) suggested, STRUT is stable in apparent time (cf. Hypothesis 3), which is in line with the results of most other studies on the Canadian Shift (cf. e.g. Labov, Ash and Boberg 2006b).

In terms of research questions b) and c), the findings do not support the hypotheses I have formulated in Section 3.4 as clearly. Hypothesis 7 states that social variables exert a stronger influence on the innovations than the linguistic variables. In terms of the latter, their influence is variably significant and generally affects the innovative variants in the fashion suggested by Clarke, Elms and Youssef (1995) for Ontario and to some degree by D’Arcy (2005) for St. John’s. However, my combination of manner and place of articulation was necessary due to some empty cells and thus makes a direct comparison impossible. In terms of the former, their stronger influence only holds true for age, as suggested by Clarke (1991: 112), when all the front lax vowels are considered, regardless of the method I use: Young speakers always lead in the adoption of the Canadian Shift (cf. Hypothesis 8). For gender, the results are rather mixed: While females clearly lead the change with regard to TRAP retraction (i.e. middle-aged females have significantly higher application rates than young males), as is to be expected in a regular sound change, gender is not significant in the lowering of DRESS. The tendency of the latter is, however, the same as for the former: young females lower DRESS at higher rates than males. The lowering of KIT shows the exact same pattern, but only in the linear regression: gender is not significant, but young females lead. In monosyllabic lexical items only, the female lead is significant again – at least when tested with logistic regression. This may, however, also be an artifact of categorizing the variables age and formant one or of the use of sum contrasts. Bearing these gender-related caveats in mind, Hypothesis 8 holds true.

Style is the variable with the least interpretable pattern. Having stated that, Hypothesis 9 seems to hold true as well: very generally, TRAP is more retracted and DRESS is more lowered in the vernacular of the speakers, whereas KIT is lowered the most in the formal styles. The fact that Clarke (1991: 116) found TRAP lowering and retraction to be highest in the formal styles leads me to conclude that the KIT vowel will show the same pattern as TRAP and DRESS already display in a generation from now. This interpretation receives support from the role of style across the vowels and the multiple regressions (linear and logistic): TRAP is most often retracted in spontaneous speech according to both regression analyses; KIT is most often lowered in formal style according to both regression models; DRESS is most often lowered in spontaneous speech in the linear regression, but in formal style in the logistic one (possibly due to categorization of originally continuous variables). Putting Clarke’s findings and mine together seems to suggest that the shift has started as a change from above in the 1980s and has since begun to change gradually into a regular sound change from below. However, the pattern of style across the different regression analyses is not a reliable indicator of such an assumption. Only the fact that number of following syllables constrains the lowering of KIT in a highly significant
fashion provides reliable evidence across both the linear and logistic regression that, for this vowel, there seems to be an intermediate combination of characteristics from lexical diffusion change and from regular sound change (cf. Labov 1994: 542): Monosyllabic lexical items are affected first and within them, similar social and linguistic variables as for DRESS constrain the lowering of KIT within monosyllabic lexical items (cf. Bybee 2002; Oliveira 1991). The lowering of DRESS is favored in interview style in the linear regression, i.e. it is below the level of social awareness. Such data are not available or not reported for the Canadian mainland so that it is unclear whether the shift behaves in a similar manner there. The data that are available for the mainland suggest characteristics which are all shared by my young speakers from St. John’s (cf. Subsection 5.6.2).

The pattern of a combination of lexical diffusion change and regular sound change also seems to make sense in light of the population make-up of present-day St. John’s, attitudes, linguistic identity and other recent changes in the speech community outlined in Chapter 3. Newfoundland’s loss of its isolation has started in the decade of World War II and its joining of the Canadian Confederation in 1949. Through the American, British and Canadian military personnel being stationed on the island and in St. John’s, Newfoundlanders increasingly came into regular and long-term contact with North Americans and the British (cf. Subsection 3.1). If not pre-existing, most of the negative attitudes towards conservative Newfoundland varieties may have originated in that time. Over time both parties have learned to live with the situation, and several linguistic innovations had entered the local speech of St. John’s residents by the early 1980s (Clarke 1991), including the retraction of TRAP. The negative attitudes have, however, been internalized by the locals and even nurtured by comedy groups and other artists from Newfoundland (e.g. recently, the Gazeebow Unit; cf. Subsection 3.3.1).

After the 1950s, the government-run resettlement program, the cod moratorium in 1992 and the development of the oil industry have caused many rural Newfoundlanders from formerly isolated outport communities (baymen) to migrate into St. John’s and other urban areas, overtly attributing prestige to salient non-local linguistic forms and seeking education as well as employment (cf. 3.3.1). The historical urban/rural divide in Newfoundland, in which the townies have similar but weaker views about the baymen as North Americans about Newfoundlanders, has thus become reinforced, leading to an emphasis of the social distinctiveness of townies from baymen via, among others, linguistic means. I consider these community-internal forces as one cause for the increase of mainland CE innovations, such as GOOSE-fronting, Canadian Raising, quotative be like and most importantly the Canadian Shift.

Very recently, a refocusing on the pride of having a Newfoundland accent has been publicly debated in the local newspapers of St. John’s, which does, however, not result in an increased use of traditional dialect features (cf. Martha’s Vineyard; Labov 1963), as shown in the analysis of this thesis, since these are attributed to baymen speech, as well as urban working-class sket speech (cf. Subsection 3.3.1). For more than a generation,
mainland Canadians have migrated to and remained in St. John’s due to the development of the oil industry. Socially, these migrants are perceptually much closer to St. John’s residents-by-birth, so that the former have become part of the social networks of the latter by now and consequently also of the linguistic identity of the local middle class. In fact, in the friends section of my interviews, no young and very few of the middle-aged respondents have told me that their friends are exclusively Newfoundlanders by birth. The majority of them even spent more time with Canadians formerly residents somewhere on the mainland than with locals in an average day in their lives. Although my respondents were aware that these were mainlanders, they still considered them to be close friends and community insiders, be it for hobbies, at work or in their leisure time. This seems to be further supported by the fact that local-ness (LI_total) is never a significant factor in any of the statistical models. The tendency is, however, pointing in a not unexpected direction: those with more contact to Canadians on the mainland prefer the innovation slightly more. Modern St. John’s is perceptually characterized by the presence of mainland Canadians as members of the speech community. This is also indicative when considering the fact that the formerly supra-local innovative TRAP variant is predominantly used in the vernacular (spontaneous speech) of my respondents, i.e. it is not a salient variant. I consider these predominantly community-external forces acting on the locals to converge with formerly innovative mainland Canadian standard variants as a second cause. I interpret the forces as predominantly but not completely external, or a hybrid of community-internal and community-external forces, due to the recent convergence of mainland Canadian residents of St. John’s and residents-by-birth in terms of linguistic and cultural norms.

In terms of the remaining two innovative variants, lowered DRESS and KIT, I hypothesize that they will be used at higher application rates in the vernacular in the course of one more generation; for DRESS, this is already the case in the linear regression. Although there is no evidence for this assumption, in combination with Clarke’s results my data seem to point in this direction. While innovative retracted TRAP was a formal response in her data collected a generation ago (1981-1982; Clarke 1991: 111), mainland Canadians were then not part of the speech community of St. John’s in the way they are today. In addition, present-day DRESS seems to be on the verge of entering the vernacular, given that a categorization of age into three groups and of formant one into two groups is primarily responsible for the different style result in the logistic regression. Lowered KIT is still a salient minority variant in the speech community when it occurs in multisyllabic lexical items. The innovative variant seems to be constrained, however, by the same linguistic and social variables as DRESS when analyzed in monosyllabic lexical items only, which I understand as indicative of following the trajectory of DRESS in a generation or more from now.

For all the innovative variants and particularly for the latter two, the role of the linguistic marketplace cannot be underestimated, given the economic situation in St.
John’s. As mentioned above, the development of the oil industry and its very recent economic results that have changed the status of Newfoundland from a ‘have not’ to a ‘have’ province are by and large strongly enforcing the recent migration patterns from rural to urban Newfoundland and onto the island, which, of course, increase the competitive economic climate for young middle-class residents of St. John’s. I consider this increasing competition for jobs as the final cause, also a community-external one, for the adoption of innovative variants such as TRAP retraction and DRESS as well as KIT lowering. The force being community external is rooted in the fact that many young Newfoundlanders have to leave the province for their profession and yet decide to commute instead of moving to the mainland permanently. Such commuters usually work for several consecutive weeks off the island and then return, again for several weeks. The contacts with the mainlanders on the Canadian mainland thus result in the adoption of innovations such as the Canadian Shift, which are consequently brought back home to their families. All three forces change the social make-up of Newfoundland’s capital and other urban areas – and consequently its linguistic identity and vernacularizing choices – into a relatively modern and open metropolitan area not unlike much larger cities on the Canadian mainland.

This conformity of the contemporary linguistic behavior of the middle-class in the province’s capital to the national pattern is surprising, because in-migration of the United Empire Loyalists after the American War of Independence, who settled Ontario and spread westwards, never was a factor in the settlement history of Newfoundland. At the same time, the capital of Nova Scotia, Halifax, participates in the Canadian Shift as well. The national picture that emerges is thus similar to the one that was outlined in the 1990s (e.g. Hung, Davison and Chambers 1993): the urban middle classes of Canada’s cities from coast to coast speak rather homogeneously.

Based on most of the literature, it could be argued that St. John’s also shows many of the innovative CE features beyond the Canadian Shift such as GOOSE-fronting, on-glide deletion (yod-dropping) in lexical items such as news, centralization of FOOT, Canadian Raising, raising and fronting of TRAP before nasals and /g/, unretracted START, use of vocal fry, use of high rise terminals and quotative be like. In terms of GOOSE-fronting, the low-back merger, the Canadian Shift and Canadian Raising, St. John’s seems to fit the characteristics of the Canadian dialect region outlined by Labov, Ash and Boberg (2006b: 146). This may suggest that the isoglosses they defined are extendable to the east of Montreal to include St. John’s.

6.2 Limitations

This dissertation is limited in several respects. I will outline the limitations in a bottom-up fashion in that I start at the level of vowel measurement and then move towards larger units such as individual linguistic segments, sentence level, normalization, topical sections and structure of the interview and sampling of the speakers, in order to finally arrive at
the level of the theoretical and methodological framework: variationist sociolinguistics (VS). However, as should already be apparent from this list, these limitations are not easily teased apart.

The goal of my study is to fill a gap in research regarding the Canadian Shift in St. John’s, Newfoundland, and consequently the Canadian Shift in general. Its absence was attested most recently by the studies of Boberg (2008b, 2010) and Labov, Ash and Boberg (2006b), based on six and two speakers from Newfoundland, respectively. Since my study thus aims at contradicting their findings, I decided to adapt my methodology as closely as possible to theirs, which specifically concerns the choice of vowel measurement points and normalization technique. I begin with the former, as it is at the lowest level in my structure of outlining the limitations. Both studies manually measured at the maximum value of formant one (or minimum F2) during non-transitional stages, i.e. the target, of the vowel, based on a visual inspection of the spectrogram. Instead of measuring the vowels in my data set similarly, I automatically measured them at 33% of the vowels’ duration, based on Evanini’s (2009) quantification of Labov, Ash and Boberg’s (2006b) manual measurements in the Atlas of North American English (ANAE). As Evanini (2009: 65) shows, the automatically-derived formant readings differ by no more than 13% from the manually-derived ones. The comparability of my findings to the ones of Boberg and Labov, Ash and Boberg is thus slightly reduced, but not substantially. A selection of a traditional mid-point measurement (50% of the vowel’s duration) would have been likely to yield different findings. In addition, recent studies in acoustic phonetics measure vowels at more than one temporal location, because they consider the information about the respective vowel targets in one F1/F2 pair as insufficient. While this may or may not be the case, it is unclear how to clean automatically-derived vowel measurements from several temporal locations in the vowel and how to use the additional information in more than one F1/F2 pair in analytical statistics. In other words, is it socially meaningful, for instance, to have a significant effect for age at the 20% and 30% measurement points but one for gender at the 40% measurement point? Or, if modeled with repeated measures, what constitutes the intervention (i.e. the hypothesis) between the measures? In medical studies, this is usually the non-/administration of a treatment or different treatments. In biology, this may be the development of plants over time, measured at different temporal locations after each has received a certain fertilizer. Due to the focus on comparability outlined above, I decided to measure the vowels at a single temporal location and consequently find no answers to these questions, leaving them as such. Concerning the choice of normalization technique, I consider the comparability unaffected, as I used Labov, Ash and Boberg’s procedure via NORM (Thomas and Kendall 2007), based on the Telsur G value for 345 and more speakers (normalization will be outlined below).

In terms of linguistic segments, my study is limited 1) in the total number of vowels that were manually segmented, as well as 2) in relation to the linguistic environments the vowels occur in and 3) in relation to the tense vowels, phonological diphthongs and
vowels before /r/ that I excluded from the analysis although they are also part of the vowel inventory of St. John’s English. Very recently, online tools such as FAVE (Rosenfelder et al. 2011) and MAUS (Schiel n.y.) have become available that automatically segment and extract formant readings from transcribed interview audio files, which saves much time and resources on the one hand, and allows to extract substantially more vowel tokens than manual segmentation on the other (e.g. six to seven times as many tokens as in the present study). These tools thus allow for a more thorough extraction of the vowel tokens for the analysis and help balance the issues I raised in 2) and 3). That is, if the stressed vowels of all interviews were extracted and collected in a database, I could have randomly sampled an equal amount of vowel tokens per preceding and following linguistic environment in each of the three styles included, word list, reading passage and interview. In addition, I would have been able to find more stylistic differentiation in the spontaneous speech section via data exploration methods. In that way, the total number of vowels would be much less limited by the temporal, financial and resource constraints that are accompanying studies like the present one, than by the smallest rate of occurrence of a vowel in a particular context. For instance, a stressed vowel in a $bVb$ (voiced bilabial plosive, vowel, voiced bilabial plosive) or $bVd$ environment may be much more common in connected speech than one in a $vVd$ or $vVv$ environment. The resulting database would be much more balanced for statistical evaluation, e.g. exactly ten vowels per environment. Such attempts have usually only been made when a particular statistical method or normalization technique was evaluated or described (e.g. Schützler 2011). I decided against the use of such online tools for two reasons: First, I wanted to gain experience in reading and interpreting spectrograms, segmenting vowels and extracting formants; and second, the resulting segmented data set requires thorough examination and post-editing of the segmentation boundaries, e.g. for consistency, which would have taken the same amount of time or more than manual segmentation and automatic formant extraction (particularly since I lacked the experience of working with such data at the outset of this thesis). In order to control for the imbalance of tokens mentioned above, I worked with a vowel space protocol to make sure that the following contexts are included with a minimum number of tokens per speaker. The minimum threshold ranged from one to five tokens, depending on the importance of the vowel-consonant combination. For instance, since I excluded TRAP tokens before nasals from the analysis, I considered these allophones as less important than TRAP before other following environments (except /g/) and consequently sampled a minimum of two per speaker and style. This discussion leads to another, directly related limitation, namely the exclusion of preceding linguistic environment as a predictor variable in the statistical analysis of this study. I did so because Labov, Ash and Boberg (2006b) did not mention any phonological constraints or tendencies in this regard for the Canadian Shift, although they included this variable. Although no large effect is to be expected, the investigation of this variable would certainly be interesting in future studies.
Like Labov, Ash and Boberg (2006b), I determined phrasal-accented lexical items impressionistically. This may be understood as a severe limitation, because the perceptions of analysts in terms of sentence stress may differ across studies. Thus, they may not be replicable in this regard, although I argue that these differences are not substantial. Online tools such as FAVE, which, to the best of my knowledge, return three levels of stress marking (primary, secondary and no stress), may help to account for perceptual differences between analysts in terms of determining a vowel’s sentence stress. Having stated that, this help may, of course, be reduced by post-editing.

Another limitation is the choice of the normalization technique. Following Labov, Ash and Boberg (2006b) and most other studies on the Canadian Shift, I used a modified, speaker-extrinsic version of Nearey’s (1977) procedure without testing whether other procedures such as Watt & Fabricius’ (2002) or Lobanov’s (1971) would perform better. The modified version scales the resulting normalized values in order to convert them back into Hertz values. Thomas and Kendall (2007) caution that scaling may undo most of the normalization algorithm when only some, but not all, of the vowels of a speaker’s vowel space are included in the scaling equation – which does not apply here. In order to preclude confounding effects on possibly different results, I decided to use the procedure most common in studies on North American Englishes.

Vowel duration is also a very important effect regarding the quality of vowels, especially when certain vowel qualities are suspected to be changing. I did not include vowel duration directly as a variable, but tried to control for it in terms of excluding vowel tokens of less than 70 milliseconds in length. Further, I indirectly included the effect of vowel duration via the predictor variable number of following syllables, as multisyllabic words tend to have shorter vowels in the stressed syllable than monosyllabic words. However, the combination of time-normalized vowel duration and/or speech rate (e.g. stressed syllables per seconds) should be included in studies like mine in future research. The lack of it in this dissertation certainly constitutes a limitation, and yet, many other studies on the Canadian Shift did not look at linguistic variables at all (e.g. Boberg 2005; Hoffman 2010).

Another predictor variable that may have had an effect on the change in vowel quality in my data is lexical frequency. Although it is generally believed that it does not play a role in sound change, the studies that have looked at the effects of lexical frequency in sociophonetics in general and sound change in particular yield results that are diametrically opposed: some find that high-frequency items are affected first by innovative vowel and/or consonant realizations, others find that low-frequency items are affected first, and still others find no effect of frequency whatsoever. Again for the sake of comparability, I did not include lexical frequency as a variable in this study, but it will be investigated in future research, using the same data set (cf. Hofmann and Wagner in prep.).

The following three limitations are concerned with the sociolinguistic interview. 1) The word list I have used contains some lexical sets which are numerically underrepresented
(e.g. FACE) so that the list is skewed towards the Canadian Shift vowels. It also contains no vowels in a \(hVd\) or \(hVt\) context, which are usually included in acoustic phonetic studies in order to collect vowel tokens in citation form that are least influenced by the neighboring linguistic context. I did not include this consonant-vowel-consonant cluster, because it would have yielded too many low-frequency items that are likely to never occur in the reading passage or connected speech (e.g. \(\text{who'd}\)). I designed the word list primarily to contain high-frequency tokens for normalization, as the lexical items in the list are uttered in citation form. At the same time, I wanted to focus on spontaneous speech and thus keep the list rather short. Other researchers would have chosen other lexical items to arrive at the same goal. A related limitation is the inclusion of multisyllabic lexical items on the word list. Usually, such lists contain stressed vowels in mono- and bisyllabic lexical items only. I, however, wanted to include number of following syllables as a predictor variable and thus did not control for multisyllabic words. 2) The issue of which reading passage to include is – for sociophonetic studies – largely raised by Labov. When he initially used such a passage, authored by himself, he wanted the readers to relax and enjoy the humorous and fictitious story (cf. Labov 1963, 1972b). I used a pre-existing passage instead, *Comma Gets a Cure*, which may not pass his standards of being humorous. I chose it for the relatively large number of stressed \(\text{kit}\) and \(\text{dress}\) vowels it contains. The choice of a different (maybe better) reading passage, or even giving up the reading passage altogether for an extended word list, would not have changed the results of this study substantially, but constitutes a limitation nevertheless. 3) The determination of the perceptions regarding the presence of the low-back merger is not based on a sophisticated methodology. I simply asked the respondents whether \(\text{cot}\) and \(\text{caught}\) as well as \(\text{don}\) and \(\text{dawn}\) rhymed. The determination of the perception of the merger was, however, not the primary concern of this study. Instead, I focused on the production of the vowels. Likewise, the perceptions about the status of linguistic variables in general does not follow a thorough methodology. I am convinced that the field of (social) psychology is a rich source on such methodology, particularly concerning the design of questionnaires for studies on and the analysis of (subconscious) attitudes and linguistic identity (cf. e.g. Bortz and Döring 2006: 213-236, 253-262). Using these, however, would have made the sociolinguistic interview much shorter or maybe even impossible. In addition, I considered methodologically sound research into attitudes and identity to be another study rather than a complement to the present one. For the interpretation of my results regarding linguistic attitudes and identity, I referred to the literature instead (e.g. Clarke 2010a).

A related issue is the structure of the interview. Traditionally, the modules eliciting more formal styles (the word list and the reading passage) have been conducted at the end of the interview after the module on language (e.g. Trudgill 1974). More recent research, however, showed that this is not necessary (e.g. Schilling-Estes 1998). Consequently, the order I used in my interviews should not reduce the comparability with other studies on the Canadian Shift in general and with those of Boberg (2008b, 2010) and Labov, Ash...
and Boberg (2006b) in particular. It might be argued that this order affects the results severely – a claim I do not share, looking at the results and listening to the differences in word lists and spontaneous speech in my data. For those in doubt, I will add that Boberg (2010: 144) also presented the word list to his respondents before the connected speech section of his interview.

One of the more severe limitations in my opinion is the amount of respondents in the judgment sample of this dissertation. Although the total number of 34 informants is comparable to those 33 Canadians from coast to coast Labov, Ash and Boberg (2006b) analyzed acoustically and those 35 middle-class residents of Montreal Boberg (2005) included in his investigation of the Canadian Shift, particularly in light of the size of St. John’s by comparison, I originally planned the number of older female respondents to be higher, at least by one informant. In addition, it would have been interesting to see whether respondents of ages 65 and older showed an even more traditional pattern than those of ages 58 to 60 included in my study. At the same time, the age range of the old respondents in my data is clearly more than appropriate for the purpose of this study, as the old speakers behave significantly different from the middle-aged and young speakers with regard to the innovative vowel variants, i.e. they suffice to establish an apparent time trajectory. It would further have been interesting to study the linguistic behavior of working-class respondents. It would have particular consequences for the interpretation I offer for the behavior of the middle-class respondents, since they historically have been distinguishing themselves linguistically from the working class. Moreover, I include the variable style in my analysis, which is the one with the least interpretable pattern. If my number of participants was as high as that of traditional “first wave” VS studies (e.g. 120; Clarke 1991), the usually few tokens in word-list and reading-passage styles would have been more balanced by similarly few tokens in spontaneous speech. In total, however, the number of vowel tokens for analysis would have been the same as it is in this study. The total number of respondents included has been constrained by external forces such as finances, time, accessibility and similarly limited resources.

A final remark concerning limitations of this study relates to the issues outlined in the Prologue of this dissertation. Throughout the decades, much of the criticism of quantitative studies in general and the VS approach to researching sociolinguistic phenomena in particular has been brought forward by qualitative sociolinguists – not without reason. As has become apparent, this study has been conducted in a VS framework, which precludes all of the methodological options that a less formalized or a qualitative framework would have offered. Thus, any possible results of this study are limited to those that the VS framework allows one to find. For instance, one method in qualitative data collection is case study observation, in which an individual (or family, group, institution) is thoroughly researched and described in order to find answers to individual processes and to capture the complexity of a case comprehensively (similar to Eckert 1988, 1989a). The observation usually concentrates on larger units, systems and behavioral patterns.
rather than measuring individual variables, which only provide a hypothetical snapshot of the social reality of the respondents (cf. Bortz and Döring 2006: 321-323). With such a method, the investigation of whether young middle-class residents of St. John’s adopt the Canadian Shit in their vernaculars would include an observation of 12 to 16 hours per day in different social spheres of one representative respondent at a time. After I would have gathered all the necessary sociodemographic information from my respondent, I would have accompanied them to work or university in the mornings, watching their interactions with fellow students, staff, faculty, co-workers, superiors and customers. Depending on the respective interlocutor, I would have noted (most likely not recorded) the quality of the lax vowels’ realizations based on an ad hoc impressionistic analysis and asked for the specific biography of the interlocutors after the communicative situations between them and my respondent. Additionally, I could have evaluated, for instance, the emotional distance to the interlocutor, emotional state of the respondent during the conversation, stress level, level of familiarization with and sympathy for the interlocutor, perceived in-/formality of the situation, perceived focus on in-/formal speech, perceived social attractiveness to co-workers/fellow students/customers, social and linguistic perceptions about the interlocutor, etc. in a post-hoc conversation with my respondent. The same evaluation would have followed each conversation that has certain characteristics, such as particularly in-/formal/ or un-/usual, in different social environments (leisure, sports, hobbies, family, friends, etc.). While such an approach would have yielded much more data in terms of linguistic attitudes and identity, social networks, sociocommunal change, daily routines, socioeconomic profile, etc., it may yield a lot of data that is not comparable across informants. Furthermore, data derived from using such a methodology is difficult to relate to data that has been derived in a VS approach. All of the studies on the Canadian Shift that have been conducted, however, use a VS approach, so that I consider a qualitative research design, data collection and evaluation for the Canadian Shift in St. John’s, Newfoundland to be a second choice, i.e. a more refined follow-up study or complement to the general picture provided in this dissertation.

6.3 Outlook

As I have concluded in Section 6.1, kit is following the trajectory established by dress, which is in turn following that established by trap. I arrived at this position by comparing the status of trap in my data with that of trap in Clarke’s (1991), collected approximately 30 years prior to mine. In order to verify or falsify my claim, a follow-up study in the same speech community would thus be necessary in one generation from now.

Future studies in the same speech community should include those linguistic variables that I did not investigate in my study. These are in particular the role of phone duration and preceding phonological environment. While the former might not necessarily be too
revealing, the latter may provide a fruitful area of investigation, particularly in combination with the variable style. That is, the original study attesting the Canadian Shift in Ontario for the first time (Clarke, Elms and Youssef 1995) already emphasized the extremely important role of following phonological environment. From this perspective, they assumed that such language-internal forces could at least in part account for the fact that the Canadian Shift is so similar to the Californian Shift, although both shifts are separated by thousands of miles (1995: 224).

In terms of the recent change in the linguistic identity that promotes the adoption of innovative linguistic forms, much more work remains to be done. As I have mentioned with regard to the limitations of this study, especially the methods that have been used in psychology seem a promising framework for measuring subconscious attitudes towards innovative forms in particular, but also for investigating the linguistic identity of young urban Newfoundlanders thoroughly. In this regard, the analysis of communities of practice in St. John’s may shed some more light on the status of born-and-raised mainland Canadians in terms of being community insiders or outsiders.

Another promising but rather methodological area of interest is the inclusion of more vowels via such online tools as FAVE. As outlined in the preceding subsection, this may also help to shed more light on the behavior of style when the multitude of vowel tokens in spontaneous speech that occur in linguistic environments other than those present in word list and reading passage are removed from the data set via random sampling.

The role of lexical frequency on a sound change in progress will be investigated in Hofmann and Wagner (in prep.). Preliminary results indicate that a traditional statistical modeling of frequency does not help to determine the influence of this variable. Traditional models produce results that are hardly replicable across studies; certain methodological decisions may even yield results that are simply wrong. In summary, much work is yet to be done, with St. John’s definitely remaining a fruitful location for sociolinguistic research for years to come.
Abdi, Hervé

Adank, Patti

Adank, Patti, Roel Smits and Roeland van Hout

Adank, Patti, Roeland van Hout and Roel Smits

Adank, Patti, Roeland van Hout and Velde, Hans van de

Adger, David and Graeme Trousdale

Allen, Harold B.

Ammon, Ulrich, Norbert Dittmar, Klaus J. Mattheier and Peter Trudgill (eds.)
Ash, Sharon

Auer, Peter

Avis, Walter S.

Baayen, R. Harald

Bailey, Guy

Bailey, Guy and Erik R. Thomas

Bailey, Richard W.

Baranowski, Maciej

Bayley, Robert

Bayley, Robert, Richard Cameron and Ceil Lucas

Bayley, Robert, Richard Cameron and Ceil Lucas (eds.)

Bayley, Robert and Ceil Lucas (eds.)
Bell, Allan

Bergman, Brian, Deana Stoakes Sullivan and Luke Fisher

Bergsma, Wicher P., Marcel A. Croon and Jacques A. Hagenaars

Bernolet, Sarah, Robert J. Hardsuiker and Martin J. Pickering

Berry, David A.

Biber, Douglas and Edward Finegan (eds.)

Bigham, Douglas

Bigham, Douglas, Jessica White-Sustaita and Lars Hinrichs

Boberg, Charles


Bod, Rens, Jennifer Hay and Stefanie Jannedy


Bod, Rens, Jennifer Hay and Stefanie Jannedy (eds.)


Boersma, Paul


2011a Sound: Get Power... In Boersma (2011d), n.p.

2011b Sound: To Pitch... In Boersma (2011d), n.p.

2011c Sound: To Spectrogram... In Boersma (2011d), n.p.


Boersma, Paul (ed.)

2011d *Praat Manual*.

Boersma, Paul and David Weenink


Bortz, Jürgen and Nicola Döring


Bourdieu, Pierre

Bucholtz, Mary, Nancy Bermuda, Victor Fung, Lisa Edwards and Rosalva Vargas  
2007  Hella Nor Cal or Totally So Cal?: The Perceptual Dialectology of California.  

Bucholtz, Mary and Kira Hall  

Bühner, Markus and Matthias Ziegler  

Bybee, Joan L.  
2002  Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change.  

Cameron, Deborah  
1990  Demythologizing Sociolinguistics: Why Language Does not Reflect Society. In  


Canadian Government  
2013  Department of Transportation and Works. Retrieved 12 September 2014 from  

Caspary, Wilhelm  

Cedergren, Henrietta  

Cedergren, Henrietta and David Sankoff  

Chambers, Jack K.  


1980  Linguistic Variation and Chomsky’s ‘Homogeneous Speech Community’. In  

1989  Canadian Raising: Blocking, Fronting, etc.  


Chambers, Jack K. and Margaret Hardwick

Chambers, Jack K. and Natalie Schilling (eds.)

Chambers, Jack K. and Peter Trudgill

Chatters, Jack K., Peter Trudgill and Natalie Schilling-Estes (eds.)

Chatterjee, Samprit and Ali S. Hadi

Chatterjee, Samprit and Jeffrey S. Simonoff

Chen, Nancy F., Wade Shen, Joseph Campbell and Reva Schwartz

Cheshire, Jenny


Cheshire, Jenny (ed.)


Childs, Rebecca and Sandra Clarke


Childs, Rebecca, Paul M. De Decker, Rachel Deal, Tyler Kendall, Jennifer Thorburn, Maia Williamson and Gerard Van Herk

2010 *Stop Signs: The Intersection of Interdental Fricatives and Identity in Newfoundland*. *University of Pennsylvania Working Papers in Linguistics (PWPL)* 16 (2; Selected Papers from NWAV 38): 26–35.

Childs, Rebecca, Gerard Van Herk and Jennifer Thorburn


Chomsky, Noam


City of St. John’s


Clark, John, Collin Yallop and Janet Fletcher


Clarke, Sandra


Clarke, Sandra (ed.)

1993 Focus on ... Canada. Amsterdam: John Benjamins.

Clarke, Sandra, Ford Elms and Amani Youssef


Clarke, Sandra and Philip Hiscock


Cleophas, Ton J. and Aeilko H. Zwinderman


Clopper, Cynthia G.


Colbourne, B. Wade

Collier, Keith

Coulmas, Florian

Coulmas, Florian (ed.)

Coupland, Nikolas

Coupland, Nikolas and Adam Jaworski (eds.)

Crawley, Michael J.

Crystal, David

Cuerie, Haver C.

Cukor-Avila, Patricia and Guy Bailey


D’Arcy, Alexandra


Davies, Mark

Davis, Donna L.

Davison, John

De Decker, Paul M.

De Decker, Paul M. and Sara Mackenzie

de Jong, Kenneth

Di Paolo, Marianna and Malcah Yaeger-Dror

Di Paolo, Marianna and Malcah Yaeger-Dror (eds.)

Di Paolo, Marianna, Malcah Yaeger-Dror and Alicia B. Wassink

Disner, Sandra F.

DNE
Dollinger, Stefan

Dollinger, Stefan and Sandra Clarke

Dunphy, Carla and Will Oxford (eds.)
2008 *Memorial University of Newfoundland Occasional Student Papers in Linguistics*. St. John’s: Memorial University of Newfoundland.

Eckert, Penelope


2005 Variation, Convention, and Social Meaning. Paper Presented at the Annual Meeting of the Linguistic Society of America, Oakland Marriott City Center, Oakland.

Eckert, Penelope and Sally McConnell-Ginet


Eckert, Penelope and John R. Rickford (eds.)
2001 *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Eddington, David
Edwards, Lloyd J., Keith E. Muller, Russell D. Wolfinger, Bahjat F. Qaqish and Oliver Schabenberger

Esling, John H.

Esling, John H. and Jerold A. Edmondson

Esling, John H. and Henry J. Warkentyne

Evanini, Keelan

Evanini, Keelan and Becky Huang

Fabricius, Anne


Fabricius, Anne H., Dominic Watt and Daniel Ezra Johnson

Fairclough, Norman

Fasold, Ralph W.

Fayyad, Usama M. and Keki B. Irani

Feagin, Crawford

Ferguson, Charles A.

Finegan, Edward and Douglas Biber

Fischer, John L.

Fishman, Joshua A.

Fishman, Joshua A. and Ofelia Garcia (eds.)

Flynn, Nicholas

Fox, John

Fox, Robert A. and Ewa Jacewicz
2008 Analysis of Total Vowel Space Areas in three Regional Dialects of American English. In Proceedings of Acoustics ’08: Joint Meeting of the Acoustical Soci-

Fusco, Leah

Gagnon, David J.

Gibbons, Robert D., Dulal Bhaumik and Subhash Aryal

Giles, Howard

Giles, Howard and Peter F. Powesland

Gorman, Kyle and Daniel E. Johnson
Graham, Dougal

Gravetter, Frederick J. and Larry B. Wallnau

Gries, Stefan Th

Gumperz, John J.

Gumperz, John J. and Jenny Cook-Gumperz

Guy, Gregory R.

Hagiwara, Robert E.

Hall, Erin
fc. Style and Substance in the Canadian Shift (Manuscript). Language Variation and Change 1–43.

Hall-Lew, Lauren
Halliday, Michael A. K.
Handcock, W. Gordon
Harrell, Frank E.
Harrington, Jonathan
Harrington, Jonathan and Steve Cassidy
Hay, Jennifer, Paul Warren and Katie Drager
Hazen, Kirk
Hedeker, Donald
Heiman, Gary W.
Herbert, A. P.
1950 Independent Member. London: Methuen.
Hickey, Raymond
Higgins, Jenny

Hillenbrand, James, Laura A. Getty, Michael J. Clark and Kimberlee Wheeler

Hillenbrand, James M., Michael J. Clark and Terrance M. Nearey
2001 Effects of Consonant Environment on Vowel Formant Patterns. The Journal

Hiller, James
2014 Newfoundland and Labrador: History. In Encyclopaedia
Britannica Online. Retrieved 12 September 2014 from
http://www.britannica.com/EBchecked/topic/412929/Newfoundland-and-
Labrador/261687/British-fishery-to-colony.

Hindle, Donald
1978 Approaches to Vowel Normalization in the Study of Natural Speech. In Sankoff

Hirayama, Manami
2000 Vowels of Western Canadian English: Changes in Progress. Master’s thesis,
Tokyo University of Foreign Studies, Tokyo.

Hockett, Charles F.

Hoffman, Michol F.
at New Ways of Analyzing Variation 27, University of Georgia, Athens.
1999 Really Expansive: The Progress of Lax Vowel Lowering in a Chain Shift. Paper
presented at New Ways of Analyzing Variation 28, University of Toronto,
Toronto.
2010 The Role of Social Factors in the Canadian Vowel Shift: Evidence from
2014 Sociolinguistic Interviews. In Research Methods in Sociolinguistics: A Prac-
tical Guide, Janet Holmes and Kirk Hazen (eds.), 25–41. Chichester: Wiley-
Blackwell.

Hoffman, Michol F. and James A. Walker

Hoffmann, Thomas

Hofmann, Matthias
2011 Sociolinguistic Interview 20: Demography.
Hofmann, Matthias and Susanne Wagner
in prep. The Urban Middle Class and their many KITS: Frequency Effects in St. John’s, NL (working title).

Hollett, Pauline

Honorof, Douglas N.

Honorof, Douglas N., Jill McCullough and Barbara Somerville
2000 Comma Gets a Cure.

Horvath, Barbara

Hosmer, David W. and Stanley Lemeshow

Hung, Henrietta

Hung, Henrietta, John Davison and Jack K. Chambers

Hymes, Dell H.

Idsardi, William J.
Ivezic, Zeljko, Andrew J. Connolly, Jacob T. VanderPlas and Alexander Gray

Jacewicz, Ewa, Robert A. Fox and Joseph Salmons

Jackman, Simon

Jaeger, T. Florian
2008 Categorical Data Analysis: Away from ANOVAs (Transformation or not) and towards Logit Mixed Models. Journal of Memory and Language 59 (4): 434–446.

Johnson, Daniel E.

Joos, Martin

Joseph, John E.

Kanjee, Raageen, Scott Watter, Alexandre Sevigny and Karin R. Humphreys

Kendall, Tyler and Erik R. Thomas


Kenyon, John S. and Thomas A. Knott
Kerswill, Paul  
Kerswill, Paul, Barbara Johnstone and Ruth Wodak (eds.)  
Kieft, Michael and Elizabeth Kay-Raining Bird  
Kiesling, Scott F.  
King, Ruth and Sandra Clarke  
King, Ruth and Jennifer Wicks  
Kinloch, A. Murray and Fazilah M. Ismail  
Kirwin, William J.  
Kleinbaum, David G., Lawrence Kupper, Azhar Nizam and Eli Rosenberg  
Kramer, Matthew  
Labov, William  


Labov, William, Sharon Ash and Charles Boberg


Labov, William, Sharon Ash, Charles Boberg and Jürgen Handke


Labov, William and Maciej Baranowski


Labov, William, Maciej Baranowski and Aaron Dinkin

Labov, William, Malcah Yaeger and Richard C. Steiner

Ladefoged, Peter

Ladefoged, Peter and Sandra F. Disner

Ladefoged, Peter and Keith Johnson

Lanari, Catherine E. P.

Lave, Jean and Etienne Wenger

Lawrance, Erika

Lehiste, Ilse

Lehiste, Ilse and Gordon E. Peterson

Lennig, Matthew

Lindblom, Björn

Llamas, Carmen

Llamas, Carmen, Louise Mullany and Peter Stockwell (eds.)
Lobanov, Boris M.

Lomax, Richard G. and Debbie L. Hahs-Vaughn

Macaulay, Ronald K. S.

Mallinson, Christine and Tyler Kendall

Mannila, Heikki, Terttu Nevalainen and Helena Raumolin-Brunberg

Mayer, Joerg

Mayers, Andrew

McCullagh, Peter and John A. Nelder

McDavid, Raven I.

McDougall, Kirsty and Francis Nolan

McKinnie, Meghan and Jennifer Dailey-O’Cain

McMahon, April M. S.
Bibliography

Meechan, Marjory

Mendoza-Denton, Norma

Mendoza-Denton, Norma, Jennifer Hay and Stefanie Jannedy

Mermelstein, Paul

Mesthrie, Rajend (ed.)

Milroy, James

Milroy, James and Lesley Milroy

Milroy, Lesley

Milroy, Lesley and Matthew J. Gordon

Milroy, Lesley and Carmen Llamas

Milroy, Lesley and James Milroy
Murray, Stephen O.  
Amsterdam/Philadelphia: John Benjamins.

Nakagawa, Shinichi and Holger Schielzeth  

Nearey, Terrance M.  

Nelder, John A. and Robert W. M. Wedderburn  

Newhook, Amanda  
2002 A Sociolinguistic Study of Burnt Islands, Newfoundland. Master’s thesis, Memorial University of Newfoundland, St. John’s.

NL Heritage Website  

Noseworthy, Ronald G.  

Oakes, Michael P.  

O’Flaherty, Patrick  

Oliveira, Marco A. de  

Paddock, Harold  

Paolillo, John C.  

Paulston, Christina Bratt and G. Richard Tucker (eds.)  
Peña, Edsel A. and Elizabeth H. Slate

Peterson, Gordon E. and Harold L. Barney

Phillips, Betty S.

Pierrehumbert, Janet B., Tessa Bent, Benjamin Munson, Ann R. Bradlow and J. Michael Bailey

Plichta, Bartlomiej

Poplack, Shana

Poplack, Shana and Sali Tagliamonte

Popper, Karl R.

Potter, Mary C. and Linda Lombardi

Pratt, Terrence K.

Preston, Dennis R.

Preston, Dennis Richard and John G. Fought (eds.)

Pringle, Ian
Quinn, Gerry P. and Michael J. Keough

R Development Core Team

Reckling, Heather

Rickford, John R.

Rickford, John R. and Faye McNair-Knox

Riggs, Bertram G.

Roeder, Rebecca V.

Roeder, Rebecca V. and Matthew Hunt Gardner

Roeder, Rebecca V. and Lidia G. Jarmasz

Romaine, Suzanne
Rosenfelder, Ingrid, Joesef Fruehwald, Keelan Evanini and Jiahong Yuan


Rousseau, Pascale and David Sankoff


Sadlier-Brown, Emily


Sadlier-Brown, Emily and Meredith Tamminga


Saito, Hidetoshi


Sankoff, David


Sankoff, David (ed.)


Scherre, Maria Marta Pereira and Anthony J. Naro


Schiel, Florian


Schilling, Natalie

Schilling-Estes, Natalie

Schmied, Josef

Schneider, Edgar W. (ed.)

Schreier, Daniel, Peter Trudgill, Edgar W. Schneider and Jeffrey Williams (eds.)

Schuchardt, Hugo

Schützler, Ole

Scobbie, James M., Alice Turk, Christian Geng, Simon King, Robin Lickley and Korin Richmond
Seier, Edith

Seltman, Howard

Shapiro, Samuel S. and Martin B. Wilk

Shuy, Roger W.

Sigley, Robert

Smitherman, Geneva

Spolsky, Bernard

SPSS

Statistics Canada

Steinlen, Anja K.

Story, George M., William J. Kirwin and John D. A. Widdowson (eds.)
Sullivan, Gail M. and Richard Feinn

Summers, W. Van

Syrdal, Ann K. and Hundraj S. Gopal

Tagliamonte, Sali A.

Tagliamonte, Sali A. and R. Harald Baayen

Tagliamonte, Sali A. and Alexandra D’Arcy

The Daily

Thomas, Erik R.


Thomas, Erik R. and Tyler Kendall


TransCanadaHighway.com


Trouvain, Jürgen and William J. Barry (eds.)


Trudgill, Peter


Turk, Alice, James M. Scobie, Christian Geng, Cedric Macmartin, Ellen Bard, Barry Campbell, Catherine Dickie, Eddie Dubourg, Bill Hardcastle, Phil Hoole, Evia Kanaida, Robin Lickley, Satsuki Nakai, Marianne Pouplier, Simon King, Steve Renals, Korin Richmond, Sonja Schaeffler, Ronnie Wiegand, Kevin White and Alan Wrench


Van Herk, Gerard


Van Herk, Gerard, Rebecca Childs and Jennifer Thorburn


Van Herk, Gerard and Sarah Knee


Wenger, Etienne

1999  *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Wickens, Thomas D.


Wodak, Ruth and Martin Reisigle


Wolfram, Walt


Woods, Howard B.


Xu, Ronghui


Zeller, Christine


Zhang, Qing


Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev and Graham M. Smith

Appendix A

Interview Questionnaire

A.1 The Word List

Please read the following list of words from left to right carefully and clearly.

<table>
<thead>
<tr>
<th>fair</th>
<th>food</th>
<th>good</th>
<th>bit</th>
<th>beat</th>
</tr>
</thead>
<tbody>
<tr>
<td>taught</td>
<td>sought</td>
<td>thought</td>
<td>sure</td>
<td>foot</td>
</tr>
<tr>
<td>understood</td>
<td>loot</td>
<td>dance</td>
<td>pool</td>
<td>these</td>
</tr>
<tr>
<td>sheep</td>
<td>cheap</td>
<td>cot</td>
<td>cash</td>
<td>start</td>
</tr>
<tr>
<td>flare</td>
<td>flair</td>
<td>then</td>
<td>ten</td>
<td>zed</td>
</tr>
<tr>
<td>card</td>
<td>journey</td>
<td>further</td>
<td>machine</td>
<td>shrink</td>
</tr>
<tr>
<td>dash</td>
<td>udder</td>
<td>other</td>
<td>measure</td>
<td>let’s say</td>
</tr>
<tr>
<td>full</td>
<td>fool</td>
<td>Chevy</td>
<td>sherry</td>
<td>cheap</td>
</tr>
<tr>
<td>male</td>
<td>mail</td>
<td>kit</td>
<td>seat</td>
<td>sheet</td>
</tr>
<tr>
<td>Jews</td>
<td>juice</td>
<td>feel</td>
<td>bench</td>
<td>colonel</td>
</tr>
<tr>
<td>wash</td>
<td>watch</td>
<td>were</td>
<td>steep</td>
<td>figure</td>
</tr>
<tr>
<td>fill</td>
<td>feel</td>
<td>waist</td>
<td>west</td>
<td>showed</td>
</tr>
<tr>
<td>heart</td>
<td>hurt</td>
<td>hart</td>
<td>fort</td>
<td>fourth</td>
</tr>
<tr>
<td>mission</td>
<td>laughed</td>
<td>raft</td>
<td>worse</td>
<td>ago</td>
</tr>
<tr>
<td>dress</td>
<td>dawn</td>
<td>shoes</td>
<td>show</td>
<td>there</td>
</tr>
<tr>
<td>father</td>
<td>spoon</td>
<td>let</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>pull</td>
<td>pool</td>
<td>fell</td>
<td>fail</td>
<td>don</td>
</tr>
<tr>
<td>bliss</td>
<td>bet</td>
<td>difficult</td>
<td>bag</td>
<td>keg</td>
</tr>
<tr>
<td>lack</td>
<td>but</td>
<td>stud</td>
<td>float</td>
<td>miss</td>
</tr>
<tr>
<td>mow</td>
<td>interesting</td>
<td>sum</td>
<td>sun</td>
<td>sung</td>
</tr>
<tr>
<td>sinner</td>
<td>kid</td>
<td>cotton</td>
<td>sudden</td>
<td>often</td>
</tr>
<tr>
<td>merry</td>
<td>earthen</td>
<td>park</td>
<td>shrug</td>
<td>would</td>
</tr>
<tr>
<td>lead</td>
<td>need</td>
<td>long</td>
<td>fin</td>
<td>fill</td>
</tr>
</tbody>
</table>
### A.2 The Vowel Space Protocol

<table>
<thead>
<tr>
<th>Lexical set</th>
<th>Linguistic Context</th>
<th>WL</th>
<th>RP</th>
<th>IS</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fleece</strong> (14)</td>
<td>2 tokens before [l] (heel, feel)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Near</strong></td>
<td>2 tokens (hear, beer)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5-10 tokens in other contexts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kit</strong> (14)</td>
<td>2 tokens before [l] (bill, hill)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 tokens before nasals (bin, hint)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>at least 10 tokens in other contexts, except [r]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Face</strong> (8)</td>
<td>2 tokens before [l] (bale, hail)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 tokens in aCe/eCe context (pane, late, lane, re, cafe, Santa Fe)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 tokens in other positions (10 when Fleece shifts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dress</strong> (16)</td>
<td>2 tokens before [l] (bell, hell)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Square</strong></td>
<td>2 tokens (bear, hair)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 tokens before nasals (pen, bend)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>at least 10 tokens in other contexts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trap</strong> (25)</td>
<td>10 tokens before voiceless stops (back, tap) at least 1 velar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 token before voiced fricatives (No polysyllabic words)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2-5 tokens before voiceless fricatives (bash, bass)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 token before velar nasal (bang, bank, hang)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 tokens of heavily stressed ‘weak words’ (have, has, am, and)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 token before front nasals (ban, hand, ham, lamb)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 before voiced stops (bad, cab, non-velar)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 before voiced velar (bag, tag)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bath</strong> (5)</td>
<td>5 tokens known to be pronounced [a]-like in BrE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Start</strong> (2)</td>
<td>2 tokens (bar, hard, hard)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lot</strong> (12)</td>
<td>2 tokens before [l] (doll)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>at least 10 tokens in other contexts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>North</strong> (14)</td>
<td>2 tokens (here, horse, more; or#; ar#; orC, uar; aur)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Force</strong></td>
<td>2 tokens (four, hoarse, ore; oar#, oar#, our#, ourC, ourC, ourC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Thought</strong></td>
<td>at least 10 tokens in other contexts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Goat</strong> (2)</td>
<td>2 tokens in other environments, where it is fronting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Foot</strong> (9)</td>
<td>2 tokens before [l] (pull, full)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cure</strong></td>
<td>2 tokens before [r] (boor, poor)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A.3 The Reading Passage

Please read the following text slowly and loudly, including the copyright.

Well, here’s a story for you: Sarah Perry was a veterinary nurse who had been working daily at an old zoo in a deserted district of the territory, so she was very happy to start a new job at a superb private practice in North Square near the Duke Street Tower. That area was much nearer for her and more to her liking. Even so, on her first morning, she felt stressed. She ate a bowl of porridge, checked herself in the mirror and washed her face in a hurry. Then she put on a plain yellow dress and a fleece jacket, picked up her kit and headed for work.

When she got there, there was a woman with a goose waiting for her. The woman gave Sarah an official letter from the vet. The letter implied that the animal could be suffering from a rare form of foot-and-mouth disease, which was surprising, because normally you would only expect to see it in a dog or a goat. Sarah was sentimental, so this made her feel sorry for the beautiful bird.

Before long, that itchy goose began to strut around the office like a lunatic, which made an unsanitary mess. The goose’s owner, Mary Harrison, kept calling, “Comma, Comma”, which Sarah thought was an odd choice for a name. Comma was strong and huge, so it would take some force to trap her, but Sarah had a different idea. First she tried gently stroking the goose’s lower back with her palm, then singing a tune to her. Finally, she administered ether. Her efforts were not futile. In no time, the goose began to tire, so Sarah was able to hold onto Comma and give her a relaxing bath.

Once Sarah had managed to bathe the goose, she wiped her off with a cloth and laid her on her right side. Then Sarah confirmed the vet’s diagnosis. Almost immediately, she remembered an effective treatment that required her to measure out a lot of medicine. Sarah warned that this course of treatment might be expensive—either five or six times
the cost of penicillin. I can’t imagine paying so much, but Mrs. Harrison—a millionaire lawyer—thought it was a fair price for a cure.

Copyright 2000 Douglas N. Honorof, Jill McCullough & Barbara Somerville. All rights reserved.

A.4 The Interview

A.4.1 Section One – Demography

The questions have been adapted from Tagliamonte (2006: 38, Appendix B).

1. What is your name?
2. Where were you born?
3. Where have you lived during the first six to twelve years of your life?
4. When is your birthday?
   a) What is the best birthday party you ever had?
   b) What is the best birthday party you ever went to?
5. (Has anyone ever held a surprise birthday party for you?)
   a) Who did it?
   b) Were you really surprised or did you pretend?
6. Has anyone ever forgotten your birthday?
7. Which schools did you go to? (elementary to senior high school)
   a) Where were they (in St. John’s)?
   b) How far were they from your house?
   c) What do you remember about your teachers? Were they nice/strict (unfair or tough)?
   d) (What was the worst thing you ever saw a teacher do to a kid?)
   e) Did you ever get blamed for something you didn’t do?
   f) Were you/your friends a troublemaker?
   g) Do you remember such people from school?
8. How many semesters of college/university did you have a chance to finish?
   a) What did/do you study?/What’s your degree?
   b) Anything else?
9. Are you working now?
   a) What?
   b) Where?
   c) Is that what you wanted to do for a living? What would you like to do?
10. What was your first job?
    a) How old where you when you started to work?
    b) Do you remember what you were excited to spend your hard-earned money on?
    c) Do young people feel the same way about working that they did in your day?
11. Did you ever have to leave St. John’s for a job?
    a) When?
    b) For how long?
    c) Why?
12. Are you religious?
    a) Which religion?
    b) Same as your parents’? (Why not?)
    c) Did you ever get into an argument because of that?
A.4. The Interview

13. Are you married/engaged/in a relationship?
   a) Is your wife/husband/partner working?
   b) What? Where?

14. (How did you meet your wife/husband/partner?)

15. (How did the marriage proposal happen?)
   a) Can you remember what you said?
   b) Can you remember how your wife/husband reacted?

16. (What was your wedding like? Did anything funny/interesting happen?)

17. Where was your father/mother born?

18. What did your parents do to earn a living?
   a) What did your parents want you to do for a living?
   b) Where did father/mother live?

19. Do you know where his/her father lived (grandfather)? Where?

20. Do you know where his/her ancestors originally came from? Irish/English/Other?

A.4.2 Section Two – Open Questions

The questions have been adapted from Labov, Ash and Boberg (2006b: 32) and Tagliamonte (2006: Appendix B). Not all of them were used in the interviews, depending on the topics the interviewees chose to talk about. Subordinate questions in particular were only asked when the interviewee simply answered the main question.

Neighborhood

1. What’s St. John’s like for you?
   a) Is it a nice place to live?
   b) What kind of people live on your street? In this area?

2. What made your parents/you move here?

3. How has your neighborhood changed in your lifetime?

4. Is this the kind of neighborhood where people talk to each other?
   a) Do you know any of your neighbors? What are they like?
   b) Some people say that nowadays everybody’s just too busy to just stop by to chat. What do you think? Why do you think that has changed?
   c) Is there anyone around here you know well enough, just to walk in?
   d) Who would invite you in for coffee, just talk?
   e) Do people from around here drop by to visit?

5. Do you think the neighborhood/community could be closer together? How?

6. What do you like best about your neighborhood?

7. What are the things that make you feel good/bad about your neighborhood?

Parents and Family

1. What kind of upbringing did you have?

2. What kinds of traditions can you remember growing up with in your family?

3. Do you (plan to) keep these traditions alive with your own family?

4. What kind of kid were you when you were growing up?
   a) Were you a troublemaker?
   b) What kinds of things did you do to get into trouble?
   c) How where you punished? By who? Were you ever grounded?

5. If you got into trouble from your parents could you talk to them?
   a) Which parent would you choose to talk to?
   b) Why?
A. Interview Questionnaire

6. Where your parents really strict?
   a) What sort of person is your father?
   b) What is your mother like?

7. Do you have siblings? How many?
   a) How did being the youngest/oldest/in the middle effect how you were treated?
   b) Do you feel that your siblings got away with things that you never did or did you get away with things that they didn’t? What kinds of things?
   c) Were you close to your siblings growing up or did you fight a lot?
   d) How about now?

8. (Have you ever been really embarrassed by something your parents/siblings said or did?)
   a) What happened?
   b) How did you react?

9. (Do you ever play tricks on your sister/brother?)
   a) What’s the worst thing you ever did?
   b) … funniest thing you ever did?

10. A lot of people say that the children today aren’t like they used to be when they were growing up, do you think so? What’s the difference? Why?

Community Events

1. Did anything really big ever happen around here that you remember? Like a big fire? Or a house burned down? Or a murder?
   a) Where?
   b) Did you see it?
   c) Did people in the neighborhood help out? With food, clothes, place to stay?
   d) What about accidents or police investigations?

2. Do you remember the flood in 2006?
   a) Where were you when it happened?
   b) What did you do when it happened?
   c) How did it effect you and your family/friends? Your neighborhood?
   d) Did people in the neighborhood help out? With food, clothes, place to stay?

3. Do you watch/like The Republic of Doyle?
   a) Why (not)? It is shot in St. John’s.
   b) How do you feel about that?
   c) Are you going to/Did you go meet the cast next/last Saturday (August 13, 2011)?

4. Do you feel as a Newfoundlander or as a Canadian? Why?

5. Some people told me St. John’s is getting more and more (mainland) Canadian. Do you think that true? Why?

6. Are people moving in or moving out of St. John’s? Why?

Friends and Hobbies

1. What do you do for fun on the weekends?
2. What do you usually do with your friends?
   a) Play cards? Bowling? Go to games?
   b) Get together on holidays?
3. How many friends do you have?
   a) Do you have different circles of friends? How many?
   b) Who are your closest friends/best friend (names and addresses)? Why/How can you tell?
4. Do you consider your family your friends as well?
5. Do you have different friends at work than elsewhere (school/hobbies)?
6. How far do your friends live away from you? Are they from St. John’s?
   a) In this neighborhood?
   b) How often do you seem them?
7. Do you have/like to chat with friends online (Facebook, MySpace, MSN)?
   a) How much time do you spend online?
   b) Did you ever have an argument with someone online? What happened?
   c) Do you distinguish online friends from real friends? How are they different?

8. What are your hobbies?
   a) How did you get into that?
   b) Did you ever go into competitions? Win a competition? What happened?

9. Do you play any musical instruments?
   a) If yes, which ones? For how long? What made you start? e.g. school, parents
   b) If no, is there an instrument you would like to learn to play? Why?

10. Do you like to party/celebrate with your friends? What do you do?
    a) Did you ever wake up and not know where you were? Not remember what you’d done the night before?
    b) What was the dumbest/silliest thing you ever did when you were drinking?

11. Do you ever go to clubs?
    a) What kind of music do they play?
    b) Is that the kind you like? Just go with friends?
    c) What is your favorite song/artist to dance to?
    d) Has anything interesting/funny happened at a club you were at? What happened?

12. Have you ever witnessed a fight?
    a) Where was it?
    b) What was it about?

13. Do you ever have fights around here? How do they start?

A.4.3 Section Three – Language

Part I: Map-Drawing Task

- Have you noticed any interesting things about the way people speak English around here? [Where do you think people in Newfoundland speak differently?]

The following instructions are adapted from Bucholtz et al. (2007: 329):

Please draw the boundaries for each region in the map below. Please provide examples for the speech in these areas, if you can think of any, and add labels you would give to these areas to describe what they sound like. I am interested in your opinion based on your knowledge and your experiences, regardless of what experts or books may say. Although you may not have visited all of Newfoundland, you may have heard people from the different areas. Feel free to draw as many areas as you like and add everything that you think is important about Newfoundland English.
Part II: Rhymes

Please read the question and answer you choose out loud.

1. What vowel do you use in bag?
   a) /æ/ as in staff
   b) /ɛ/ as in beg
   c) /ɛɪ/ as in say
   d) other:

2. How do you pronounce the vowels roof, room, broom, root?
   a) /uː/ as in boot
   b) /ɑ/ as in good
   c) These words all have different vowels:

3. What vowel do you use in aunt?
   a) /æ/ as in staff
   b) /ɔː/ as in dawn
   c) /ɑ/ as in don
A.4. The Interview

4. How do you pronounce been?
   a) /iː/ as in keen
   b) /ɪ/ as in bin
   c) /ɛ/ as in Ben
   d) other:

5. Are the vowels in Mary, marry and merry all the same?
   a) Yes
   b) No, they are all different:

6. Do cot and caught sound the same for you (rhyme)?
   a) Yes
   b) No
   c) If no, do they sound similar or completely different?

7. Do don and dawn sound the same for you (rhyme)?
   a) Yes
   b) No
   c) If no, do they sound similar or completely different?

Part III: Perceptions

The questions have been adapted from Tagliamonte (2006: Appendix B).

1. Have you noticed any changes in the way people talk and sound around here?
   a) Can you tell by the way people talk around here that they come from St. John’s/a different town on the island?
   b) Do people in St. John’s sound different? E.g. in contrast to rural areas around St. John’s, in this neighborhood, etc.?

2. How can you distinguish baymen from one another?

3. How about the difference between old and young speakers?
   a) Do you sound the same as your parents? Why not?
   b) Do your parents sound the same as you? Do your kids?

4. How was it when school teachers tried to change the way you talk? What did they/you do?

5. Has anyone ever given you a hard time about the way you talk?
   a) What did they say?
   b) What did you think/do about that?

6. Do you think that how you sound plays a role in how others perceive you?
   a) Do you think that you try to change how you sound when you are in certain environments?
   b) Which ones? Why?

7. Do you speak the same way as your friends?
   a) What kinds of differences do you notice?
   b) Can you draw your circles of friends on a sheet of paper? (show figure)

8. Do you have a lot of friends from the mainland? Only from St. John’s or Newfoundland?
   a) In which circles?
   b) How much time/often do you spend/talk with them?
   c) Do these circles overlap? Which ones/How?
   d) Are there hobbies you share with your work friends?
   e) Are they really close to you? E.g. like family/confide in them
A. Interview Questionnaire

Please describe your circle of friends as shown below. You can add as much information as you like.

- Me
- work
data
- hobbies
- family
- other hobbies

A.5 Data Collection

A.5.1 In-class Advertisement Speech

- Are you born and raised in St. John’s? Are your parents Newfoundlanders?
- participate in my study “Contemporary St. John’s English”
- the interview
  - read a text and a word list
  - “draw” a map
    * circle areas of Newfoundland
    * Where do you think people talk differently?
– let’s talk about you
  * friends, community, neighborhood
  * school days, childhood
  * family, traditions, life in St. John’s

• the interviewing process

  – 1 - 1.5 hours, or however long you wish
  – a time and place convenient for you, anywhere quiet you like
  – by participating, you transfer important cultural and linguistic information
    that will help to understand, describe and preserve Newfoundland English

• all of your personal information and you identity is kept confidential

• contact me for an interview

  – visit me at: https://sites.google.com/site/stjohnsenglish
  – call me at: (709) 351-7082 (cell)
  – email me at: matt.hofmann...@...
A. Interview Questionnaire

A.5.2 Participant Background Information Form

Background Information Form

For the purpose of the study it is important that you try and answer the following questions as truthfully as possible. If you feel uncomfortable in any way answering any of the questions, please just leave them blank. Thank you very much.

1. Your name: _______________________________________

2. Date of birth (mm/dd/yyyy): ___________________________

3. Place of birth: _______________________________________

4. Communities lived in during first 6 – 12 years of your life, if other than St. John’s: _______________________________________

5. For how long have you lived in them?

6. Did you go to school in St. John’s (circle one)? Yes / No

7. What is your level of education (circle one)?
   a. some high school
   b. high school diploma
   c. some college/university
   d. college/university degree

8. If you went to college/university, was it in St. John’s (circle one)? Yes / No

9. What’s your occupation? ___________________________; In St. John’s? Yes / No

10. Your Mother’s occupation? _________________________; In St. John’s? Yes / No

11. Her level of education? ___________________________; In St. John’s? Yes / No

12. Your Father’s occupation? _________________________; In St. John’s? Yes / No

13. His level of education? ___________________________; In St. John’s? Yes / No

14. Your spouse’s (if any) occup.? _______________________; In St. John’s? Yes / No

15. Your spouse’s education? ___________________________; In St. John’s? Yes / No

16. Are you religious (circle one)? Yes / No; If yes, which religion? ___________________________

17. Where did your ancestors originally come from?
   a. Mother’s: ________________________________
   b. Father’s: ________________________________
A.5. **Data Collection**

### A.5.3 Informed Consent Form

**Certification of Informed Consent**

Title: Contemporary Urban St. John’s English  
Researcher: Matthias Hofmann, M.A., Chemnitz University of Technology, Germany.

You are invited to take part in a research project entitled “Contemporary Urban St. John’s English.”

This form is part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, you should feel free to ask at any time during the interview. Please take the time to read this carefully and to understand any other information given to you by the researcher.

The purpose of this project is to record and understand Newfoundland English. The project will look at Newfoundland English today and also at language changes. The project’s purpose is to describe Newfoundland English, not to evaluate it. The study is part of a Ph.D. Thesis.

I, _______________________________________, agree to take part in this research project.

You understand that to take part in this study you will participate in an audio-recorded conversation, involving discussion of topics such as the following:

- your neighborhood/your community/your friends  
- your school days/your childhood/life when you were younger  
- your family and your traditions/your peer group  
- life, work, leisure, etc. in Pouch Cove/St. John’s  
- important events you remember that happened to you in your life time  
- differences between generations in Pouch Cove/St. John’s

You understand that:

- the interview will take about an hour and a half or however long you wish  
- the interview can occur at a time and place that is convenient for you  
- it is entirely up to you to decide whether to take part in this research  
- you may refuse to answer any questions, to stop the interview at any time, or withdraw from the study at any time without negative consequences for you, now or in the future  
- you will not benefit directly from the study  
- by participating, you transfer important cultural and linguistic information that will help to understand, describe, and preserve Newfoundland English  
- you may obtain information of the results of this study by contacting the principal investigator

Your check mark before each part of the form and your signature at the end of the form indicate that you consent to that part of the study. You may choose to check only some parts of the form and not others.

The proposal for this research has been reviewed by the Interdisciplinary Committee on Ethics in Human Research and found to be in compliance with Memorial University’s ethics policy. If you have ethical concerns about the research (such as the way you have been treated or your rights as a participant), you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at (709) 864-2861.
I have been advised of the purpose(s) of the research for which you have interviewed me and:

___ A. I am fully aware that the interview is being digitally recorded, and that I have the right to request deletion of (any portion of) the recorded interview that I am uncomfortable with.

___ B. I agree to A. above and I understand that all the information provided will be kept confidential and that my identity and personal information will be known only to the present investigator and his research team. I also understand that my participation is voluntary. I further grant you permission to use the interview material for any academic purposes such as a PhD thesis, discussions, presentations, or any published or unpublished works.

___ C. I agree to A. and B. above and I grant you permission to deposit the digitally recorded material with the Department of Linguistics, Memorial University of Newfoundland, and the Department of English Language and Linguistics, Chemnitz University of Technology, Germany. I also grant you and the research team permission to use the stored data for other research. I understand that the material will be kept under lock and key and that all the personal information will be removed from the interviews before they are deposited in the archives. The data will be deleted after it has served its purpose.

Your signature on this form means that you have read the information about the research and have been able to ask questions about this study. You are satisfied with the answers to all of your questions and you have been given a copy of this consent form. If you sign this form, you do not give up your legal rights, and do not release the researcher(s) from their professional responsibilities.

Your Signature: ______________________________________ Date: _________________

Please print in block letters

Name: __________________________________________________________

Address: __________________________________________________________________

Phone/email: __________________________________________________________

Signature of investigator: ____________________________ Date: _________________

If you would like more information about this study, please contact:

Doctor Gerard Van Herk, Ph.D., Office: SN-3050D, Department of Linguistics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1B 3X9, Telephone: (709)864-7632, Email: gvanherk@mun.ca

or:

Matthias Hofmann, Chemnitz University of Technology, English Language and Linguistics, Reichenhainerstrasse 39, Office: 39/220, D-09126 Chemnitz, GERMANY, Email: matthias.hofmann@phil.tu-chemnitz.de
Appendix B

Normality Tests per Speaker and Age Group

B.1 Lot per Speaker

<table>
<thead>
<tr>
<th>Speaker</th>
<th>First Formant Statistic W</th>
<th>p-Value</th>
<th>Second Formant Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.868</td>
<td>0.047</td>
<td>02PDMM</td>
<td>0.959</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.865</td>
<td>0.036</td>
<td>03PSMF</td>
<td>0.947</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.914</td>
<td>0.048</td>
<td>05GKYM</td>
<td>0.905</td>
</tr>
<tr>
<td>07HPOm</td>
<td>0.969</td>
<td>0.338</td>
<td>07HPOm</td>
<td>0.974</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.919</td>
<td>0.522</td>
<td>08GCOF</td>
<td>0.844</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.969</td>
<td>0.532</td>
<td>09PDOM</td>
<td>0.931</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.969</td>
<td>0.324</td>
<td>11CDMF</td>
<td>0.978</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.878</td>
<td>0.301</td>
<td>12RDOM</td>
<td>0.897</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.925</td>
<td>0.097</td>
<td>13FCYF</td>
<td>0.914</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.826</td>
<td>0.159</td>
<td>14HLYF</td>
<td>0.941</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.982</td>
<td>0.838</td>
<td>15CSYM</td>
<td>0.984</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.798</td>
<td>0.078</td>
<td>16KPYM</td>
<td>0.936</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.986</td>
<td>0.94</td>
<td>17LEMM</td>
<td>0.976</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.98</td>
<td>0.542</td>
<td>20SCOF</td>
<td>0.97</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.954</td>
<td>0.335</td>
<td>21FJOM</td>
<td>0.96</td>
</tr>
<tr>
<td>22KCMM</td>
<td>0.891</td>
<td>0.002</td>
<td>22KCMM</td>
<td>0.962</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.938</td>
<td>0.159</td>
<td>23PMMF</td>
<td>0.933</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.976</td>
<td>0.663</td>
<td>24PSMM</td>
<td>0.976</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.955</td>
<td>0.503</td>
<td>25SAYF</td>
<td>0.974</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.852</td>
<td>0.001</td>
<td>26PRYF</td>
<td>0.959</td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results.
### B. Normality Tests per Speaker and Age Group

#### First Formant (Shapiro-Wilk of Lot)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>27WLMF</td>
<td>0.856</td>
<td>0.047</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.96</td>
<td>0.806</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.891</td>
<td>0.363</td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.912</td>
<td>0.08</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.943</td>
<td>0.097</td>
</tr>
<tr>
<td>32RROM</td>
<td>0.912</td>
<td>0.166</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.861</td>
<td>0.049</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.982</td>
<td>0.878</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.847</td>
<td>0.186</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.914</td>
<td>0.049</td>
</tr>
<tr>
<td>37GRYMY</td>
<td>0.825</td>
<td>0.128</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.947</td>
<td>0.249</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.971</td>
<td>0.62</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.965</td>
<td>0.146</td>
</tr>
</tbody>
</table>

**Values in italics represent marginally insignificant results**

#### Second Formant (Shapiro-Wilk of Lot)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>27WLMF</td>
<td>0.893</td>
<td>0.048</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.877</td>
<td>0.296</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.942</td>
<td>0.683</td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.978</td>
<td>0.911</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.949</td>
<td>0.148</td>
</tr>
<tr>
<td>32RROM</td>
<td>0.895</td>
<td>0.096</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.973</td>
<td>0.429</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.973</td>
<td>0.627</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.983</td>
<td>0.951</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.944</td>
<td>0.046</td>
</tr>
<tr>
<td>37GRYMY</td>
<td>0.908</td>
<td>0.453</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.836</td>
<td>0.002</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.94</td>
<td>0.12</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.955</td>
<td>0.055</td>
</tr>
</tbody>
</table>

**Values in italics represent marginally insignificant results**

### B.2 THOUGHT per Speaker

#### First Formant (Shapiro-Wilk of THOUGHT)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.957</td>
<td>0.763</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.905</td>
<td>0.048</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.927</td>
<td>0.138</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.889</td>
<td>0.049</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.902</td>
<td>0.441</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.959</td>
<td>0.647</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.954</td>
<td>0.693</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.903</td>
<td>0.445</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.907</td>
<td>0.263</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.95</td>
<td>0.719</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.972</td>
<td>0.908</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.962</td>
<td>0.79</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.964</td>
<td>0.566</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.937</td>
<td>0.23</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.928</td>
<td>0.394</td>
</tr>
</tbody>
</table>

#### Second Formant (Shapiro-Wilk of THOUGHT)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.854</td>
<td>0.24</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.954</td>
<td>0.773</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.949</td>
<td>0.346</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.925</td>
<td>0.18</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.95</td>
<td>0.715</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.893</td>
<td>0.063</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.983</td>
<td>0.979</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.908</td>
<td>0.047</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.91</td>
<td>0.282</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.995</td>
<td>0.981</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.867</td>
<td>0.048</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.917</td>
<td>0.518</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.963</td>
<td>0.548</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.973</td>
<td>0.831</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.93</td>
<td>0.407</td>
</tr>
</tbody>
</table>

**Values in italics represent marginally insignificant results**
## B.3 Strut per Speaker

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
<th>First Formant (Shapiro-Wilk of Thought)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22KCMM</td>
<td>0.943</td>
<td>0.175</td>
<td></td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.973</td>
<td>0.942</td>
<td></td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.959</td>
<td>0.772</td>
<td></td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.959</td>
<td>0.815</td>
<td></td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.907</td>
<td>0.102</td>
<td></td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.929</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.935</td>
<td>0.622</td>
<td></td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.816</td>
<td>0.134</td>
<td></td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.963</td>
<td>0.823</td>
<td></td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.966</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>32RROM</td>
<td>0.836</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.942</td>
<td>0.448</td>
<td></td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.873</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.955</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.86</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.891</td>
<td>0.389</td>
<td></td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.872</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.947</td>
<td>0.636</td>
<td></td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.964</td>
<td>0.574</td>
<td></td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results
### B. Normality Tests per Speaker and Age Group

#### First Formant (Shapiro-Wilk of STRUT)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic $W$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>15CSYM</td>
<td>0.958</td>
<td>0.748</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.905</td>
<td>0.402</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.943</td>
<td>0.225</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.956</td>
<td>0.622</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.947</td>
<td>0.635</td>
</tr>
<tr>
<td>22KCMM</td>
<td>0.946</td>
<td>0.333</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.908</td>
<td>0.068</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.879</td>
<td>0.048</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.964</td>
<td>0.849</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.965</td>
<td>0.645</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.834</td>
<td>0.117</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.894</td>
<td>0.342</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.978</td>
<td>0.941</td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.810</td>
<td>0.047</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.906</td>
<td>0.138</td>
</tr>
<tr>
<td>32RRROM</td>
<td>0.945</td>
<td>0.685</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.942</td>
<td>0.341</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.908</td>
<td>0.125</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.993</td>
<td>0.971</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.966</td>
<td>0.492</td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.911</td>
<td>0.446</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.941</td>
<td>0.273</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.971</td>
<td>0.893</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.924</td>
<td>0.194</td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results

#### Second Formant (Shapiro-Wilk of STRUT)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic $W$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>15CSYM</td>
<td>0.914</td>
<td>0.273</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.978</td>
<td>0.943</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.974</td>
<td>0.799</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.951</td>
<td>0.537</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.901</td>
<td>0.227</td>
</tr>
<tr>
<td>22KCMM</td>
<td>0.966</td>
<td>0.695</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.972</td>
<td>0.821</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.962</td>
<td>0.701</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.961</td>
<td>0.815</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.955</td>
<td>0.444</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.917</td>
<td>0.484</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.948</td>
<td>0.723</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.876</td>
<td>0.249</td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.937</td>
<td>0.482</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.976</td>
<td>0.947</td>
</tr>
<tr>
<td>32RRROM</td>
<td>0.884</td>
<td>0.244</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.916</td>
<td>0.127</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.955</td>
<td>0.606</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.817</td>
<td>0.135</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.969</td>
<td>0.579</td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.896</td>
<td>0.352</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.971</td>
<td>0.801</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.925</td>
<td>0.26</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.967</td>
<td>0.792</td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results

### B.4 TRAP per Speaker

#### First Formant (Shapiro-Wilk of TRAP)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic $W$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.943</td>
<td>0.429</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.922</td>
<td>0.523</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.941</td>
<td>0.346</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.916</td>
<td>0.063</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.928</td>
<td>0.089</td>
</tr>
</tbody>
</table>

#### Second Formant (Shapiro-Wilk of TRAP)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic $W$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.915</td>
<td>0.17</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.902</td>
<td>0.384</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.93</td>
<td>0.246</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.987</td>
<td>0.987</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.897</td>
<td>0.167</td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results
<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>09PDOM</td>
<td>0.982</td>
<td>0.845</td>
<td>09PDOM</td>
<td>0.956</td>
<td>0.207</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.982</td>
<td>0.927</td>
<td>11CDMF</td>
<td>0.922</td>
<td>0.065</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.855</td>
<td>0.354</td>
<td>12RDOM</td>
<td>0.899</td>
<td>0.231</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.939</td>
<td>0.075</td>
<td>13FCYF</td>
<td>0.963</td>
<td>0.345</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.913</td>
<td>0.091</td>
<td>14HLYF</td>
<td>0.941</td>
<td>0.641</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.928</td>
<td>0.144</td>
<td>15CSYM</td>
<td>0.943</td>
<td>0.269</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.93</td>
<td>0.521</td>
<td>16KPYM</td>
<td>0.976</td>
<td>0.301</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.981</td>
<td>0.812</td>
<td>17LEMM</td>
<td>0.966</td>
<td>0.35</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.981</td>
<td>0.719</td>
<td>20SCOF</td>
<td>0.948</td>
<td>0.066</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.943</td>
<td>0.535</td>
<td>21FJOM</td>
<td>0.945</td>
<td>0.56</td>
</tr>
<tr>
<td>22KCMM</td>
<td>0.922</td>
<td>0.048</td>
<td>22KCMM</td>
<td>0.965</td>
<td>0.233</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.889</td>
<td>0.047</td>
<td>23PMMF</td>
<td>0.94</td>
<td>0.24</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.962</td>
<td>0.613</td>
<td>24PSMM</td>
<td>0.889</td>
<td>0.049</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.857</td>
<td>0.142</td>
<td>25SAYF</td>
<td>0.948</td>
<td>0.713</td>
</tr>
<tr>
<td>26PYRF</td>
<td>0.923</td>
<td>0.047</td>
<td>26PYRF</td>
<td>0.847</td>
<td>0.049</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.945</td>
<td>0.065</td>
<td>27WLMF</td>
<td>0.897</td>
<td>0.078</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.918</td>
<td>0.109</td>
<td>28HLMF</td>
<td>0.976</td>
<td>0.081</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.983</td>
<td>0.241</td>
<td>29CCOF</td>
<td>0.899</td>
<td>0.199</td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.953</td>
<td>0.48</td>
<td>30PRYF</td>
<td>0.954</td>
<td>0.486</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.969</td>
<td>0.568</td>
<td>31GEMF</td>
<td>0.978</td>
<td>0.814</td>
</tr>
<tr>
<td>32RROM</td>
<td>0.831</td>
<td>0.191</td>
<td>32RROM</td>
<td>0.996</td>
<td>0.881</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.946</td>
<td>0.047</td>
<td>33GPMM</td>
<td>0.978</td>
<td>0.583</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.952</td>
<td>0.144</td>
<td>34VJMF</td>
<td>0.976</td>
<td>0.63</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.908</td>
<td>0.177</td>
<td>35DJMM</td>
<td>0.891</td>
<td>0.798</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.969</td>
<td>0.108</td>
<td>36HJYF</td>
<td>0.95</td>
<td>0.049</td>
</tr>
<tr>
<td>37GGRYM</td>
<td>0.943</td>
<td>0.901</td>
<td>37GGRYM</td>
<td>0.877</td>
<td>0.051</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.975</td>
<td>0.843</td>
<td>38GMYF</td>
<td>0.924</td>
<td>0.105</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.934</td>
<td>0.346</td>
<td>39RMYM</td>
<td>0.962</td>
<td>0.762</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.96</td>
<td>0.301</td>
<td>40ESOF</td>
<td>0.944</td>
<td>0.116</td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results
### B. Normality Tests per Speaker and Age Group

#### B.5 BATH per Speaker

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.913</td>
<td>0.771</td>
<td>02PDMM</td>
<td>0.954</td>
<td>0.181</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.909</td>
<td>0.132</td>
<td>03PSMF</td>
<td>0.903</td>
<td>0.116</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.978</td>
<td>0.154</td>
<td>05GKYM</td>
<td>0.991</td>
<td>0.649</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.898</td>
<td>0.781</td>
<td>07HPOM</td>
<td>0.978</td>
<td>0.187</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.987</td>
<td>0.099</td>
<td>08GCOF</td>
<td>0.954</td>
<td>0.231</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.988</td>
<td>0.747</td>
<td>09PDOM</td>
<td>0.933</td>
<td>0.512</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.975</td>
<td>0.722</td>
<td>11CDMF</td>
<td>0.959</td>
<td>0.177</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.935</td>
<td>0.504</td>
<td>12RDOM</td>
<td>0.869</td>
<td>0.16</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.978</td>
<td>0.413</td>
<td>13FCYF</td>
<td>0.981</td>
<td>0.487</td>
</tr>
<tr>
<td>14HLVF</td>
<td>0.998</td>
<td>0.637</td>
<td>14HLVF</td>
<td>0.934</td>
<td>0.354</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.965</td>
<td>0.611</td>
<td>15CSYM</td>
<td>0.965</td>
<td>0.179</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.991</td>
<td>0.632</td>
<td>16KPYM</td>
<td>0.923</td>
<td>0.759</td>
</tr>
<tr>
<td>17LEM</td>
<td>0.927</td>
<td>0.421</td>
<td>17LEM</td>
<td>0.912</td>
<td>0.243</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.967</td>
<td>0.367</td>
<td>20SCOF</td>
<td>0.997</td>
<td>0.091</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.977</td>
<td>0.061</td>
<td>21FJOM</td>
<td>0.984</td>
<td>0.499</td>
</tr>
<tr>
<td>22KCMM</td>
<td>0.922</td>
<td>0.258</td>
<td>22KCMM</td>
<td>0.975</td>
<td>0.317</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.918</td>
<td>0.531</td>
<td>23PMMF</td>
<td>0.919</td>
<td>0.511</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.945</td>
<td>0.387</td>
<td>24PSMM</td>
<td>0.982</td>
<td>0.347</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.934</td>
<td>0.465</td>
<td>25SAYF</td>
<td>0.918</td>
<td>0.049</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.998</td>
<td>0.421</td>
<td>26PRYF</td>
<td>0.912</td>
<td>0.831</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.919</td>
<td>0.369</td>
<td>27WLMF</td>
<td>0.943</td>
<td>0.566</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.918</td>
<td>0.109</td>
<td>28HLMF</td>
<td>0.976</td>
<td>0.081</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.973</td>
<td>0.064</td>
<td>29CCOF</td>
<td>0.966</td>
<td>0.851</td>
</tr>
<tr>
<td>30PRF</td>
<td>0.992</td>
<td>0.373</td>
<td>30PRF</td>
<td>0.982</td>
<td>0.321</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.879</td>
<td>0.047</td>
<td>31GEMF</td>
<td>0.881</td>
<td>0.799</td>
</tr>
<tr>
<td>32RROM</td>
<td>0.967</td>
<td>0.522</td>
<td>32RROM</td>
<td>0.971</td>
<td>0.019</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.971</td>
<td>0.455</td>
<td>33GPMM</td>
<td>0.933</td>
<td>0.103</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.925</td>
<td>0.076</td>
<td>34VJMF</td>
<td>0.915</td>
<td>0.544</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.952</td>
<td>0.169</td>
<td>35DJMM</td>
<td>0.994</td>
<td>0.881</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.985</td>
<td>0.099</td>
<td>36HJYF</td>
<td>0.943</td>
<td>0.087</td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.935</td>
<td>0.532</td>
<td>37GRYM</td>
<td>0.889</td>
<td>0.243</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.955</td>
<td>0.554</td>
<td>38GMYF</td>
<td>0.958</td>
<td>0.311</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.988</td>
<td>0.405</td>
<td>39RMYM</td>
<td>0.978</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Values in italics represent marginally insignificant results
<table>
<thead>
<tr>
<th>First Formant (Shapiro-Wilk of BATH)</th>
<th>Second Formant (Shapiro-Wilk of BATH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>Statistic W</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.879</td>
</tr>
</tbody>
</table>

Values in italics represent marginally significant results

## B.6 DRESS per Speaker

<table>
<thead>
<tr>
<th>First Formant (Shapiro-Wilk of DRESS)</th>
<th>Second Formant (Shapiro-Wilk of DRESS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>Statistic W</td>
</tr>
<tr>
<td>02PDMM</td>
<td>0.885</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.892</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.974</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.987</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.899</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.992</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.975</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.856</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.888</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.992</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.929</td>
</tr>
<tr>
<td>16KPYM</td>
<td>0.939</td>
</tr>
<tr>
<td>17LEMMD</td>
<td>0.979</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.99</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.952</td>
</tr>
<tr>
<td>22KCMM</td>
<td>0.973</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.985</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.985</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.905</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.957</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.839</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.904</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.914</td>
</tr>
<tr>
<td>30PRYF</td>
<td>0.927</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.951</td>
</tr>
<tr>
<td>32RRROM</td>
<td>0.866</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.975</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Values in italics represent marginally significant results
### B. Normality Tests per Speaker and Age Group

#### First Formant (Shapiro-Wilk of Dress)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>35DJMM</td>
<td>0.899</td>
<td>0.311</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.975</td>
<td>0.346</td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.989</td>
<td>0.798</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.946</td>
<td>0.049</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.981</td>
<td>0.748</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.977</td>
<td>0.225</td>
</tr>
</tbody>
</table>

#### Second Formant (Shapiro-Wilk of Dress)

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>35DJMM</td>
<td>0.976</td>
<td>0.079</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.961</td>
<td>0.088</td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.905</td>
<td>0.403</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.979</td>
<td>0.608</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.973</td>
<td>0.481</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.955</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Values in italics represent marginally significant results

#### B.7 Kit per Speaker

<table>
<thead>
<tr>
<th>F1 (Krr)</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.905</td>
<td>0.359</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.912</td>
<td>0.094</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.972</td>
<td>0.181</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.978</td>
<td>0.341</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.924</td>
<td>0.498</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.943</td>
<td>0.049</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.938</td>
<td>0.048</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.88</td>
<td>0.228</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.963</td>
<td>0.197</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.856</td>
<td>0.176</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.979</td>
<td>0.625</td>
</tr>
<tr>
<td>16KPMF</td>
<td>0.911</td>
<td>0.405</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.981</td>
<td>0.55</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.921</td>
<td>0.049</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.974</td>
<td>0.374</td>
</tr>
<tr>
<td>22KCMF</td>
<td>0.982</td>
<td>0.375</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.927</td>
<td>0.047</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.989</td>
<td>0.915</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.984</td>
<td>0.563</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.973</td>
<td>0.251</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.981</td>
<td>0.966</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.875</td>
<td>0.207</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.943</td>
<td>0.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F2 (Krr)</th>
<th>Statistic W</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>02PDMM</td>
<td>0.883</td>
<td>0.24</td>
</tr>
<tr>
<td>03PSMF</td>
<td>0.956</td>
<td>0.524</td>
</tr>
<tr>
<td>05GKYM</td>
<td>0.982</td>
<td>0.525</td>
</tr>
<tr>
<td>07HPOM</td>
<td>0.919</td>
<td>0.047</td>
</tr>
<tr>
<td>08GCOF</td>
<td>0.952</td>
<td>0.746</td>
</tr>
<tr>
<td>09PDOM</td>
<td>0.957</td>
<td>0.049</td>
</tr>
<tr>
<td>11CDMF</td>
<td>0.951</td>
<td>0.047</td>
</tr>
<tr>
<td>12RDOM</td>
<td>0.938</td>
<td>0.618</td>
</tr>
<tr>
<td>13FCYF</td>
<td>0.973</td>
<td>0.408</td>
</tr>
<tr>
<td>14HLYF</td>
<td>0.911</td>
<td>0.443</td>
</tr>
<tr>
<td>15CSYM</td>
<td>0.931</td>
<td>0.049</td>
</tr>
<tr>
<td>16KPMF</td>
<td>0.857</td>
<td>0.141</td>
</tr>
<tr>
<td>17LEMM</td>
<td>0.95</td>
<td>0.048</td>
</tr>
<tr>
<td>20SCOF</td>
<td>0.945</td>
<td>0.047</td>
</tr>
<tr>
<td>21FJOM</td>
<td>0.964</td>
<td>0.159</td>
</tr>
<tr>
<td>22KCMF</td>
<td>0.917</td>
<td>0.048</td>
</tr>
<tr>
<td>23PMMF</td>
<td>0.982</td>
<td>0.569</td>
</tr>
<tr>
<td>24PSMM</td>
<td>0.894</td>
<td>0.049</td>
</tr>
<tr>
<td>25SAYF</td>
<td>0.969</td>
<td>0.107</td>
</tr>
<tr>
<td>26PRYF</td>
<td>0.962</td>
<td>0.077</td>
</tr>
<tr>
<td>27WLMF</td>
<td>0.95</td>
<td>0.733</td>
</tr>
<tr>
<td>28HLMF</td>
<td>0.95</td>
<td>0.727</td>
</tr>
<tr>
<td>29CCOF</td>
<td>0.958</td>
<td>0.804</td>
</tr>
</tbody>
</table>

Values in italics represent marginally significant results
### B.9 All Vowels for Middle-aged Speakers

<table>
<thead>
<tr>
<th>Speaker</th>
<th>F1 (Kit)</th>
<th>Shapiro-Wilk</th>
<th>F2 (Kit)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>30PRYF</td>
<td>0.923</td>
<td>0.049</td>
<td>30PRYF</td>
<td>0.967</td>
</tr>
<tr>
<td>31GEMF</td>
<td>0.984</td>
<td>0.649</td>
<td>31GEMF</td>
<td>0.983</td>
</tr>
<tr>
<td>32RROM</td>
<td>0.958</td>
<td>0.59</td>
<td>32RROM</td>
<td>0.959</td>
</tr>
<tr>
<td>33GPMM</td>
<td>0.987</td>
<td>0.775</td>
<td>33GPMM</td>
<td>0.955</td>
</tr>
<tr>
<td>34VJMF</td>
<td>0.985</td>
<td>0.589</td>
<td>34VJMF</td>
<td>0.946</td>
</tr>
<tr>
<td>35DJMM</td>
<td>0.967</td>
<td>0.872</td>
<td>35DJMM</td>
<td>0.893</td>
</tr>
<tr>
<td>36HJYF</td>
<td>0.978</td>
<td>0.485</td>
<td>36HJYF</td>
<td>0.976</td>
</tr>
<tr>
<td>37GRYM</td>
<td>0.848</td>
<td>0.117</td>
<td>37GRYM</td>
<td>0.939</td>
</tr>
<tr>
<td>38GMYF</td>
<td>0.983</td>
<td>0.47</td>
<td>38GMYF</td>
<td>0.97</td>
</tr>
<tr>
<td>39RMYM</td>
<td>0.959</td>
<td>0.053</td>
<td>39RMYM</td>
<td>0.953</td>
</tr>
<tr>
<td>40ESOF</td>
<td>0.988</td>
<td>0.745</td>
<td>40ESOF</td>
<td>0.964</td>
</tr>
</tbody>
</table>

Values in italics represent marginally significant results

### B.8 All Vowels for Young Speakers

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>F1 (Young)</th>
<th>Shapiro-Wilk</th>
<th>F2 (Young)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOT</td>
<td>0.989</td>
<td>0.045</td>
<td>LOT</td>
<td>0.98</td>
</tr>
<tr>
<td>THOUGHT</td>
<td>0.985</td>
<td>0.252</td>
<td>THOUGHT</td>
<td>0.931</td>
</tr>
<tr>
<td>STRUT</td>
<td>0.989</td>
<td>0.308</td>
<td>STRUT</td>
<td>0.989</td>
</tr>
<tr>
<td>TRAP</td>
<td>0.993</td>
<td>0.373</td>
<td>TRAP</td>
<td>0.96</td>
</tr>
<tr>
<td>BATH</td>
<td>0.898</td>
<td>0.15</td>
<td>BATH</td>
<td>0.903</td>
</tr>
<tr>
<td>DRESS</td>
<td>0.996</td>
<td>0.449</td>
<td>DRESS</td>
<td>0.988</td>
</tr>
<tr>
<td>KIT</td>
<td>0.994</td>
<td>0.051</td>
<td>KIT</td>
<td>0.989</td>
</tr>
</tbody>
</table>

Values in italics represent significant results

### B.9 All Vowels for Middle-aged Speakers

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>F1 (Middle)</th>
<th>Shapiro-Wilk</th>
<th>F2 (Middle)</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOT</td>
<td>0.974</td>
<td>0</td>
<td>LOT</td>
<td>0.988</td>
</tr>
<tr>
<td>THOUGHT</td>
<td>0.996</td>
<td>0.93</td>
<td>THOUGHT</td>
<td>0.965</td>
</tr>
</tbody>
</table>

Values in italics represent significant results
### B. Normality Tests per Speaker and Age Group

**F1 (Middle)**  | **Shapiro-Wilk** | **F2 (Middle)**  | **Shapiro-Wilk**  
---|---|---|---
**Phoneme** | **Statistic W** | **p-Value** | **Phoneme** | **Statistic W** | **p-Value**

| STRUT | 0.99 | 0.215 | STRUT | 0.991 | 0.33 |
| TRAP | 0.972 | 0 | TRAP | 0.988 | 0.032 |
| BATH | 0.912 | 0.143 | BATH | 0.945 | 0.446 |
| DRESS | 0.997 | 0.684 | DRESS | 0.985 | 0 |
| KIT | 0.994 | 0.048 | KIT | 0.994 | 0.003 |

Values in italics represent significant results.

### B.10 All Vowels for Old Speakers

**F1 (Old)**  | **Shapiro-Wilk** | **F2 (Old)**  | **Shapiro-Wilk**  
---|---|---|---
**Phoneme** | **Statistic W** | **p-Value** | **Phoneme** | **Statistic W** | **p-Value**

| LOT | 0.994 | 0.544 | LOT | 0.967 | 0 |
| THOUGHT | 0.988 | 0.462 | THOUGHT | 0.98 | 0.118 |
| STRUT | 0.992 | 0.843 | STRUT | 0.985 | 0.352 |
| TRAP | 0.987 | 0.207 | TRAP | 0.984 | 0.101 |
| BATH | 0.954 | 0.555 | BATH | 0.977 | 0.933 |
| DRESS | 0.988 | 0.013 | DRESS | 0.98 | 0 |
| KIT | 0.983 | 0 | KIT | 0.961 | 0 |

Values in italics represent significant results.
Appendix C

Vowel Plot of Median Formant Values

Median Vowel Formant Values (ANAE Hz)

- middle
- old
- young
Appendix D

Results for the Assumptions of T-tests

D.1 Trap and Bath

D.1.1 Normality and Homoscedasticity of Formant Mean Values

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phone</th>
<th>Shapiro-Wilk Statistic W</th>
<th>p-Value</th>
<th>F-test Statistic F</th>
<th>Df</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>Trap</td>
<td>0.931</td>
<td>[0.586]</td>
<td>1.805</td>
<td>8</td>
<td>[0.533]</td>
</tr>
<tr>
<td></td>
<td>Bath</td>
<td>0.913</td>
<td>[0.455]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>Trap</td>
<td>0.934</td>
<td>[0.382]</td>
<td>2.594</td>
<td>12</td>
<td>[0.112]</td>
</tr>
<tr>
<td></td>
<td>Bath</td>
<td>0.943</td>
<td>[0.491]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>Trap</td>
<td>0.921</td>
<td>[0.398]</td>
<td>1.035</td>
<td>11</td>
<td>[0.963]</td>
</tr>
<tr>
<td></td>
<td>Bath</td>
<td>0.976</td>
<td>[0.939]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean F2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>Trap</td>
<td>0.878</td>
<td>[0.261]</td>
<td>5.458</td>
<td>8</td>
<td>[0.086]</td>
</tr>
<tr>
<td></td>
<td>Bath</td>
<td>0.946</td>
<td>[0.707]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>Trap</td>
<td>0.915</td>
<td>[0.217]</td>
<td>0.867</td>
<td>12</td>
<td>[0.808]</td>
</tr>
<tr>
<td></td>
<td>Bath</td>
<td>0.934</td>
<td>[0.382]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>Trap</td>
<td>0.873</td>
<td>[0.131]</td>
<td>2.686</td>
<td>11</td>
<td>[0.184]</td>
</tr>
<tr>
<td></td>
<td>Bath</td>
<td>0.912</td>
<td>[0.327]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### D.1.2 T-test Results of Formant Mean Values

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phone</th>
<th>T-test for independent samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRAP</td>
<td>Bath</td>
</tr>
<tr>
<td>Middle</td>
<td>Mean</td>
<td>769</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>Sd</td>
<td>38</td>
</tr>
<tr>
<td>Middle</td>
<td>Mean</td>
<td>1823</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>Sd</td>
<td>128</td>
</tr>
</tbody>
</table>

### D.2 LOT and THOUGHT

#### D.2.1 Normality and Homoscedasticity excluding Contexts before Nasals

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phone</th>
<th>Shapiro-Wilk</th>
<th>F-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Statistic W</td>
<td>p-Value</td>
</tr>
<tr>
<td>Mean F1</td>
<td>Old</td>
<td>LOT</td>
<td>0.967</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thought</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>LOT</td>
<td>0.924</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thought</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td>LOT</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thought</td>
<td>0.937</td>
</tr>
<tr>
<td>Mean F2</td>
<td>Old</td>
<td>LOT</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thought</td>
<td>0.865</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>LOT</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thought</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>Young</td>
<td>LOT</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thought</td>
<td>0.937</td>
</tr>
</tbody>
</table>
D.2.2 Normality and Homoscedasticity including Contexts before Nasals

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phone</th>
<th>Shapiro-Wilk Statistic</th>
<th>p-Value</th>
<th>F-test Statistic</th>
<th>F-test df</th>
<th>F-test p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean F1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old</td>
<td>LOT</td>
<td>0.953</td>
<td>0.725</td>
<td>0.803</td>
<td>8</td>
<td>0.764</td>
</tr>
<tr>
<td></td>
<td>THOUGHT</td>
<td>0.934</td>
<td>0.519</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>LOT</td>
<td>0.922</td>
<td>0.26</td>
<td>0.602</td>
<td>12</td>
<td>0.392</td>
</tr>
<tr>
<td></td>
<td>THOUGHT</td>
<td>0.897</td>
<td>0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>LOT</td>
<td>0.961</td>
<td>0.788</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>THOUGHT</td>
<td>0.97</td>
<td>0.884</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Mean F2   |       |                         |         |                |             |               |
| Old       | LOT   | 0.945                   | 0.637   | 1.603          | 8           | 0.52          |
|           | THOUGHT | 0.868                  | 0.116   |                |             |               |
| Middle    | LOT   | 0.973                   | 0.927   |                |             |               |
|           | THOUGHT | 0.954                  | 0.664   |                |             |               |
| Young     | LOT   | 0.945                   | 0.578   | 1.053          | 11          | 0.937         |
|           | THOUGHT | 0.973                  | 0.918   |                |             |               |

D.2.3 T-test Results of Formant Mean Values including Contexts before Nasals

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Phone</th>
<th>T-test for independent samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LOT</td>
</tr>
<tr>
<td>Old</td>
<td>F1 (Hz)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd</td>
</tr>
<tr>
<td>Middle</td>
<td>F1 (Hz)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd</td>
</tr>
<tr>
<td>Young</td>
<td>F1 (Hz)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd</td>
</tr>
<tr>
<td></td>
<td>F2 (Hz)</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sd</td>
</tr>
</tbody>
</table>
D.3 TRAP Retraction

D.3.1 Normality and Homoscedasticity excluding Pre-nasal Environments (and /g/)
E.1 LOT-THOUGHT

E.1.1 Non-pruned Decision Tree Importance for LOT-THOUGHT

The variables entered into the Decision Tree are: the individual second formant values (dependent variable), age, sex, style, local-ness index total (LItotal), phoneme label, vowel duration, number of following syllables, preceding sounds’ voicing, place and manner of articulation (PoA and MoA) and following sounds’ voicing, place and manner of articulation (PoA and MoA). F1 is not independent of F2 and has thus been disregarded.

The importance of the significant independent variables is as follows:
### E. Results from Decision Trees and Optimal Binning

**E.1.2 Pruned Decision Tree for LOT-THOUGHT**

The independent variables are the same as outlined in Appendix E.1.1 above. The importance of the significant independent variables is as follows:

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>2133.525</td>
<td>100.0%</td>
</tr>
<tr>
<td>Preceding PoA</td>
<td>1548.967</td>
<td>72.6%</td>
</tr>
<tr>
<td>Sex</td>
<td>1307.394</td>
<td>61.3%</td>
</tr>
<tr>
<td>LItotal</td>
<td>1064.171</td>
<td>49.9%</td>
</tr>
<tr>
<td>Following PoA</td>
<td>833.688</td>
<td>39.1%</td>
</tr>
<tr>
<td>Preceding MoA</td>
<td>601.052</td>
<td>28.2%</td>
</tr>
<tr>
<td>Preceding Voice</td>
<td>572.883</td>
<td>26.9%</td>
</tr>
<tr>
<td>Following MoA</td>
<td>569.471</td>
<td>26.7%</td>
</tr>
<tr>
<td>Duration</td>
<td>169.410</td>
<td>7.9%</td>
</tr>
<tr>
<td>Style</td>
<td>134.323</td>
<td>6.3%</td>
</tr>
<tr>
<td>Following Voice</td>
<td>103.326</td>
<td>4.8%</td>
</tr>
<tr>
<td>Following Syllable</td>
<td>66.106</td>
<td>3.1%</td>
</tr>
</tbody>
</table>

Growing Method: CRT   Dependent Variable: F2

The importance of each predictor is based on its overall importance for the model (sums). For my purposes here, it is only relevant that exclusion of Speaker 20SCOF
improves the Decision Tree insofar, as there are reasonable splits in age with regard to the second formant. The importance reported above shows that age has the strongest effect on formant two. The graphic representation of the resulting pruned Decision Tree, excluding speaker 20SCOF is shown here:

Without speaker 20SCOF the classification of the data makes much more sense. I take the threshold of 1385 Hz in the second formant to be a reasonable threshold for transforming that formant in a categorical variable. Unlike optimal binning, Decision
Trees split the individual second formant values based on continuous age values (20-59 years of age).

### E.1.3 Optimal Binning for LOT-THOUGHT

<table>
<thead>
<tr>
<th>Bin</th>
<th>End Point</th>
<th>Number of Cases by Level of Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower</td>
</tr>
<tr>
<td>1</td>
<td>Unbounded</td>
<td>1520.62</td>
</tr>
<tr>
<td>2</td>
<td>1520.62</td>
<td>Unbounded</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>172</td>
</tr>
</tbody>
</table>

This table includes all 34 speakers per age group. The dependent variable is formant two. The independent variables sex and age and sex crossed does not yield better results: The categories derived in order to transform formant two into a discrete variable yields values that are too large and thus includes too many tokens below the threshold. This may be understood as a confirmation that the low-back merger is not a change in progress in St. John’s English.

### E.2 Trap

#### E.2.1 Pruned Decision Tree for Trap

The variables entered into the Decision Tree are: the individual second formant values (dependent variable), age, sex, style, local-ness index total (LITotal), vowel duration, lexical item, number of following syllables, preceding sounds’ voicing, place and manner of articulation (PoA and MoA) and following sounds' voicing, place and manner of articulation (PoA and MoA). F1 is not independent of F2 and has thus been disregarded.

The importance of the significant independent variables is as follows:

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>10252.513</td>
<td>100%</td>
</tr>
<tr>
<td>Sex</td>
<td>6241.708</td>
<td>60.9%</td>
</tr>
<tr>
<td>LITotal</td>
<td>1728.490</td>
<td>16.9%</td>
</tr>
<tr>
<td>Duration</td>
<td>386.823</td>
<td>3.8%</td>
</tr>
<tr>
<td>Following Voice</td>
<td>152.510</td>
<td>1.5%</td>
</tr>
<tr>
<td>Preceding PoA</td>
<td>100.195</td>
<td>1.0%</td>
</tr>
<tr>
<td>Preceding MoA</td>
<td>0.942</td>
<td>0%</td>
</tr>
<tr>
<td>Style</td>
<td>0.078</td>
<td>0%</td>
</tr>
<tr>
<td>Following PoA</td>
<td>0.015</td>
<td>0%</td>
</tr>
</tbody>
</table>

Growing Method: CRT  Dependent Variable: F2
The other independent variables have not been selected as significant predictors. The first split is made in age, supporting the fact that it has the strongest effect on formant two. Juxtapose of the result for formant two of optimal binning (1786 Hz; cf. Appendix E.2.2) and the one of the Decision Tree (mean of 1725 Hz for speakers younger than 37) reveals a difference of 61 Hz. If the dependent variable in the Decision Tree is changed to age, instead of F2, the first split is made at 1786 Hz. The similar result indicates that the determined age groups (discretized based on a priori reasoning) are suitable categorizations of the individual ages per speaker for TRAP. The graphic representation of the resulting pruned Decision Tree is shown here:

### E.2.2 Optimal Binning for TRAP

The categorized dependent variable (formant two) is based on the three age groups. Optimal binning identified three bins:

<table>
<thead>
<tr>
<th>Bin</th>
<th>End Point</th>
<th>Number of Cases by Level of Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Old</td>
</tr>
<tr>
<td>1</td>
<td>Unbounded</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>1786.62</td>
<td>65</td>
</tr>
<tr>
<td>3</td>
<td>1935.55</td>
<td>77</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>155</td>
</tr>
</tbody>
</table>
Only most of the youngest speakers’ tokens conform to the threshold of 1786 Hz and lower, i.e. it is a very conservative value. This threshold is much lower than the one reported by Labov, Ash and Boberg (2006b: 130, 151, 219). Their threshold of 1825 Hz will result in a higher number of tokens that conform to it in my data.

**E.3 Dress**

### E.3.1 Pruned Decision Tree for Dress

The variables entered into the Decision Tree are: the individual first formant values (dependent variable), age, sex, style, local-ness index total (LItotal), vowel duration, lexical item, number of following syllables, preceding sounds’ voicing, place and manner of articulation (PoA and MoA) and following sounds’ voicing, place and manner of articulation (PoA and MoA). F2 is not independent of F1 and has thus been disregarded.

The importance of the significant independent variables is as follows:

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>916.484</td>
<td>100.0%</td>
</tr>
<tr>
<td>Preceding PoA</td>
<td>840.332</td>
<td>91.7%</td>
</tr>
<tr>
<td>Following PoA</td>
<td>264.699</td>
<td>28.9%</td>
</tr>
<tr>
<td>Style</td>
<td>129.553</td>
<td>14.1%</td>
</tr>
<tr>
<td>LItotal</td>
<td>67.778</td>
<td>7.4%</td>
</tr>
<tr>
<td>Duration</td>
<td>43.527</td>
<td>4.7%</td>
</tr>
<tr>
<td>Following PoA</td>
<td>24.899</td>
<td>2.7%</td>
</tr>
<tr>
<td>Sex</td>
<td>23.089</td>
<td>2.5%</td>
</tr>
<tr>
<td>Preceding MoA</td>
<td>0.228</td>
<td>0%</td>
</tr>
</tbody>
</table>

Growing Method: CRT  
Dependent Variable: F1

The other independent variables have not been selected as significant predictors. The first split is made in age, supporting the fact that it has the strongest effect on formant two. Juxtaposition of the result for formant one of optimal binning (670 Hz; cf. Appendix E.3.2) and the one of the Decision Tree (mean of 683 Hz for speakers younger than 37) reveals a difference of 13 Hz. The graphic representation of the resulting pruned Decision Tree is shown here:
E.3.2 Optimal Binning for DRESS

The categorized dependent variable (formant one) is based on the three age groups. Optimal binning identified two bins:

<table>
<thead>
<tr>
<th>Bin</th>
<th>End Point</th>
<th>Number of Cases by Level of Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Middle</td>
</tr>
<tr>
<td>1</td>
<td>Unbounded</td>
<td>669.54</td>
</tr>
<tr>
<td>2</td>
<td>669.54</td>
<td>Unbounded</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>302</td>
</tr>
</tbody>
</table>

With a threshold of 670 Hz and higher most old and middle-aged speaker tokens fall in a conservative realization (higher DRESS realizations, i.e. lower F1 values). Only for the youngest age group, the majority of the tokens conform to this threshold. This threshold
is higher than the one reported by Labov, Ash and Boberg (2006b: 130, 151, 219). Their threshold of 660 Hz will result in a higher number of tokens that conform to it in my data.

E.4 KIT

E.4.1 Pruned Decision Tree for KIT

The variables entered into the Decision Tree are: the individual first formant values (dependent variable), age, sex, style, local-ness index total (LItotal), vowel duration, lexical item, number of following syllables, preceding sounds’ voicing, place and manner of articulation (PoA and MoA) and following sounds’ voicing, place and manner of articulation (PoA and MoA). F2 is not independent of F1 and has thus been disregarded.

The importance of the significant independent variables is as follows:

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Importance</th>
<th>Normalized Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>320.920</td>
<td>100.0%</td>
</tr>
<tr>
<td>Duration</td>
<td>150.873</td>
<td>47.0%</td>
</tr>
<tr>
<td>Style</td>
<td>38.248</td>
<td>11.9%</td>
</tr>
<tr>
<td>Following PoA</td>
<td>9.244</td>
<td>2.9%</td>
</tr>
<tr>
<td>LItotal</td>
<td>7.518</td>
<td>2.3%</td>
</tr>
<tr>
<td>Sex</td>
<td>2.390</td>
<td>0.7%</td>
</tr>
<tr>
<td>Preceding PoA</td>
<td>0.430</td>
<td>0.1%</td>
</tr>
<tr>
<td>Preceding MoA</td>
<td>0.063</td>
<td>0%</td>
</tr>
</tbody>
</table>

Growing Method: CRT  Dependent Variable: F1

The other independent variables have not been selected as significant predictors. The first split is made in age, supporting the fact that it has the strongest effect on formant two. Juxtaposition of the result for formant one of optimal binning (525 Hz; cf. Appendix E.4.2) and the one of the Decision Tree (mean of 523 Hz for speakers younger than 35) reveals a difference of 2 Hz. The graphic representation of the resulting pruned Decision Tree is shown here:
E.4.2 Optimal Binning for Kit

The categorized dependent variable (formant one) is based on the three age groups. Optimal binning identified two bins:

<table>
<thead>
<tr>
<th>Bin</th>
<th>End Point</th>
<th>Number of Cases by Level of Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Middle</td>
</tr>
<tr>
<td>1</td>
<td>Unbounded</td>
<td>209</td>
</tr>
<tr>
<td>2</td>
<td>486.28</td>
<td>114</td>
</tr>
<tr>
<td>3</td>
<td>525.58</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>355</td>
</tr>
</tbody>
</table>

With a threshold of 525 Hz and higher most old and middle-aged speaker tokens fall in a conservative realization (higher KIT realizations, i.e. lower F1 values). Even for the youngest age group, only 47% of the tokens (230) conform to this threshold, indicating that the shift of KIT constitutes the last stage of the chain shift.
## Appendix F

### Results from Regression Analyses

#### F.1 Linear Regression of Lot-Thought

##### F.1.1 Main Model

Fixed effects:

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic $F$</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>3.018</td>
<td>36</td>
<td>1044</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Phoneme Label</strong></td>
<td><strong>1.082</strong></td>
<td>1</td>
<td>1044</td>
<td><strong>[0.299]</strong></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td><strong>1.092</strong></td>
<td>1</td>
<td>1044</td>
<td><strong>[0.296]</strong></td>
</tr>
<tr>
<td>Sex</td>
<td>1.816</td>
<td>1</td>
<td>1044</td>
<td>[0.178]</td>
</tr>
<tr>
<td># Foll. Syllable</td>
<td>0.838</td>
<td>1</td>
<td>1044</td>
<td>[0.360]</td>
</tr>
<tr>
<td>Foll. Voice</td>
<td>0.635</td>
<td>1</td>
<td>1044</td>
<td>[0.426]</td>
</tr>
<tr>
<td>Foll. PoA</td>
<td>0.878</td>
<td>3</td>
<td>1044</td>
<td>[0.854]</td>
</tr>
<tr>
<td>LItotal</td>
<td>0.936</td>
<td>1</td>
<td>1044</td>
<td>[0.334]</td>
</tr>
<tr>
<td><strong>Phoneme Label*Sex</strong></td>
<td><strong>10.079</strong></td>
<td>1</td>
<td>1044</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td>Phoneme Label*# Foll. Syll.</td>
<td>0.042</td>
<td>1</td>
<td>1044</td>
<td>[0.838]</td>
</tr>
<tr>
<td>Phoneme Label*Foll. Voice</td>
<td>1.431</td>
<td>1</td>
<td>1044</td>
<td>[0.232]</td>
</tr>
<tr>
<td>Phoneme Label*Foll. PoA</td>
<td>1.862</td>
<td>2</td>
<td>1044</td>
<td>[0.156]</td>
</tr>
<tr>
<td>Phoneme Label*Foll. MoA</td>
<td>2.525</td>
<td>2</td>
<td>1044</td>
<td>[0.081]</td>
</tr>
<tr>
<td>Sex*# Foll. Syllable</td>
<td>2.523</td>
<td>1</td>
<td>1044</td>
<td>[0.112]</td>
</tr>
<tr>
<td>Sex*Foll. Voice</td>
<td>0.609</td>
<td>1</td>
<td>1044</td>
<td>[0.436]</td>
</tr>
<tr>
<td>Sex*Foll. PoA</td>
<td>1.515</td>
<td>3</td>
<td>1044</td>
<td>[0.209]</td>
</tr>
<tr>
<td><strong>Sex*Foll. MoA</strong></td>
<td><strong>8.608</strong></td>
<td>2</td>
<td>1044</td>
<td><strong>0.000</strong></td>
</tr>
<tr>
<td># Foll. Syll.*Foll. Voice</td>
<td>0.224</td>
<td>1</td>
<td>1044</td>
<td>[0.636]</td>
</tr>
<tr>
<td># Foll. Syll.*Foll. PoA</td>
<td>0.047</td>
<td>3</td>
<td>1044</td>
<td>[0.986]</td>
</tr>
<tr>
<td># Foll. Syll.*Foll. MoA</td>
<td>0.046</td>
<td>2</td>
<td>1044</td>
<td>[0.955]</td>
</tr>
</tbody>
</table>

Values discussed in boldface; Variable (style) excluded
F. Results from Regression Analyses

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic F</th>
<th>( Df_1 )</th>
<th>( Df_2 )</th>
<th>( p )-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foll. Voice*Foll. PoA</td>
<td>2.016</td>
<td>3</td>
<td>1044</td>
<td>0.11</td>
</tr>
<tr>
<td>Foll. Voice*Foll. MoA</td>
<td>1.08</td>
<td>1</td>
<td>1044</td>
<td>0.299</td>
</tr>
<tr>
<td>Foll. PoA*Foll. MoA</td>
<td>0.619</td>
<td>1</td>
<td>1044</td>
<td>0.432</td>
</tr>
</tbody>
</table>

Values discussed in boldface; Variable (style) excluded

Random effects:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Z</th>
<th>( p )-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>7414.777</td>
<td>2239.547</td>
<td>3.311</td>
<td>0.001</td>
<td>4102.066 - 13402.738</td>
</tr>
<tr>
<td>Word</td>
<td>3346.561</td>
<td>802.371</td>
<td>4.171</td>
<td>0.000</td>
<td>2901.774 - 5354.054</td>
</tr>
<tr>
<td>Speaker*Phoneme Label</td>
<td>1016.579</td>
<td>482.501</td>
<td>2.107</td>
<td>0.035</td>
<td>400.991 - 2577.201</td>
</tr>
</tbody>
</table>

Values discussed in boldface

F.1.2 LOT

Fixed effects (due to spatial constraints, the following tables show only those interactions [IAs] that are significant):

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic F</th>
<th>( Df_1 )</th>
<th>( Df_2 )</th>
<th>( p )-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>3.377</td>
<td>28</td>
<td>705</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>4.597</td>
<td>1</td>
<td>705</td>
<td>0.032</td>
</tr>
<tr>
<td>Sex</td>
<td>2.788</td>
<td>1</td>
<td>705</td>
<td>0.095</td>
</tr>
<tr>
<td># Foll. Syllable</td>
<td>0.104</td>
<td>1</td>
<td>705</td>
<td>0.748</td>
</tr>
<tr>
<td>Foll. Voice</td>
<td>0.054</td>
<td>1</td>
<td>705</td>
<td>0.816</td>
</tr>
<tr>
<td>Foll. PoA</td>
<td>2.194</td>
<td>3</td>
<td>705</td>
<td>0.087</td>
</tr>
<tr>
<td>Ltotal</td>
<td>1.489</td>
<td>1</td>
<td>705</td>
<td>0.223</td>
</tr>
<tr>
<td>Sex*Foll. MoA</td>
<td>11.864</td>
<td>2</td>
<td>705</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Values discussed in boldface; Variable (style) & insign. IAs excluded

Random effects:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Z</th>
<th>( p )-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>8364.029</td>
<td>2466.133</td>
<td>3.392</td>
<td>0.001</td>
<td>4692.874 - 14907.065</td>
</tr>
<tr>
<td>Word</td>
<td>4130.870</td>
<td>1014.023</td>
<td>4.074</td>
<td>0.000</td>
<td>2553.250 - 6683.281</td>
</tr>
</tbody>
</table>
F.2. **Linear Regression of Trap**

### F.1.3 Stops

**Fixed effects:**

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic $F$</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>2.605</td>
<td>26</td>
<td>628</td>
<td>0.000</td>
</tr>
<tr>
<td>Phoneme Label</td>
<td>2.479</td>
<td>1</td>
<td>628</td>
<td>[0.09]</td>
</tr>
<tr>
<td>Age</td>
<td>0.943</td>
<td>1</td>
<td>628</td>
<td>[0.332]</td>
</tr>
<tr>
<td>Sex</td>
<td>2.458</td>
<td>1</td>
<td>628</td>
<td>[0.117]</td>
</tr>
<tr>
<td># Foll. Syllable</td>
<td>1.462</td>
<td>1</td>
<td>628</td>
<td>[0.227]</td>
</tr>
<tr>
<td>Foll. Sound</td>
<td>1.193</td>
<td>5</td>
<td>628</td>
<td>[0.311]</td>
</tr>
<tr>
<td>LItotal</td>
<td>0.778</td>
<td>1</td>
<td>628</td>
<td>[0.378]</td>
</tr>
<tr>
<td>Phoneme Label*Sex</td>
<td>12.002</td>
<td>1</td>
<td>628</td>
<td>0.001</td>
</tr>
<tr>
<td>Sex*Foll. Sound</td>
<td>4.182</td>
<td>5</td>
<td>628</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Random effects:**

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald $Z$</th>
<th>$p$-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>9098.677</td>
<td>2927.936</td>
<td>3.108</td>
<td>0.002</td>
<td>4842.874-17,095.948</td>
</tr>
<tr>
<td>Word</td>
<td>3583.146</td>
<td>1200.336</td>
<td>2.985</td>
<td>0.003</td>
<td>1858.302-6908.960</td>
</tr>
<tr>
<td>Speaker*Phoneme Label</td>
<td>1948.762</td>
<td>937.166</td>
<td>2.079</td>
<td>0.038</td>
<td>759.300-5001.546</td>
</tr>
</tbody>
</table>

### F.2 Linear Regression of Trap

#### F.2.1 Main Model

This model is run in SPSS. **Fixed effects:**

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic $F$</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>5.229</td>
<td>11</td>
<td>634</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>26.772</td>
<td>1</td>
<td>634</td>
<td>0.000</td>
</tr>
<tr>
<td>Sex</td>
<td>6.041</td>
<td>1</td>
<td>634</td>
<td>0.014</td>
</tr>
<tr>
<td>Voice PoA MoA</td>
<td>2.154</td>
<td>4</td>
<td>634</td>
<td>[0.073]</td>
</tr>
<tr>
<td>Style</td>
<td>0.602</td>
<td>2</td>
<td>634</td>
<td>[0.548]</td>
</tr>
<tr>
<td>LItotal</td>
<td>0.129</td>
<td>1</td>
<td>634</td>
<td>[0.720]</td>
</tr>
<tr>
<td># Foll. Syllable</td>
<td>0.017</td>
<td>2</td>
<td>634</td>
<td>[0.983]</td>
</tr>
</tbody>
</table>

Values discussed in boldface.
Random effects:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Z</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>4619.799</td>
<td>1705.606</td>
<td>2.709</td>
<td>0.007</td>
<td>2240.598 9525.380</td>
</tr>
<tr>
<td>Word</td>
<td>2167.674</td>
<td>764.277</td>
<td>2.836</td>
<td>0.005</td>
<td>1086.120 4326.235</td>
</tr>
</tbody>
</table>

Values discussed in boldface

F.2.2 Young Females

This model is run in SPSS. Fixed effects:

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic F</th>
<th>$D_f_1$</th>
<th>$D_f_2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>1.378</td>
<td>10</td>
<td>163</td>
<td>[0.190]</td>
</tr>
<tr>
<td>Style</td>
<td>4.382</td>
<td>2</td>
<td>163</td>
<td>0.014</td>
</tr>
<tr>
<td>Age</td>
<td>0.227</td>
<td>1</td>
<td>163</td>
<td>[0.635]</td>
</tr>
<tr>
<td>LItotal</td>
<td>0.093</td>
<td>1</td>
<td>163</td>
<td>[0.760]</td>
</tr>
<tr>
<td># Foll. Syllable</td>
<td>0.275</td>
<td>2</td>
<td>163</td>
<td>[0.760]</td>
</tr>
<tr>
<td>Voice PoA MoA</td>
<td>0.069</td>
<td>4</td>
<td>163</td>
<td>[0.991]</td>
</tr>
</tbody>
</table>

Values discussed in boldface

Random effects are insignificant and do not change the $p$-values of the fixed effects when excluded. Individual factors:

<table>
<thead>
<tr>
<th>Model Term</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Statistic $t$</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1693.224</td>
<td>32.967</td>
<td>51.361</td>
<td>0.000</td>
<td>1628.127 1758.321</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Passage</td>
<td>33.823</td>
<td>21.909</td>
<td>1.544</td>
<td>[0.125]</td>
<td>-9.438 77.085</td>
</tr>
<tr>
<td>Word List</td>
<td>278.280</td>
<td>105.081</td>
<td>2.648</td>
<td>0.009</td>
<td>70.786 485.775</td>
</tr>
<tr>
<td>LItotal</td>
<td>-1.961</td>
<td>6.414</td>
<td>-0.306</td>
<td>[0.760]</td>
<td>-14.627 10.704</td>
</tr>
</tbody>
</table>

Values discussed in boldface
F.3 Logistic Regression of Trap

This model is run in Rbrul. Fixed effects: Age and sex crossed ($p < 0.001$), Voice PoA MoA ($p = 0.223$), LItotal ($p = 0.331$), style ($p = 0.431$), number of following syllables ($p = 0.86$). Random effects: speaker and lexical item (Rbrul does not return $p$-values for random effects).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Log Odds</th>
<th>Tokens</th>
<th>Application Rates</th>
<th>Centered Factor Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age and Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young Females</td>
<td>2.653</td>
<td>175</td>
<td>0.663</td>
<td>0.934</td>
</tr>
<tr>
<td>Middle Females</td>
<td>1.842</td>
<td>121</td>
<td>0.413</td>
<td>0.863</td>
</tr>
<tr>
<td>Young Males</td>
<td>1.423</td>
<td>54</td>
<td>0.296</td>
<td>0.806</td>
</tr>
<tr>
<td>Old Females</td>
<td>-0.253</td>
<td>77</td>
<td>0.078</td>
<td>0.437</td>
</tr>
<tr>
<td>Old Males</td>
<td>-2.583</td>
<td>78</td>
<td>0.013</td>
<td>0.07</td>
</tr>
<tr>
<td>Middle Males</td>
<td>-3.082</td>
<td>141</td>
<td>0.014</td>
<td>0.044</td>
</tr>
<tr>
<td>Voice PoA MoA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless Labiodentals</td>
<td>0.714</td>
<td>44</td>
<td>0.477</td>
<td>[0.671]</td>
</tr>
<tr>
<td>Voiceless Apicals</td>
<td>0.147</td>
<td>191</td>
<td>0.298</td>
<td>[0.537]</td>
</tr>
<tr>
<td>Voiceless Velar Stops</td>
<td>-0.134</td>
<td>193</td>
<td>0.275</td>
<td>[0.467]</td>
</tr>
<tr>
<td>Voiced Fricatives</td>
<td>-0.252</td>
<td>72</td>
<td>0.236</td>
<td>[0.437]</td>
</tr>
<tr>
<td>Voiced Stops</td>
<td>-0.475</td>
<td>146</td>
<td>0.295</td>
<td>[0.383]</td>
</tr>
<tr>
<td>Style</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>0.127</td>
<td>503</td>
<td>0.30</td>
<td>[0.532]</td>
</tr>
<tr>
<td>Formal</td>
<td>-0.127</td>
<td>143</td>
<td>0.28</td>
<td>[0.468]</td>
</tr>
<tr>
<td>Number of Foll. Syllables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.120</td>
<td>313</td>
<td>0.291</td>
<td>[0.53]</td>
</tr>
<tr>
<td>1</td>
<td>-0.020</td>
<td>235</td>
<td>0.285</td>
<td>[0.495]</td>
</tr>
<tr>
<td>2+</td>
<td>-0.101</td>
<td>98</td>
<td>0.337</td>
<td>[0.475]</td>
</tr>
<tr>
<td>LItotal +1</td>
<td>-0.304</td>
<td></td>
<td>n.s.</td>
<td></td>
</tr>
</tbody>
</table>

deviance = 466, AIC = 498, intercept = -1.483, grand mean = 0.296; application value < 1725 Hz

F.4 Linear Regression of DRESS

This main model is run in SPSS. Fixed effects:

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic $F$</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>$p$-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>3.619</td>
<td>9</td>
<td>1036</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>9.789</td>
<td>1</td>
<td>1036</td>
<td>0.002</td>
</tr>
<tr>
<td>Glottal State</td>
<td>4.901</td>
<td>1</td>
<td>1036</td>
<td>0.027</td>
</tr>
<tr>
<td>PoA MoA</td>
<td>2.245</td>
<td>3</td>
<td>1036</td>
<td>[0.081]</td>
</tr>
</tbody>
</table>

Values discussed in boldface
F. Results from Regression Analyses

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># Foll. Syllable</td>
<td>1.730</td>
<td>1</td>
<td>1036</td>
<td>[0.189]</td>
</tr>
<tr>
<td>LItotal</td>
<td>0.424</td>
<td>1</td>
<td>1036</td>
<td>[0.515]</td>
</tr>
<tr>
<td>Sex</td>
<td>0.246</td>
<td>1</td>
<td>1036</td>
<td>[0.620]</td>
</tr>
<tr>
<td>Style</td>
<td>0.013</td>
<td>2</td>
<td>1036</td>
<td>[0.908]</td>
</tr>
</tbody>
</table>

Random effects:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Z</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaker</td>
<td>1361.060</td>
<td>422.370</td>
<td>3.222</td>
<td>0.001</td>
<td>740.848 - 2500.493</td>
</tr>
<tr>
<td>Word</td>
<td>1317.914</td>
<td>264.359</td>
<td>4.985</td>
<td>0.000</td>
<td>889.499 - 1952.670</td>
</tr>
</tbody>
</table>

Values discussed in boldface

F.5 Linear Regression of Kit

F.5.1 Main Model

This model is run in SPSS. Fixed effects:

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>6.560</td>
<td>9</td>
<td>1359</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>16.956</td>
<td>1</td>
<td>1359</td>
<td>0.000</td>
</tr>
<tr>
<td># Foll. Syllable</td>
<td>13.785</td>
<td>1</td>
<td>1359</td>
<td>0.000</td>
</tr>
<tr>
<td>Style</td>
<td>7.273</td>
<td>1</td>
<td>1359</td>
<td>0.007</td>
</tr>
<tr>
<td>PoA MoA</td>
<td>2.431</td>
<td>3</td>
<td>1359</td>
<td>[0.064]</td>
</tr>
<tr>
<td>Glottal State</td>
<td>2.773</td>
<td>1</td>
<td>1359</td>
<td>[0.096]</td>
</tr>
<tr>
<td>LItotal</td>
<td>1.428</td>
<td>1</td>
<td>1359</td>
<td>[0.232]</td>
</tr>
<tr>
<td>Sex</td>
<td>0.246</td>
<td>1</td>
<td>1359</td>
<td>[0.620]</td>
</tr>
</tbody>
</table>

Values discussed in boldface
Random effects:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Z</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker</td>
<td>304.764</td>
<td>98.296</td>
<td>3.100</td>
<td>0.002</td>
<td>161.966-573.460</td>
</tr>
<tr>
<td>Word</td>
<td>395.984</td>
<td>90.733</td>
<td>4.364</td>
<td>0.000</td>
<td>252.719-620.463</td>
</tr>
</tbody>
</table>

Values discussed in boldface

F.5.2 Kit in Monosyllabic Lexical Items

This model is run in SPSS. Fixed effects:

<table>
<thead>
<tr>
<th>Source</th>
<th>Statistic</th>
<th>$Df_1$</th>
<th>$Df_2$</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>4.380</td>
<td>9</td>
<td>627</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>15.687</td>
<td>1</td>
<td>627</td>
<td>0.000</td>
</tr>
<tr>
<td>Glottal State</td>
<td>3.963</td>
<td>1</td>
<td>627</td>
<td>0.047</td>
</tr>
<tr>
<td>PoA MoA</td>
<td>2.308</td>
<td>3</td>
<td>627</td>
<td>[0.075]</td>
</tr>
<tr>
<td>Style</td>
<td>2.237</td>
<td>2</td>
<td>627</td>
<td>[0.108]</td>
</tr>
<tr>
<td>LItotal</td>
<td>1.589</td>
<td>1</td>
<td>627</td>
<td>[0.208]</td>
</tr>
<tr>
<td>Sex</td>
<td>1.019</td>
<td>1</td>
<td>627</td>
<td>[0.313]</td>
</tr>
</tbody>
</table>

Values discussed in boldface

Random effects:

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Wald Z</th>
<th>p-Value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaker</td>
<td>408.533</td>
<td>140.534</td>
<td>2.907</td>
<td>0.004</td>
<td>208.169-801.748</td>
</tr>
<tr>
<td>Word</td>
<td>149.903</td>
<td>89.498</td>
<td>1.675</td>
<td>[0.094]</td>
<td>46.517-483.071</td>
</tr>
</tbody>
</table>
Résumé

Personal Data

Name           Matthias Hofmann
Address        (withdrawn for privacy)
Phone          (withdrawn for privacy)
Email          matthias.hofmann@phil.tu-chemnitz.de
Day of Birth   23 July 1980
Place of Birth Schlema
Marital Status (withdrawn for privacy)
Nationality    German

Work Experience

since 12/2009  Research fellow, lecturer, Ph.D. candidate
               English Language and Linguistics
               Department of English and American Studies
               Technical University Chemnitz

09/2008 – 12/2009 Teacher of English, information processing & pre-vocational training
                   Berufsschule, Internationaler Bund
                   Schörderzentrum Berufsbildungswerk Chemnitz gGmbH

04/2006 – 09/2008 Adult educator, tutor (English)
                   Foneta GbR, Chemnitz
2003 – 2007 Tutor of English and Latin in private lessons
ABACUS, Chemnitz

Education

2001–2008 Magister Artium in English and American Studies, Intercultural Communication & Psychology
Technical University Chemnitz
Grade: 1.2

09 – 11/2007 Internship as translator
Germans from Russia Heritage Collection
North Dakota State University Library, Fargo, ND, USA

2005 – 2006 Semester abroad
Universiteit Utrecht, Netherlands

2003 Latin Proficiency Certificate
Technical University Chemnitz

1997 – 1998 Student Exchange Program
Bismarck High School, Bismarck, ND, USA
High School Diploma

Teaching

WS 2014/15 S Using and Learning English World-Wide (MA1)
SS 2014 S U.S. Varieties of English (BA4)
SS 2014 S Translation Theory and Technology (MA2)
SS 2013 S U.S. Urban Black English and the Rap Game: Lexis, Phonology and Morphosyntax (BA4)
WS 2012/13 S Corpus Linguistics (BA3)
SS 2012 S Semantics (BA2)
WS 2011/12 S Using and Learning English World-Wide (MA1)
SS 2011 S Sociolinguistics (BA4)
WS 2010/11 S Sociolinguistics (BA3)
SS 2010 S Phonetics and Phonology (BA2)
WS 2009/10 T Applied Linguistics (BA3)
Conferences


2013 The Urban Middle Class and their Many kīts: Frequency Effects in St. John’s, Newfoundland. 5th International Conference on the Linguistics of Contemporary English (ICLCE 5). Austin, TX, USA (with Susanne Wagner). Sep 28.


Invited Talks


Publications

in prep. The Urban Middle Class and their Many kīts: Frequency Effects in St. John’s, NL. (with Susanne Wagner)


**Language Skills**

<table>
<thead>
<tr>
<th>Language</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>highly proficient in reading, writing, listening, speaking and translating</td>
</tr>
<tr>
<td>Latin</td>
<td>proficient in reading and translating</td>
</tr>
<tr>
<td>French</td>
<td>basic knowledge of grammar and vocabulary</td>
</tr>
<tr>
<td>Spanish</td>
<td>fundamental knowledge</td>
</tr>
</tbody>
</table>

**Computer Skills**

<table>
<thead>
<tr>
<th>Software</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS Office</td>
<td>highly proficient in using Word, Excel, Powerpoint and Access</td>
</tr>
<tr>
<td>L\LaTeX</td>
<td>proficient in creating documents and presentations</td>
</tr>
<tr>
<td>Speech Analyzer</td>
<td>proficient in transcription</td>
</tr>
<tr>
<td>Praat</td>
<td>proficient in plotting, transcription, segmentation and scripting</td>
</tr>
<tr>
<td>R</td>
<td>proficient in plotting and statistical modeling</td>
</tr>
<tr>
<td>SPSS</td>
<td>proficient in statistical modeling and data exploration</td>
</tr>
</tbody>
</table>

Chenmitz, den ____________  

Signature: ____________
H.1 Einleitung


Im Fokus dieser Arbeit steht die nicht umstrittene Antwort auf die Frage, ob sich Neufundlands junge Hauptstadtbewohner am Lautwandel, der in den 1990er Jahren auf dem kanadischen Festland identifiziert, postuliert und unter dem Namen The Canadian Shift bekannt wurde, beteiligen. Die Ursachen des umstrittenen Charakters der Beantwortung sind nach Meinung des Autors in zwei Tatsachen zu suchen: Zum Einen existiert keine soziolinguistische Variationsforschung, die sich elaborierter, akustisch-phonetischer Methodologie bedient, und die der statistischen Auswertung eine insofern aussagekräftige Stichprobengröße zu Grunde legt, als dass Idiosynkratien weniger Sprecher ausgeschlossen werden können; zum Anderen sind die Ergebnisse der wenigen verfügbaren Studien widersprüchlich.
H.2 Kanadas Vokalverschiebungen


Wortlisten. Die Stichproben dieser Untersuchungen umfassten hauptsächlich und teilweise ausschließlich junge Frauen der sozialen Mittelschicht.


H.3 Neufundlands Varietäten


H.4 Datensatz und Methodologie


Mit diesen Informanten wurden formalisierte soziolinguistische Interviews durchgeführt, die aus drei variationslinguistischen Stilen bestehen (Wortlisten, Lesepassagen, Interviewsprache) und von einer durchschnittlichen Dauer von 90 Minuten sind. Aus diesen Interviews wurden mehr als 10.000 Vokale für die Normalisierung extrahiert, von denen ca. 6000 in den linguistischen Kontexten auftreten, die in die statistische Analyse aufgenommen werden; von diesen ungefähr 6000 Vokalen sind 5000 Vokale die des Canadian Shifts. Alle Vokale treten in satzbetonten (phrasal-accented) Worten auf und sind mindestens 70 Millisekunden lang, um Zentralisierungstendenzen zu kurz realisierter Vokale auszuschließen.

Die vom Autor gewählte Methodologie orientiert sich aus oben genannten Gründen sehr stark an der des ANAE (Labov, Ash and Boberg 2006b). Die weitgehende methodologische Ähnlichkeit betrifft besonders die Wahl des Messpunktes der Vokale bei 33% ihrer Gesamtdauer, die sprecher-extrinsische Normalisierung der gemessenen Formanten basierend auf Nearey (1977) und die Wahl der unabhängigen sozialen und linguistischen Variablen für die statistische Modellierung. Wie bereits erwähnt umfassen die sozialen Variablen Alter, Geschlecht, LItotal und Stil im Labov’schen Sinne. Die linguistischen
Variablen umfassen die Anzahl der dem betonten Vokal nachfolgenden Silben, die Artikulationsart und den Artikulationsort des dem Vokal folgenden Konsonanten, sowie seine Stimmhaftigkeit.


**H.5 Ergebnisse**

Die Ergebnisse der Arbeit sprechen für sich. Zusammengefasst können so gut wie alle Forschungsfragen positiv, d.h. hypothesengetreu, beantwortet werden mit Ausnahme der, die das stilistische Profil des Canadian Shifts in St. John’s, Neufundland, betrifft. Für Stil zeigt sich über die Triangulation der Ergebnisse kein einheitliches Muster, so dass Interpretationen basierend auf Stil sehr kritisch hinterfragt werden müssen. Die eindeutigsten Ergebnisse ergeben sich für die gleichzeitig wichtigste soziale Variable Alter: unabhängig von der Methode ist Alter immer die wichtigste Variable, die die Variation in den Vokalen des Canadian Shifts bei jüngeren Sprechern und besonders bei Frauen erklärt. Die Teilnahme der neufundländischen Informanten am festlandkanadischen Lautwandel ist also hauptsächlich eine Funktion ihres Alters – ein für Lautwandel sehr typisches Ergebnis. Geschlecht spielt keine so wichtige Rolle wie Alter, weil dessen Effekte nicht ähnlich uniforme Gültigkeit haben wie die des Alters. Geschlecht spielt nur für TRAP und KIT eine signifikante Rolle. Viel stärker stehen aber linguistische Faktoren bezüglich der Unterdrückung oder Förderung der innovativen Vokalvarianten im Vordergrund. Zwar ist die Artikulationsart TRAP nachfolgender Konsonanten nicht signifikant in der dieser Arbeit zugrundeliegenden Modellierung, aber deren wichtiger Einfluss wird dennoch deut-
lich. Für dress und kit ist die Variable Stimmhaftigkeit des folgenden Konsonanten der wichtigste Konditionierungsfaktor nach Alter.

H.6 Schlussbetrachtung


Chemnitz, den __________
Unterschrift: __________