The variable pronunciations of word-final consonant clusters in a force aligned corpus of spoken French

Peter Milne

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Linguistics

Department of Linguistics
Faculty of Arts
University of Ottawa

Abstract

This thesis project examined both schwa insertion and simplification following word-final consonant clusters in a large corpus of spoken French. Two main research questions were addressed. Can a system of forced alignment reliably reproduce pronunciation judgments that closely match those of a human researcher? How do variables, such as speech style, following context, motivation for simplification and speech rate, affect the variable pronunciations of word-final consonant clusters? This project describes the creation and testing of a novel system of forced alignment capable of segmenting recorded French speech. The results of comparing the pronunciation judgments between automatic and manual methods of recognition suggest that a system of forced alignment using speaker adapted acoustic models performed better than other acoustic models; produced results that are likely to be similar to the results produced by manual identification; and that the results of forced alignment are not likely to be affected by changes in speech style or speech rate. This project also described the application of forced alignment on a corpus of natural language spoken French. The results presented in this large sample corpus analysis suggest that the dialectal differences between Québec and France are not as simple as "simplification in Québec, schwa insertion in France". While the results presented here suggest that the process of simplification following a word-final consonant cluster is similar in both dialects, the process of schwa insertion is likely to be different in each dialect. In both dialects, word-final consonant cluster simplification is more frequent in a preconsonantal context; is most likely in a spontaneous or less formal speech style and in that speech style is positively associated with higher speaking rates. Schwa insertion following a word-final consonant cluster displays much stronger dialectal differences. Schwa insertion in the dialect from France is strongly affected by following context and possibly speech style. Schwa insertion in the dialect from Québec is not affected by following context and is strongly predicted by a lack of consonant cluster simplification.

Contents

Chapter 1

Introduction

This thesis project is interested in the variable pronunciations of word-final consonant clusters in two dialects of French. In the French language, there are two well known phonological operations that target consonant clusters at word boundaries: cluster simplification and schwa insertion. Consonant cluster simplification involves, in most cases, the deletion of the final consonant of the cluster. Schwa insertion has occurred when a final schwa vowel is realized following the consonant cluster. Both are understood to alter a sequence of (phonological) consonants in order to either facilitate articulation or enhance perception of the sequence. The variability of word-final consonant clusters in French has been investigated either from the perspective of consonant cluster simplification or from the perspective of the word-final /ə $\sim \emptyset$/ alternation. Many of these studies have been based on data from introspection; some have used corpora of natural language; while a few more recent studies have been based on large databases of recorded speech. For word-final consonant clusters, both simplification and schwa insertion are highly variable in their application. Even when a particular motivation has been proposed for either schwa insertion or simplification following a word-final consonant cluster, most authors have acknowledged that the application of either remains optional in most circumstances. This optionality has made a concise description of the distribution of the variable pronunciations of word-final consonant clusters elusive. Notwithstanding the long history of investigation into schwa insertion and simplification following a word-final consonant cluster, no clear and complete picture has yet emerged of exactly which variables condition the choice of pronunciation: Simplification or schwa? It has been claimed that these variable pronunciatons may be sensitive to various phonological factors (e.g. perceptual needs, co-articulation, prosodic requirements). As well, the different pronunciations of word-final consonant clusters may vary according to dialect, speech style or speech rate. Additionally, variables such as position of the word in the utterance, word-length and word-frequency have all been demonstrated to have an effect on either consonant cluster simplification or the insertion of schwa at word boundaries. Sociolinguistic factors may also play a role in influencing the variable pronunciations of word-final consonant clusters. This lack of consensus may be due to the fact that most studies consider variables, and pronunciations, individually, despite the fact that they often interact or are correlated.

This research is concerned with two main questions, of which the answer to the one depends upon the other. In the first place, what effect do variables, such as speech style, following context, motivation for simplification and speech rate, have on the variable pronunciations of word-final consonant clusters in two dialects of French. This question will be addressed through the use of automatic forced alignment on a large corpus of natural language spoken French containing examples of recordings from two dialects and two speech styles. In the second place, because a new system of automatic forced alignment will be used to obtain the required data, this thesis project seeks to determine how likely is it that a system of automatic forced alignment will produce judgments on the variable pronunciations of word-final consonant clusters that agree with manually identified pronunciations. Therefore, the confidence in the interpretation of the results of automatic forced alignment with respect to the variable pronunciations of word-final consonant clusters in two dialects of French depends upon the confidence in the accuracy of the system of automatic forced alignment in reliably reproducing pronunciation judgments on the variable pronunciations of word-final consonant clusters that are similar to the judgments that would be obtained by a human annotator.

Forced alignment is typically a tool used in the process of constructing an automatic speech recognition system. During forced alignment, orthographic transcripts of the utterances are used to constrain an optimal alignment between existing speech models and the new speech data. A word network is constructed from the given transcription and the role of forced alignment is to determine, given a dictionary with (often multiple) pronunciations listed for words, the most likely sequence of phones (i.e. the most likely pronunciation of each word). While the technique of forced alignment has been used within the field of automatic speech recognition for several years, only recently has it been applied to research in phonetics and phonology. Little is known about how good these systems are, compared to humans, for annotation of phonetic and phonological variables.

Within the context of evaluating a novel system of forced alignment capable of segmenting recorded French speech two specific questions will be adressed. On the one hand, from within a selection of different types of acoustic models, which type of acoustic model was consistently more accurate at producing judgments on the variable pronunciations of word-final consonant clusters that most closely matched the judgments of a human listener? On the other hand, under which of a set of several conditions do the levels of agreement between manual and automatic coding significantly differ? The answer to the former gives a set of acoustic models to be used in forced alignment that will confidently reproduce pronunciation judgments similar to those of a human researcher. The answer to the latter gives a better understanding of the conditions under which the two methods of coding might be expected to disagree, and perhaps why.

Within the context of evaluating the effects of several variables on the variable pronunciations of word-final consonant clusters in two dialects of French, a further two specific questions will be addressed. In the first place, from the data provided by the use of forced alignment, how do variables such as speech style, following context, motivation for simplification and speech rate effect the probability of observing either schwa insertion or simplification following a word-final consonant cluster? In the second place, how do the strengths of these effects compare between two dialects of French, one of which seems to prefer consonant cluster simplification while the other seems to prefer schwa insertion? The answer to the first question helps to better describe the combined effects of speech style, following context, motivation for simplification and speech rate on the variable pronunciations of word-final consonant clusters in French. The answer to the second question gives a better understanding of the dialectal differences with respect to the variable pronunciations of word-final consonant clusters in French.

The presentation of this thesis investigation begins by exploring the nature of the problem posed by the variable pronunciations of word-final consonant clusters in the French language. Included in this exploration is a justification for the use of a natural language corpus, as opposed to an introspective or experimentally collected set of data, with which to investigate the effects of several variables on word-final consonant clusters in two dialects of French. Additionally, the motivation for adopting the relatively new approach of forced alignment to obtain the data required to answer the research questions is presented. Next, Chapter 2 presents the construction and evaluation of a novel system of automatic forced alignment capable of segmenting recorded French speech. The results of several measures of accuracy of this system presented towards the end of that chapter will provide evidence in support of the argument that the overall reliability of automatic labelling using a well trained system of forced alignment can be considered to be close to the one acheived by human annotators. Chapter 3 introduces the natural language corpus that was used to determine the effects of speech style, following context, motivation for simplification and speech rate on the variable pronunciations of word-final consonant clusters in two dialects of French. The results on comparing the individual and combined effects of these variables between the two dialects presented in that chapter will provide evidence in favour of the conclusions that the patterns of word-final consonant cluster simplification may be more similar between the two dialects than has previously been assumed but that there may be real dialectal differences in the patterns of schwa insertion following a word-final consonant cluster. The final chapter of this thesis project, Chapter 4, summarizes the main findings discovered during the course of this investigation along with a discussion of their implications for future research.

The variable pronunciations of word-final consonant clusters in French

Word-final consonant clusters are common in the modern French language. A word-final consonant cluster occurs when a word contains a final sequence of two or more phonological consonants. Some word-final consonant clusters are survivors from Old French; some have been introduced through borrowings or new lexical words; but the majority arose from the loss of word-final schwas in the pronunciation of French in the 17^{th} century. Regardless of their genesis, contemporary word-final consonant clusters are often associated with variable pronunciations. Consider the following three examples of the variable pronunciations of word-final consonant clusters in French:

(1) *"... titres de gloire..."*

"... claims to fame..."

a. [titʁ də glwaʁ]

b. [tit də glwaʁ]

c. [titʁə də glwaʁ]

(2) *"... prend contact avec..."*

"... contacts..."

a. [pʁɑ̃ kɔ̃takt avɛk]

b. [pʁɑ̃ kɔ̃tak avɛk]

c. ?[pʁɑ̃ kɔ̃taktə avɛk]

(3) *"... au premier ministre."*

"... to the Prime Minister."

a. [o pʁœmje ministʁ]

b. [o pʁœmje minist]

c. [o pʁœmje ministʁə]

In (1), the consonant cluster appears before a consonant-initial word, in (2) before a vowel-initial word and in (3) the cluster appears phrase-finally (or prepausal). In all three of these conditions, the word-final consonant cluster may be pronounced unmodified, as in (1-a), (2-a), (3-a), simplified, as in (1-b), (2-b), (3-b), or with a schwa, as in (1-c), (2-c), (3-c).[1] It is commonly acknowledged that speakers from Québec are more likely to simplify word-final consonant clusters while speakers from France are more likely to insert a schwa following a word-final consonant cluster. Although in practice, it is often the case that speakers from either dialect may instead select one of the other two available pronunciation variants, often unpredictably. This variation poses an interesting problem, both from the perspective of automatic speech recognition technology as well as from the perspective of understanding any dialectal differences that might exist.

From an automatic speech recognition (ASR) perspective, words in connected speech exhibit considerable variation (Johnson, 2004; Zwicky, 1972). The variable pronunciations of word-final consonant clusters in French are difficult to model within ASR technologies and cause many errors in text-to-speech (Lanchantin et al., 2008), in word recognition (Adda-Decker, 2007), or in phonetic alignment (de Mareüil, 2007; Bürki et al., 2008). Among the potential sources of variability are phonological processes, which lead to dramatic changes in the phonetic form of the words; segments are added, deleted, or replaced by others (Bürki et al., 2011a). Describing the distribution of the variable pronunciations of word-final consonant clusters in French offers an excellent opportunity to contribute to the practical solutions required to advance the accuracy and applicability of automatic speech recognition technologies.

From the perspective of understanding the dialectal differences that might exist with respect to these variable pronunciations, describing how the effects of several variables differ between the two dialects may help to inform a discussion of the differences in the phonologies required to predict and explain the variable pronunciations of word-final consonant clusters in French. The variability of word-final consonant clusters in French has been investigated either from the perspective of consonant cluster simplification or from the perspective of the word-final $\emptyset \sim$/ə/ alternation. Many of these studies have been based on data from introspection; some have used corpora of natural language; while a few more recent studies have been based on large databases of recorded speech.

Following a word-final consonant cluster, both consonant cluster simplification and schwa insertion are highly variable in their application. Even when a particular motivation has been proposed for the simplified productions (cf. Côté (2004a)) or for the schwa productions (cf. Tranel (1987)), most authors note that the application of either remains optional in most circumstances. As well, it has been claimed that these variable pronunciatons may be sensitive to various phonological conditions, whether these be perceptual needs (Côté, 2000, 2004b), the demands of co-articulation (Barnes and Kavitskaya, 2002),

[1] For most French speakers, a pronunciation such as (2c), without any prosodic boundary or hesitation between the two words, sounds difficult or unnatural and they are often taken to be ungrammatical or impossible in the literature. But they are indeed attested.

prosodic requirements (Eychenne, 2005), or even lexical retrieval (Pustka, 2011). Additionally, many studies have suggested that the application of both consonant cluster simplification and schwa insertion following word-final consonant clusters varies according to dialect (Milne and Côté, 2009), speech style (Armstrong, 1998, 2001) and speech rate (Hansen, 1994). Confusing the matter still further, variables such as position of the word in the utterance (Malécot, 1976), word-length (Adda-Decker et al., 2008) and word-frequency (Racine and Grosjean, 2002) have all been demonstrated to have an effect on either consonant cluster simplification or the insertion of schwa at word boundaries. Sociolinguistic factors may also play a role in influencing the variable pronunciations of word-final consonant clusters (Laks, 1977; Armstrong, 2001; Durand and Eychenne, 2004; Boughton, 2008).

Given the above factors, it seems unlikely that the dialectal differences between Québec and France, with respect to the variable pronunciations of word-final consonant clusters, will be as simple as "simplification in Québec, schwa insertion in France". It is hoped that the current study, using a large corpus of spoken speech comparing different dialects and styles, may help to provide valuable new information to the discussion.

In order to answer the research questions, a corpus analysis is proposed. A corpus, based on rigorous and explicit methodology, allows quantitative and qualitative analyses far superior to intuition alone (Durand and Eychenne, 2004). With regards to the dialect from Québec, schwa insertion at word boundaries is understood to be a rare occurence and intuitions would not capture the nuances under which it might occur. For example, Milne and Côté (2009) have demonstrated that in certain contexts, word-final schwa insertion can predictably surface in that dialect. As explained by Lyche and Durand (1996), only studies at the level of the utterance can provide a comprehensive and coherent account of these phonological alternations. The use of a corpus becomes an important instrument to examine the behaviour of words in connected speech and is, therefore, an indispensable part of the construction of phonological objects (Durand and Eychenne, 2004).

The choice of a corpus analysis, as opposed to an introspective or experimentally elicited set of data, presented the investigation with a practical problem. The data used for the current study was obtained by combining two corpora of recorded French speech. The first (AssNat) is a corpus of political debates from the national assemblies of Québec and France that occurred in the month of May, 2011. This corpus contains approximately 126 hours of audio recordings (66 hours from France and 60 hours from Québec) representing 439 different speakers (105 from Québec, 334 from France). The second corpus (PFC) contains the read text exercise and both recorded conversations from selected investigations in France and Québec through the Projet Phonologie du Français Contemporain (Durand et al., 2002, 2005, 2009). The PFC corpus collected for this project contains more than 120 hours of audio recordings (65 from France and 57 from Québec) representing 201 individual speakers (118 from France and 83 from Québec). When combined, these corpora contained several hundred hours of recorded speech of which only a small proportion had been previously analyzed to determine which one of the several possible pronunciations of the word-final consonant cluster was actually realized. Since manual segmentation is an incredibly time-consuming task and it was inconceivable that a single researcher could manually label all the phones, the practical problem was how to quickly and accurately, with some degree of objectivity, identify the variable pronunciations of the word-final consonant clusters that occurred in the corpus. The proposed solution was to deploy an automatic forced aligner.

The following Chapter 2 describes in some detail the system of automatic forced alignment that was used to harvest the variable pronunciations of word-final consonant clusters from a large corpus of natural language recorded French speech. Since the use of forced alignment in linguistic research is a relatively new phenomenon, a small explanation of the benefits of this approach may be warranted.

Forced alignment is typically a tool used in the process of constructing an automatic speech recognition system. In order to train an acoustic model (or, generally, a statistical pattern classifier), a training set of labelled examples is required. In speech recognition, acoustic models are used to classify continuous speech into a set of subword units such as phones, context-dependent subphones etc. Therefore, there is the problem of generating labels for training data in terms of these classes. This can be done by hand, at least in some cases. However, this work is incredibly time-consuming and suffers from consistency problems between the experts. For the many hundreds of hours of speech in corpora like the one described in this project, it is inconceivable that a researcher would manually label all the phones. Rather, transcribers are used to obtain word-level transcripts of the utterances, then the word sequence is used to constrain an optimal alignment between existing speech models and the new speech data.

The output of forced alignment contains the start and end times of each phone, the name of the phone and the word from which the phone was retrieved. These text based results make it relatively easy for the researcher to query many interesting variables from the data set. For example durations in milliseconds of individual phones or words; left and right contexts; group membership of sets of phones; and rates of speech. The formatted text is also easily transferred to the more familiar Praat text grids so that automatic acoustic measurements might be extracted using the Praat scripting language. The results of forced alignment displaying the three pronunciations of Example (1) given above are presented below in Figure 1.1, Figure 1.2 and Figure 1.3.

Forced aligners, originally developed for automatic speech recogntion (ASR), are increasingly being used in phonetic and phonological research. Schuppler et al. (2011) performed forced alignment on a corpus of spontaneous Dutch in order to investigate phonological, co-articulation, lenition and reduction

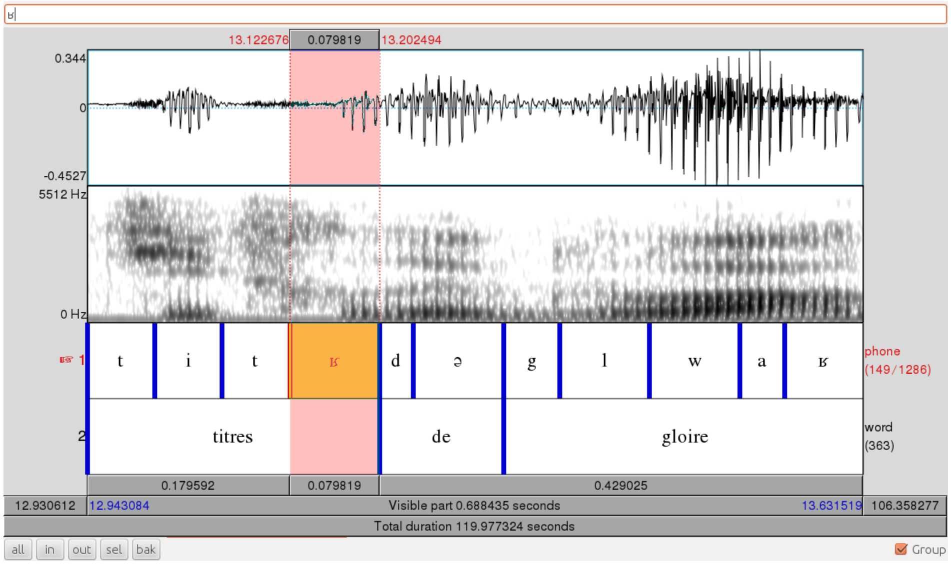

Figure 1.1: Example (1-a): *"...titres de gloire..."* canonical pronunciation

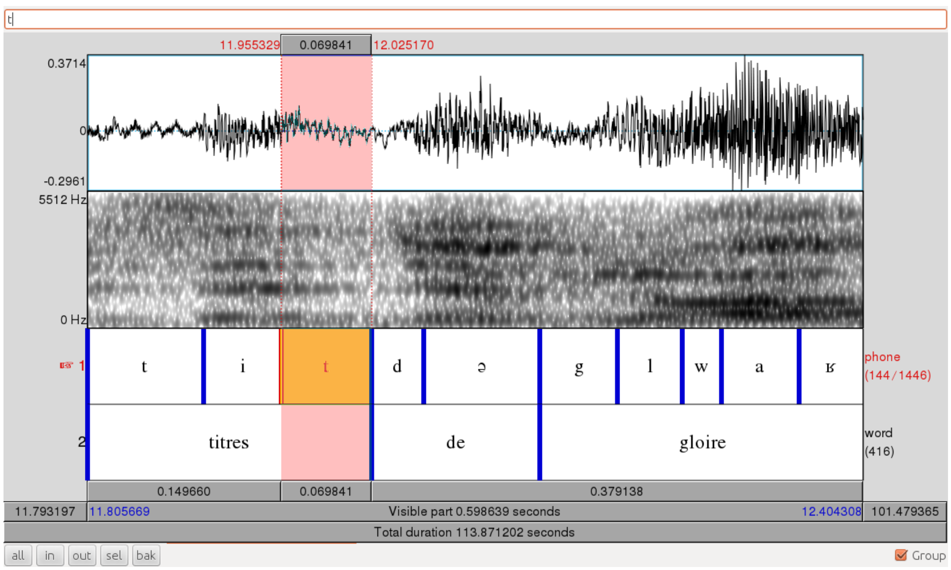

Figure 1.2: Example (1-b): *"...titres de gloire..."* pronounced simplified

rules on the basis of the automatically generated transcriptions. Their paper demonstrated how work in ASR and phonetics can benefit from each other. Their analysis of reductions was based on a speech corpus that could be automatically transcribed thanks to the availability of an ASR system. At the same time, they conclude that the ASR system will profit from incorporating the statistics about pronunciation variants that can be derived from very large corpora.

In a follow-up paper, Schuppler et al. (2012) investigated the realization of word-final /t/ in the same corpus of spontaneous Dutch. Using automatically generated transcriptions from the use of a forced aligner, they show how morphological properties of the words and their position in the utterance's

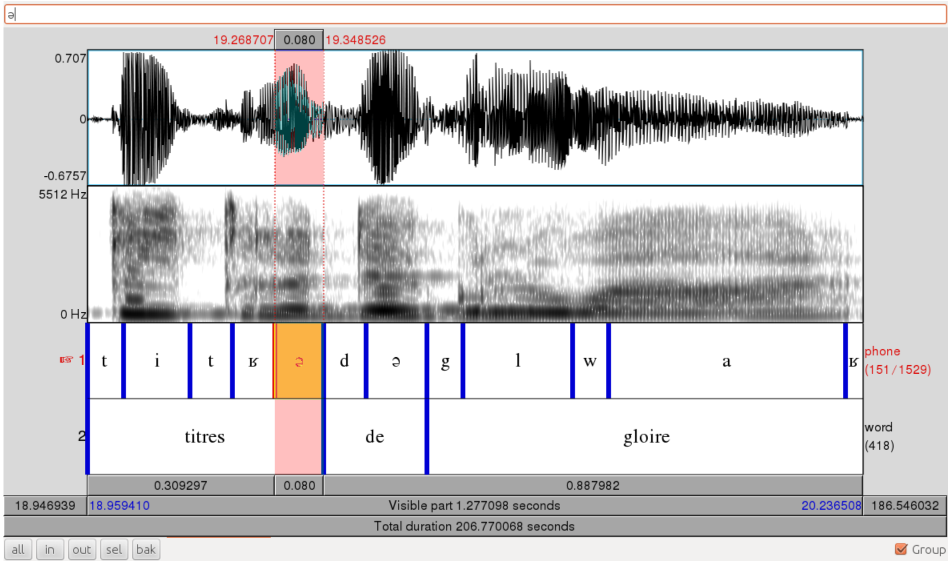

Figure 1.3: Example (1-c): *"…titres de gloire…"* pronounced with schwa

syntactic structure played a role in predicting the presence versus absence of final /t/. Additionally, their data suggested that the acoustic absence of /t/ was the natural endpoint of gradient reduction, and not its absence in the selected pronunciation variants in the mental lexicon, or from categorical phonological deletion rules.

Labov et al. (2013), studying sound changes in progress in Philadelphia, employ FAVE (Forced Alignment and Vowel Extraction), an adaptation of the Penn Phonetics Lab Forced Aligner (Yuan and Liberman, 2008), on a large corpus of neighbourhood studies which included data from 379 speakers with dates of birth from 1888 to 1991. Their detailed analysis of the progress of several vowel changes was only made possible by the development of programs for forced alignment and automatic formant measurements (Evanini, 2009). By using automatically generated transcriptions, they were able to significantly increase the number of vowel measurements obtained from a single interview from $300 \sim 500$ to $3,000 \sim 9,000$. Importantly, they were able to show that when investigating the production of speech, which is highly variable both within and across speakers, the larger sample size afforded by forced alignment helped to shrink the standard error of the mean (in their study, the mean of formant values measured in Hertz). When comparing the standard error of the mean between the automatically measured formant values with a smaller sample of manually collected values, once the number of tokens in a vowel category rose above 50, the standard error of the mean fell to less than $10Hz$ and remained there.

While investigating the adaptation of existing forced aligners to the study of endangered languages, for which little training data exists to build dedicated systems, DiCanio et al. (2013) used two forced aligners built for English to segment an isolated word corpus from Yoloxóchitl Mixtec, an Oto-Manguean language spoken in a community in Guerrero, Mexico. In general, they found that this approach had potential for making phonetic analysis of small corpora more efficient.

Among the benefits to using an accurate system of forced alignment on a large corpus of recorded speech that are relevant to the current study include the dramatic decrease in time required by manual labelling. The challenge posed by determining pronunciation judgments for every word-final consonant cluster in a very large corpus of recorded conversational speech is monumental. Even with the good fortune of having these conversations readily transcribed orthographically, it has been estimated that assisted segmentation of speech may take as long as 10 times real time (Goldman, 2011). As well, while constructing the acoustic models and language resources required for forced alignment is a serious investment of time, once they are available the time required to force align even many hours of recordings is measured in minutes. Also, like any good resource, it is available for use in all future research endeavours, thus shortening the time interval between the conception of an idea and the presentation of results.

A second benefit involves the consistent application of pronunciation judgments. While the pronunciation judgments for the two political corpora used in this investigation (from the national assemblies of both France and Québec) were manually completed by just one researcher, the pronunciation judgments for the corpora obtained through the PFC were manually completed by many different researchers, which introduces questions of consistency and reliability. While there has not been a great deal of attention paid to testing the agreement levels between human labellers with respect to variable pronunciations, what research there is suggests levels of disagreement could be significantly larger than is assumed to be the case. Using the Switchboard corpus (c.f. Godfrey et al. (1992)), a collection of spontaneous telephone conversations between American English speakers, Saraclar and Khudanpur (2004) examined 30 minutes of the 4 hour carefully labelled portion (Greenberg, 1998) to test agreement between labellers. Their analysis of the manual phonetic transcription revealed a large number ($> 20\%$) of instances where even human labellers disagreed on the actual pronunciation. They suggest that "when the deviation from the canonical pronunciation is neither small nor large, but at an intermediate level, the transcribers, who are forced to use a categorical label from the limited phonetic inventory at their disposal, may end up choosing different labels for a surface-form" (p. 376).

In a similar design, but comparing the temporal placement of phone boundaries in manually transcribed speech, Goldman (2011) found that in a 15 minute test corpus of spontaneous speech, fully manually annotated by two experts, 81% of the phone boundaries were in agreement between the two annotators with a 20ms tolerance, but only 57% of the boundaries were in agreement with a 10ms tolerance. They found these levels of agreement to be "surprisingly low despite the expertise of the annotators"(p. 4). Moreover, it is well known that human speech perception may sometimes be biased by higher-level language knowledge and understanding (c.f. Elman and McClelland (1988); Ganong (1980); Samuel and Pitt (2003)) and that this can be expected to vary from listener to listener.

Investigating inter-rater reliability in a corpus of spontaneous German, Kipp et al. (1996, 1997) observed that disagreements between human transcribers may vary between 5.6% and 21%, depending on the degree of spontaneity of the speech. Similarly, Cucchiarini and Binnenpoorte (2002) report a deviation of 12.5% for read speech and of 24.3% for spontaneous speech.

Directly relevant to the current study, Bürki et al. (2011c) found both inter- and intra-judge disagreement when determining the presence or absence of a French schwa vowel. Using a data set of words that contained an optional word-internal schwa vowel drawn from the ESTOR corpus (Galliano et al., 2005), both single and multi-judge perceptions of the presence or absence of schwa was compared. When the data were categorized based on perceptual information only, the single judge identified 8% of the tokens as ambiguous: He was unable to determine whether a schwa vowel was clearly present or clearly absent for 330 of the 4294 tokens. When the responses of a further 22 judges were considered on a subset of the data (72 items), it was found that 21 of the items were considered ambiguous in terms of the presence or absence of schwa. Furthermore, within the items classified as ambiguous, discrepancies in the responses were found both across participants as well as within the three responses to the same item when presented to each participant. Their analysis also described significant effects of schwa duration, speech rate, word length and segmental context on the judgments of participants. These results are similar to those described by Kuipers and van Donselaar (1997) who found that three phonetically trained transcribers disagreed in 10% of cases on the presence versus absence of schwa in read Dutch sentences, which is nearly twice as high as the overall disagreement between manual transcriptions of read speech (e.g. Kipp et al. (1997): 5.6%). A forced aligner, on the other hand, will consistently make the same decisions over the entire corpus.

Further benefits include the easy access to variables of interest and the possibility of performing additional acoustic measurements.

The use of forced alignment to assess the variable pronunciations of word-final consonant clusters in these data comes with some disadvantages as well. The most obvious is how to determine whether the results of forced alignment are consistent with those of human judgments. Using a collection of corpora including 360 hours of of various radio and TV shows that were used for the Technolangue-ESTER (Galliano et al., 2005) campaign, the French telephone conversation (CTS) corpus, 120 hours of LIMSI internal resources, and 10 hours of speech from the PFC, Adda-Decker and Snoeren (2011) employed the forced speech alignment tool (Adda-Decker and Lamel, 1999) based on the LIMSI speech recognition system (Gauvain et al., 1994, 2005) to produce automatic forced alignment of the data. While they do not report specifics and they are referring to comparisons of phone boundary locations (i.e. temporal agreement, not pronunciation agreement) between manual and automatic methods, they conclude that the overall reliability of automatic labelling through forced alignment can be considered close to the one achieved by human experts. In that paper they cite results from Hosom (2009) who reports over 90% agreement with a 20ms tolerance between several automatic alignments and manual labelling with respect to the placement of phone boundaries. Using the EasyAlign phonetic alignment tool, Goldman (2011) reports results on a 15 minute test corpus of spontaneous French speech indicating between 79% $\sim$ 82% agreement for the placement of phone boundaries with a 20ms tolerance between automatic and manual methods. These values were comparable with the level of agreement between two trained annotators. While one can be confident, therefore, that the automatic placement of phone boundaries, using an adequately trained forced alignment tool, is comparable to expert human performance, none of these authors report on the levels of agreement between manual and automatic methods in choosing pronunciation variants. In fact, EasyAlign requires the user to manually select the desired pronunciation *prior* to performing forced alignment, thereby reducing the role of forced alignment to determining the boundary locations of a fixed sequence of phones. Adda-Decker and Snoeren (2011) make no mention of any attempts to verify whether the pronunciation variants chosen automatically are comparable to what would have been judged manually and Hosom (2009) relies on the test partition of the TIMIT corpus of American English speech (Garofolo et al., 1990) which has been extensively labelled to represent the exact phone sequences.

A previous attempt at verifying levels of agreement between manual and automatic methods, with respect to the variable pronunciations of word-final $\emptyset \sim$/ə/ (Milne, 2011b) found that automatic pronunciation judgments agreed with the judgments of human researchers on the realizations of word final schwa in 74.3% of examined cases; was comparable in both dialects under consideration (Québec and France); and was consistent across three contexts (prevocalic, preconsonantal, prepausal). These results, however, were based on the use of a forced aligner trained for English, but modified to recognize French (Milne, 2011a). The vast majority of errors (96.9%) were of the sort where the aligner “recognized” a schwa to be realized, while the human researchers had not. It is possible that the relatively low accuracy rate, along with the over-representation of schwa realizations, could have been due to either the differences in acoustic cues between the French and English schwa vowels (phonetically, [ø/œ] vs [ə]) or the possibility that elements of word final consonant release could have been misinterpreted by the aligner as vocalic. Indeed, de Mareüil et al. (2005), using 9 hours of French TV show recordings that were both manually transcribed and automatically force aligned observed that it was often difficult to distinguish hesitations, which all corresponded to "euh" or its variants, from the pronunciation of a final schwa.

Chapter summary

This initial chapter has introduced the topic of the variable pronunciations of word-final consonant clusters in French. The two objectives in performing this research were presented. First, to produce an automatic forced alignment that faithfully represents what the results would have been had the pronunciations been manually judged by a human researcher. Second, to understand what effect variables such as, speech style, following context, motivation for simplification and speech rate have on the variable pronunciations of word-final consonant clusters in two dialects of French. These research goals were motivated by an understanding of the literature concerning the topic, as well as the practical problems of extracting meaningful linguistic data from large corpora of spoken speech. The literature on the variable pronunciations of word-final consonant clusters in French (to be presented in Section 3.1) suggests that a large-scale corpus analysis of several styles of speech from two dialects of French, examining the effects of several factors on both consonant cluster simplification and schwa insertion, may help to advance an understanding of the phonological grammar(s) that might derive these variable pronunciations.

First, the literature review will suggest that several of the factors possibly influencing schwa insertion are similar to the factors possibly influencing consonant cluster simplification. While examining the distributions of either consonant cluster simplification or schwa insertion (but never both), several researchers have associated a rising sonority contour, faster rates of speech, the amount of contrast between sequences of consonants, either dialectal or regional variation, the special status of boundaries, and a casual conversational (as opposed to formal read or broadcast news) speech style with different probabilities of consonant cluster simplification or schwa insertion following a word-final consonant cluster.

Second, the authors of several recent corpus based studies of either consonant cluster simplification or schwa insertion have highlighted the fact that it is most likely a combination of factors that will eventually best be able to explain the variable pronunciations of word-final consonant clusters in French. No single factor, whether articulatory, phonological, stylistic, or sociolinguistic appears to be able to account for all the attested distributional accounts of word-final consonant clusters.

Third, notwithstanding the possibly similar factors affecting consonant cluster simplification and schwa insertion, as well as the likelihood of interaction among them, the distribution of both consonant cluster simplification and schwa insertion together has yet to be investigated. Most published accounts having as their object of study the variable pronunciations of word-final consonant clusters in French have investigated either consonant cluster simplification or schwa insertion but not both.

Chapter 2

SPLAligner

The value in working with natural language corpora is the ability to collect large volumes of empirical data with which to test research hypotheses. The challenge in generating these data is how to quickly and accurately, with some degree of objectivity, identify linguistic units as data points. Accumulating the linguistic data required to test and evaluate hypotheses can be a time consuming and labour intensive job. While there exist many computational tools available to the professional linguist, it is often the case that the unique needs of a specific research project are not adequately covered by available software. One of the goals of this chapter is to illustrate how several freely available tools can be used to generate the data required for empirical analysis.

In this chapter is discussed the construction and evaluation of an automatic forced aligner. Forced alignment is typically a tool used in the process of constructing an automatic speech recognition system. In order to train an acoustic model (or, generally, a statistical pattern classifier), a training set of labelled examples is required. In speech recognition, acoustic models are used to classify continuous speech into a set of subword units such as phones, context-dependent subphones etc. Therefore, there is the problem of generating labels for training data in terms of these classes. This can be done by hand, at least in some cases. However, this work is incredibly time-consuming and suffers from consistency problems between the experts. For the many hundreds of hours of speech in corpora like the one described in this chapter, it is inconceivable that a researcher would manually label all the phones. Rather, transcribers are used to obtain word-level transcripts of the utterances, then the word sequence is used to constrain an optimal alignment between existing speech models and the new speech data. This process is called forced alignment.

Forced alignment is a neat way to obtain the training targets for large speech corpora. It may even be preferable to manual labelling because the alignment is applied absolutely consistently by machine.

Typically, the labeling can be refined through iterative alignment and retraining, so-called embedded training. For the purposes of the current project, forced alignment is also a neat way to generate the linguistic data required to test and evaluate research hypotheses. However, just as hand labelling suffers from inconsistency across labellers, forced alignment suffers from inconsistency when inadequately or inappropriately trained acoustic models are used. Since the research questions posed in this project are concerned with the variable pronunciations of word-final consonant clusters in two dialects of French, the aim of the current chapter is to determine which of several types of acoustic models produce, as far as is possible, an optimal alignment between speech models and the recorded speech data that most closely matches an alignment of word-final consonant clusters produced manually by a researcher.

In section 2.1 is described the set of data, using audio and transcription files obtained from the national assemblies of Québec and France, with which were trained four different sets of acoustic models. Next, section 2.2 describes in some detail the process involved in constructing the four acoustic models: a set of monophone acoustic models, a set of word-internal triphone acoustic models, a set of context-dependent, cross-word triphone acoustic models and a set of speaker adapted acoustic models. In section 2.3 the models are evaluated with respect to how closely the results of forced alignment agree with the judgments of a researcher on the variable pronunciations of word-final consonant clusters in two dialects of French. This chapter concludes with an interpretation and discussion of the results of forced alignment.

2.1 Training data

The first stage in the development of an automatic forced aligner is data preparation. Speech data is needed for both training and for testing. The training data is drawn from a corpus of political debates from the national assemblies of Québec and France that occurred in the month of May 2011. The website of the Assemblée nationale du Québec (ANQ, 2011) offers verbatim transcripts of Assembly proceedings through the Journal des débats, available online. The website of the Assemblée nationale de France (ANF, 2011) offers a similar service through the *Journal officiel*. Both audio and transcript files of the debates that occurred in both assemblies for the month of May, 2011 were obtained. This corpus contains approximately 126 hours of audio recordings (66 hours from France and 60 hours from Québec) representing 439 different speakers (105 from Québec, 334 from France). From this, the first 11.5 hours of recordings from each Assembly were selected for a total of 23 hours of data. Before training of the system could begin, the transcripts and audio recordings needed to be properly formatted for use with

HTK.[1] Additionally, a phone set had to be defined and a pronunciation dictionary constructed to cover both training and testing.

2.1.1 Audio files

In order to build the system of forced alignment, the audio files obtained from the national assemblies needed to be prepared for training and testing. Preparation proceeded in four steps. In step one, the complete audio recordings were resampled, following Yuan and Liberman (2008), to a uniform sampling rate of 11025Hz. In step two, using a Praat TextGrid, a pause detector script (Lennes, 2011) attempted to segment the full audio recording into shorter intervals corresponding to pauses in the speech. These estimated interval boundaries were then manually corrected to reflect individual speaker turns. Longer turns were divided into segments of no longer than 100 seconds using natural breaks in the speech. Step three involved indicating on the Praat TextGrid, using the notation “xxx”, those intervals that did not correspond to a clean speaker turn. Situations that required this included periods of applause, background noise of the chamber rising or sitting, the sounds of other members of the assembly heckling or interrupting the speaker, or more than one speaker talking at a time. In step four, all intervals that were not marked as “xxx” were individually extracted and saved as wav files labelled for speaker and turn identification. The end result of these operations was 1,173 individual wav files, sampled at 11025Hz, varying in length from just a few seconds up to 100 seconds, labelled for speaker and turn identification.

2.1.2 Transcription files

To train a system of forced alignment, every file of training data must have an associated phone level transcription. To test the system, every file of training data must have an associated orthographic word level transcription. Both sets of transcriptions were generated from the transcript files obtained through the national assemblies. A great deal of work was involved in cleaning and formatting the set of transcriptions provided by the assemblies. To begin with, the actual transcriptions themselves needed to be verified for accuracy. While the transcription services of the Assemblée nationale du Québec attempts to provide a verbatim transcript of the proceedings, the transcription services of the Assemblée nationale de France provides a paraphrastic transcription. Therefore, the transcripts from the Assemblée nationale de France needed to be extensively re-written to reflect, verbatim, what is present

[1] HTK, the Hidden Markov Model Toolkit, is a portable toolkit for building and manipulating hidden Markov models. Primarily used for speech recognition, HTK consists of a set of library modules and tools available in C source form. The tools provide sophisticated facilities for speech analysis, HMM training, testing and results analysis

on the audio recording. For example, the provided transcription in (1-a) is, in reality, a paraphrase of what the speaker actually said, given in (1-b):

(1) a. "Monsieur le président, monsieur le ministre, chers collègues, ce débat, qui s'inscrit dans la suite de la réforme des collectivités territoriales, nous offre en quelque sorte le spectacle de l'arroseur arrosé".

b. "Monsieur le président euh monsieur le ministre, mes chers collègues, l'ouverture de ce débat, qui fait suite euh la réforme des collectivités territoriales, j'ai l'impression euh de me retrouver dans la fable de l'arroseur arrosé"

Additionally, both sets of transcripts needed to be amended to include the inevitable hesitations, disfluencies, mispronunciations, and non-speech sounds (coughs, laughter, breaths, etc) present in natural speech. In order to reduce the amount of spurious and partial pronunciation entries in the pronunciation dictionary (since the more pronunciation possibilities per word in the dictionary, the higher the likelihood of an erroneous pronunciation being chosen), a decision was made to transcribe hesitations and disfluencies, as much as possible, as pre-existing legitimate words (see Section 2.1.4 for a complete description of the dictionaries). For example, the disfluencies expressed in (2) and (3) were transcribed as (2-a) and (3-a).

(2) ...le rapport des ***act/*** des activités ***du (du)*** commissariat des incendies

a. le rapport des *actes* des activités *du du* commissariat des incendies

(3) ...alors ***maint/*** maintenant sur la question ***sur la question*** des ***(euh)*** *de* la firme

a. alors *mains* maintenant sur la question *sur la question* des *euh de* la firme

Subsequent to these corrections, the text was further altered by

- all text in lower case,

  e.g. Monsieur le ministre→monsieur le ministre

  - so that sentence-initial capitalized words would not require separate dictionary entries

- all punctuation markings removed,

  e.g. !?,;:

  - so that the text reflects only the orthographic words

- all contractions separated,

  e.g. c'est, n'est, qu'est→c' est, n' est, qu' est

    - so that vowel-initial words would not require multiple entries with different contracted articles in the dictionary

- each line of text to represent an individual speaker's turn, with longer turns (> 100 seconds in length) broken at natural pauses,

    - so that each line of text is associated with a single audio file, and

- each speaker's turn labelled with a unique speaker identification mask and turn number.

  e.g. ANQ069-1 = Mme Marois, first turn

With these corrected, cleaned, and formatted texts it was possible to generate two sets of transcriptions: An orthographic word level transcript for testing, and an IPA phone level transcript for training. Both transcripts are in the HTK format known as a *Master Label File*, commonly referred to as an MLF. An MLF is simply a text file containing all transcriptions for every file with every word (for the word level transcript) or phone (for the phone level transcript) written on its own line and every new file indicated by a special prompt. The word level MLF was easily generated from the assembly texts and looks like the following with a unique speaker and turn file identifier as well as an optionally realized short pause (`sp`) segment inserted between words:

```
#!MLF!#
"*/FNQ275-1.lab"
la
sp
philosophie
sp
qui
...
"*/ANQ069-1.lab"
merci
sp
monsieur
sp
le
...
```

The phone level transcript was generated in three steps. In the first place, using the pronunciation dictionary (described below in section 2.1.4), the orthographic text was converted to IPA notation by selecting the first pronunciation candidate (in the dictionary, always the canonical, unmodified pronunciation). Next, all the audio files were carefully listened to and the phonetic transcriptions modified to reflect, as much as possible, the actually realized pronunciations. This included adding all optional schwa vowels (both word internal, as well as word final); removing final consonants from all simplified word final consonant clusters; and inserting all realized liaison consonants. These phonetic transcriptions only reflected the adding or deleting of segments, and do not make reference to possible changes in a segment's quality. While there is evidence to suggest that some variable pronunciations may be gradient, as opposed to categorical (c.f. Saraclar and Khudanpur (2004)), in fact, acoustic and articulatory evidence suggest that the $\emptyset \sim$/ə/ alternation in French is categorical rather than gradient (Bürki et al., 2011a). There are certainly many more modifications that could have been included to make the resulting transcript a true "close" phonetic representation, but for the purposes of the current project, these modifications suffice. Finally, the phone level *Master Label File* was generated as above. In addition to the optionally realized short pause (`sp`) segment inserted between words, there is now also an optionally realized silence (`sil`) segment marking the beginning and ending of each turn.

```
#!MLF!#
"*/FNQ275-1.lab"
sil
l
a
sp
f
i
l
o
z
o
f
i
sp
k
i
...
"*/ANQ069-1.lab"
sil
m
ɛ
ʁ
s
i
sp
m
ə
s
```

```
j
ø
sp
l
ə
...
```

This phone level transcription, used for training the system, informs the HTK tool `HERest` of the exact sequence of phones to be encountered for each audio file. The goal of training then, is to progressively discover the likely boundaries between these phones and to define a statistically based acoustic model for each individual phone to be encountered in connected speech.

2.1.3 Phone set

The phone set used for training the system was generated by extracting a unique list of every phone present in the phone level transcription. The resulting phone set is listed in table 2.1 for consonants and table 2.2 for vowels. In addition to these 39 phones, models for non-speech sounds `{cg} cough`, and `{ns} noise`, along with models for optional silence (`sil`) and short pauses (`sp`) were built. These charts will obviously raise some questions for the phonologically inclined reader and a brief explanation of several of the choices is warranted.

For the consonants, almost all the represented phones will not be considered controversial. The exceptions may be the inclusion of two palatal/velar nasal consonants (ɲ and ŋ) and the placement of the French r-sound as a voiced uvular fricative (/ʁ/). In the native inventory of French there is but one palatal nasal consonant: /ɲ/, as in '*dignité*' [diɲite] or '*signaux*' [siɲo]. The inclusion of /ŋ/ is to accomodate English loan words. In the training data they include words such as: '*computing*' [kɔ̃pytiŋ] and '*gerrymandering*' [ʤɛʁimandəʁiŋ]. While the English loan words in the data are few and the infrequent occurences of /ŋ/ might have been adequately modelled by collapsing these two nasal consonants into a single phone category of /ɲ/, the distinction was maintained for two reasons.

The first reason is that, while constructing the acoustic models and language resources required for forced alignment is a serious investment of time, once they are available they can be an indispensable resource for future research into language. English loan words containing /ŋ/ listed in the Lexique 3.55 pronunciation database (New et al., 2001, 2004) number in the hundreds. Should the forced alignment tool described in this project be made available to the research community, it is likely that it will eventually be required to handle such pronunciations, and as such it makes good sense to create both categories of dorsal nasal consonants now.

The second reason has to do with the fact that it is not at all certain the phonetic realizations of these two phones are either (a) stable or (b) similar to each other. Grevisse and Goosse (2011) suggest that English loan words ending in “ing” may sometimes be pronounced as [n], [ɲ] , or even [ɛ̃g][2]. There is also a high likelihood that /ɲ/ is frequently realized as [nj]. If either of these is correct, it makes sense to retain two categories for these nasal consonants and allow context dependent phone models to capture any variation.

With respect to the classification of the French r-sound as a voiced, uvular fricative, there is no doubt that the French r-sound, depending on a number of not completely understood factors, comes with a wide variety of phonetic realizations. It can be heard as a trill, either apical [r] or uvular [ʀ]; as a voiceless or voiced fricative, either velar [x,ɣ] or uvular [χ,ʁ]; and also as an approximant [ʁ̞] (c.f. Milne (2011a)). Nevertheless, with the exception of the apical trill, which is becoming much less frequently pronounced in any case, at least in Québec (Sankoff and Blondeau, 2007), all posterior variants share some important acoustic charateristics. They are all relatively sonorous with formant-like peaks of energy below $3000Hz$, with somewhat higher centres of gravity (owing to the presence of frication noise) for the fricatives (Milne, 2011a) which led to the selection of a single category with which to model this phone. This discussion is not meant to resolve which phonetic variant of the French rhotic is more appropriate (voiced or voiceless, trill or fricative), but rather to justify that the acoustic properties of all variants suggest that a single phone category will suffice to adequately model this phoneme.

For the vowels, at least three potential concerns should be addressed: The separation of the low back vowels into two phone categories (/a/ vs /ɑ/); the potentially slightly different vowel inventories of France and Québec; and the inclusion of /ə/ as a phone, when in all accounts this French vowel is phonetically similar to [œ] or [ø].

While the phonemic distinction between front /a/ and back /ɑ/ may not be completely maintained in all speakers of French from France (Tranel, 1987), it is clear for speakers from Québec. In particular, Dumas (1986) claims two contexts that favour back /ɑ/: In final, accented position (e.g. ‘*États*’ [etɑ] vs ‘*États Unis*’ [etazyni]) and when the vowel is lengthened, whether etymologically (e.g. ‘*haste*’>‘*hâte*’ [ɑt]) or via the effect of a lengthening consonant /z, ʁ, v, ʒ/ (e.g. ‘*gaz*’ [gɑz], ‘*barre*’ [bɑʁ]). Since occurrences of both [a] and [ɑ] were expected to be relatively frequent in the data (even if more so in the data from Québec), it was decided to maintain two phone categories for the low vowels.

In addition to the potential /a/ vs /ɑ/ difference between the vowel inventories of France and Québec, other differences exist. For example, the distinction between /ɛ/ and /ɛː/, which occurs in minimal pairs such as ‘*mettre*’ [mɛtʁ] vs ‘*maître*’ [mɛːtʁ], is not maintained to an equal degree in both France and Québec. In France, the contrast is typically not maintained and, if it is, it would be essentially by

[2]Although this is a pronunciation guide and not a controlled study.

vowel length. In Québec, this contrast is salient and quite stable and, importantly, it corresponds to clear differences in both vowel length and quality. There are also other differences between the vowel inventories in Québec and France. Québec French has two series of high vowels, either one long and one short, or one tense and one lax, depending on the analysis. When it occurs in a closed syllable, but not before one of the consonants that condition vowel lengethening (i.e. /ʁ, ʒ, z, v/), a high, tense vowel will surface as lax as in the difference between '*petit*' [pəti] and '*petite*' [pətɪt] (Dalton, 2011). The vowel quality of nasal vowels is also different. For a good description of the vowel inventory of Québec, see Côté (2012), while a good description of the vowel inventories of other varieties of French, including from France, can be found in Coveney (2001, pp. 69–137). Mainly for practical reasons, these differences were not exploited during the creation of the phone set. In the first place, the starting point for the majority of the pronunciations was the online resource Lexique 3.55 (New et al., 2001, 2004) which is based on France pronunciation patterns. Identifying all the potential variations in vowels between the two dialects in the corpus was beyond the scope of this project. In the second place, the frequency with which some of these differences might be expected to occur in the corpus was, while unknown, likely to be too small to provide adequate tokens for training. Finally, these differences in the vowel inventories of France and Québec were not expected to negatively impact the ability of a system of forced alignment at recognizing the variable pronunciations of word-final consonant clusters.

Irrespective of whether these assumptions were correct, these potential differences between the vowel inventories of France and Québec do not stand to figure in the analysis of word-final consonant clusters. This is not the case for the difference between a schwa vowel in France and a schwa vowel in Québec. Entire library shelves are devoted to the study of the French schwa and this project only deals with a relatively small part of the schwa phenomenon. Nevertheless, some principled decisions were required when situating the schwa vowel in the phone set. First, it is near universally assumed that its special phonological behaviour, namely its instability, warrants considering it as a distinct phoneme in French. All pronunciation resources encode the schwa vowel with its own unique character and for the phone set /ə/ was selected. This distinguishes it from the stable vowels /œ/ and /ø/. Second, while the schwa vowel in Standard French (i.e. similar to the Northern France variety examined here) alternates between [œ] and [ø], and speakers from Québec invariably realize schwa as [œ] (Séguin, 2010; Martin, 1998), regardless of the dialect, schwa is realized as a front, rounded vowel (Côté, 2008). Third, regardless of the rate or style of speech, optional schwa vowels are likely to be significantly shorter in duration when compared with other (stable) vowels and especially when compared with /œ/ or /ø/, which surprisingly tend to be realized with longer durations than the other stable vowels (Adda-Decker and Snoeren, 2011). Therefore, because it requires its own unique phone, is a front, rounded vowel in both dialects, and is

	Labial	Dental	Palato-alveolar	Palatal plain	Palatal lab.	Velar plain	Velar lab.	Uvular
Nasal	m	n		ɲ		ŋ		
Plosive	p b	t d				k g		
Fricative	f v	s z	ʃ ʒ					ʁ
Approximant				j	ɥ		w	
Lateral		l						
Affricate			ʧ ʤ					

Table 2.1: IPA chart: Consonant phones in training data

		Front unrounded	Front rounded	Central	Back
Close		i	y		u
Close-mid	oral	e	ø	ə	o
Open-mid	oral	ɛ	œ		ɔ
Open-mid	nasal	ɛ̃	œ̃		ɔ̃
Open	nasal				ɑ̃
Open	oral	a			ɑ

Table 2.2: IPA chart: Vowel phones in training data

likely to be, at least, shorter in duration than the other front, rounded vowels the phone set contains a single phone model for schwa, designated by /ə/.

2.1.4 Pronunciation dictionaries

The pronunciation dictionary used for the project was created using a combination of three strategies. Since the list of required words along with their corresponding pronunciations was much too large to be created by hand, a unique word list of all the words present in the corpus was generated from the word level transcript. Using this list of words, the first strategy was to obtain as many of the pronunciations as possible from an existing pronunciation dictionary. For this, Lexique 3.55 (New et al., 2001, 2004) was used. Lexique is a database that provides >135,000 French words including orthographic and phonemic representations, syllabification, parts of speech, gender and number, frequency, and associated lemmas. This information is stored in tables that can be downloaded or searched online. Lexique is an open database to which everyone is encouraged to participate. The list of $10,767$ unique words was compared with the list provided by Lexique and generated the majority of required pronunciations. A second strategy to find pronunciations for those words in the corpus which were not present in Lexique was employed. For this, the online Larousse dictionary[3] which contains more than $250,000$ words and phrases and helpfully provides an IPA pronunciation was used. The final strategy, for those words not in either Lexique or Larousse, was to compare with words having a similar orthography or to consult with native

[3]www.larousse.com/en/dictionaries/french-english/

speakers of French. The majority of these last words were proper nouns (e.g. place names: “Abitibi”) or acronyms (e.g. PIB for “produit intérieur brut”).

Once the dictionary had a canonical, unmodified pronunciation for every word in the corpus, alternative pronunciations for many words were required. Since an objective is to investigate the variable pronunciations of word final consonant clusters, each word ending in a phonemic consonant cluster would need at least three pronunciations: The unmodified form; a form with an epenthesized final schwa vowel; and a form with the final consonant cluster simplified by deleting the final consonant. Côté (2004a) provides a list, adapted from Dell (1995), of 83 attested word final consonant clusters in French. Using this list of word final consonant clusters, the dictionary was searched for words ending in one of these clusters, and appropriate alternative pronunciations were inserted. For example, the frequently occuring word *“ministre”* has four possible pronunciations:

full cluster	ministʁ
deletion, 1st iteration	minist
deletion, 2nd iteration	minis
inserted schwa	ministʁə

Finally, the pronunciations captured in the phone level transcription (i.e. the phonetic transcription that was manually modified to reflect actual realizations of optional schwa, liaison consonants, and consonant cluster simplification) were concatenated with the dictionary of canonical pronunciations to produce candidate pronunciations reflecting all realized liaison consonants as well as optional schwa vowels (both word internal, as well as word final). The final operation was to generate the pronunciation dictionary in the format expected by HTK (the speech recognition engine used by the aligner). HTK expects a dictionary in the format of

```
ortho [outsym] [probability] P1 P2 P3 ...
```

`ortho` is the orthographic entry containing no white space. `[outsym]` is the optional preferred output. When this is not specified, the orthographic form is the chosen output. `[probability]` is an optional integer from 0.0 to 1.0 indicating the probability of the pronunciation form being realized. `P1 P2 P3 ...` are the list of phones used in the pronunciation. These must be separated by whitespace. Additional entries for non-speech sounds include
`{br}{cg}{lg}{ls}{ns} sil & sp` for *breath, cough, laugh, long silence, noise, silence, and short pause* respectively. The final dictionary used for this task contains 13,176 entries. With this pronunciation dictionary and the phone set, along with audio files and corresponding word and phone level transcriptions, it is possible to train an automatic forced aligner.

2.2 Acoustic models

In this section, the progressive cycle of training and testing several iterations of acoustic models used in automatic forced alignment is described. First described is the creation of a well trained set of 8-mixture Gaussian monophone HMMs (mono). Next, the creation of a well trained set of context dependent 8-mixture Gaussian word-internal triphone HMMs (wintri). Third, the creation of a well trained set of context dependent 8-mixture Gaussian cross-word triphone HMMs (cross). Finally, the creation of a set of speaker adapted, context dependent 8-mixture Gaussian cross-word triphone HMMs (sat).

2.2.1 Monophone models

The monophone system uses continuous density 8-mixture Gaussian monophone models with 5 states, 3 of which are emitting states with associated output probability distributions. The raw speech wave forms of the training data have been parameterized into a sequence of 39 feature vectors. Mel Frequency Cepstral Coefficients (MFCC), which are derived from FFT-based log spectra, have been used. A filter bank of 26 channels was employed to extract 39 feature vectors: 12 MFCC plus an energy component and delta and acceleration coefficients. There is one HMM for each of the phones listed in the phone set. In this simple model estimation, the sequence of parameter vectors is assumed to form an exact representation of the speech waveform on the basis that for the duration covered by a single vector (10ms), the speech waveform is, for practical purposes, stationary. It is assumed that the sequence of observed speech vectors (each 10ms observation) corresponding to each phone is generated by a Hidden Markov Model as shown in figure 2.1. As explained in the HTKBook-3.4.1,

> A Markov model is a finite state machine which changes state once every time unit and each time t that a state j is entered, a speech vector o_t is generated from the probability density $b_j(o_t)$. Furthermore, the transition from state i to state j is also probabilistic and is governed by the discrete probability a_{ij}. (p4)

Ouput distributions are represented by Gaussian Mixture Densities with each observation vector at time t split into four S independent data streams o_{st} of the basic parameter vector (the MFCC), first (delta) and second (acceleration) difference coefficients and log energy. The formula for computing $b_j(o_t)$ is given by 2.1.

$$b_j(o_t) = \prod_{s=1}^{S} \left[\sum_{m=1}^{M_s} c_{jsm} \mathcal{N}(o_{st}; \mu_{jsm}, \Sigma_{jsm}) \right]^{\gamma_s} \tag{2.1}$$

Where M_s is the number of mixture components in stream s, c_{jsm} is the weight of the m'th component, and $\mathcal{N}(o; \mu, \Sigma)$ is a multivariate Gaussian with mean vector μ and covariance matrix Σ,

$$\mathcal{N}(o; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n \mid \Sigma \mid}} e^{\frac{1}{2}(o-\mu)^T \Sigma^{-1}(o-\mu)} \tag{2.2}$$

where n is the dimensionality of o.

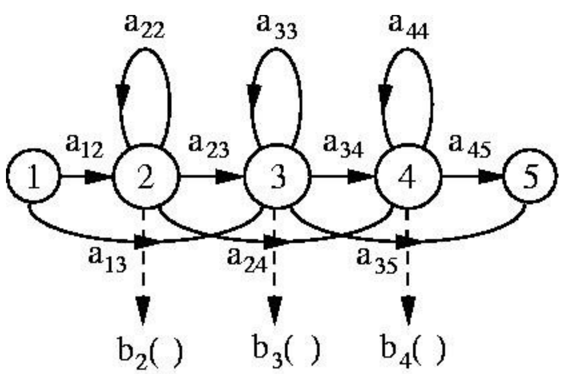

Figure 2.1: Simple Left-Right HMM

With this approach, a HMM definition is created for every phone listed in the phone set. Each definition describes:

- a model name (the phone symbol)
- type of observation vector (MFCC_0_D_A)
- number and width of each data stream (4 streams)
- number of states (5 states, 3 emitting)
- for each emitting state and each stream
  - means and covariances
  - mixture component weights
- transition matrix (5 rows and 5 columns with each row summing to 1 except for the final row, which sums to 0 since no transitions are allowed out of the final state)

Initially, a single prototype definition is constructed manually by setting all the means and covariances to 0. Next, the entire set of training files is scanned and a global speech mean and variance is computed and all of the Gaussians in the prototype definition are set to have the same mean and variance. This prototype definition is then copied, once for every phone in the phone set, and a *Master Macro File* is created containing an identical HMM definition for every required phone. The role of training, then, is to more accurately estimate the individual parameters of each phone.

Training Monophone HMM

Training involves finding more accurate parameters for each HMM definition and is accomplished by applying the Baum-Welch re-estimation formula. The Baum-Welch re-estimation formula attempts to estimate the means and variances of an HMM in which each state output distribution is a single component Gaussian,

$$b_j(o_t) = \frac{1}{\sqrt{(2\pi)^n \mid \Sigma_j \mid}} e^{-\frac{1}{2}(o_t-\mu_j)^\top \Sigma_j^{-1}(o_t-\mu_j)} \tag{2.3}$$

Since the full likelihood of each observation sequence is based on the summation of all possible state sequences, each observation vector o_t contributes to the computation of the maximum likelihood parameter values for each state j. Therefore, the Baum-Welch re-estimation formula assigns each observation to every state in proportion to the probability of the model being in that state when the vector was observed. Thus, the Baum-Welch re-estimation formula for the means and covariances of an HMM are given in 2.4 and 2.5 where $L_j(t)$ denotes the probability of being in state j at time t.

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} L_j(t) o_t}{\sum_{t=1}^{T} L_j(t)} \tag{2.4}$$

and

$$\hat{\Sigma}_j = \frac{\sum_{t=1}^{T} L_j(t)(o_t - \mu_j)(o_t - \mu_j)^\top}{\sum_{t=1}^{T} L_j(t)} \tag{2.5}$$

$L_j(t)$, the probability of being in state j at time t, is calculated using a *Forward-Backward* algorithm. The forward joint probability $\alpha_j(t)$ for some model M with N states is defined as

$$\alpha_j(t) = P(o_1, ..., o_t, x(t) = j \mid M) \tag{2.6}$$

and can be calculated recursively by

$$\alpha_j(t) = \left[\sum_{i=2}^{N-1} \alpha_i(t-1) a_{ij} \right] b_j(o_t). \tag{2.7}$$

The backward conditional probability $\beta_j(t)$ is defined as

$$\beta_j(t) = P(o_{t+1}, ..., o_T \mid x(t) = j, M) \tag{2.8}$$

and can be calculated recursively by

$$\beta_i(t) = \sum_{j=2}^{N-1} a_{ij} b_j(o_{t+1}) \beta_j(t+1). \tag{2.9}$$

Since the forward probability is a joint probability while the backward probability is a conditional probability, the probability of state occupation can be determined by taking the product of the two probabilities. Therefore,

$$\begin{aligned} L_j(t) = \alpha_j(t)\beta_j(t) &= P(x(t) = j \mid O, M) \\ &= \frac{P(O, x(t) = j \mid M)}{P(O \mid M)} \\ &= \frac{1}{P}\alpha_j(t)\beta_j(t) \end{aligned} \tag{2.10}$$

where $P = P(O \mid M)$.

Using these calculations, HMM parameter re-estimation using the Baum-Welch algorithm is performed by successive invocations of the HTK tool `HERest` by updating the parameters for all phone models encountered in the training file in parallel. Each call to `HERest` performs the following 7 steps and this was repeated 8 times to achieve the necessary convergence:

1. For every parameter vector/matrix requiring re-estimation, allocate storage for the numerator and denominator summations of the form illustrated by equations 2.4 and 2.5. The storage locations are called *accumulators*.

2. Get the first training file.

3. Construct a composite HMM by joining in sequence the HMMs corresponding to the symbol transcription of the current training file.

4. Calculate the forward and backward probabilities for the composite HMM.

5. Use the forward and backward probabilities to compute the probabilities of state occupation at each time frame (10ms) and update the accumulators.

6. Repeat from 2 until all training files have been processed.

7. Use the accumulators to calculate new parameter estimates for all of the HMM definitions.

Mixture incrementing

Multiple mixture component HMMs are constructed by increasing the number of components in a mixture by a process known as *mixture splitting*. Mixture splitting works by repeatedly splitting the heaviest mixture component until the required number of components is obtained. The actual split is performed by copying the mixture component, dividing the weights of both copies by 2, and perturbing the means by ± 0.2 standard deviations. Incrementing mixture components is done in stages by incrementing by a factor of 1, re-estimating the model parameters, and then incrementing again by 1, re-estimating, and so forth. This allows recognition performance to be monitored to find the optimum. The models were incrementally mixed from 1 to 8, successively incrementing the number of mixtures by a factor of 1 with 4 invocations of `HERest` to update the model parameters in between each mixture incrementation. With these re-estimated HMM definitions, the system using continuous density 8-mixture Gaussian monophone models was ready for evaluation.

2.2.2 Word-internal triphone models

The HMM definitions constructed in section 2.2.1 are what is known as *context-independent* models. There is a single definition for each phone listed in the phone set that is, in effect, a composite definition of every realization of that phone in the training data irrespective of the phonetic context in which it appeared. The reality of continuous speech, however, is that many phones will be realized with often significant variation in the parameter vectors depending, in large part, on the neighbouring phones. Failing to take account of this variation can have detrimental effects on the performance of a speech recognition system. Therefore, the first improvement to be made on the forced alignment system is to make the HMM definitions *context-dependent*. This will have the effect of augmenting the phone set, and subsequently, the number of HMM definitions, but also allow for more accurate re-estimation of model parameters by taking into account the effects of neighbouring phones. Creating context dependent models involves converting monophone transcriptions to triphone transcriptions and creating a set of triphone models by copying the monophone models and re-estimating. Following this, similar acoustic states of these triphones can be tied to ensure that all state distributions can be robustly estimated.

Triphone transcriptions

The phone level transcript MLF was converted to a triphone level transcript MLF by re-writing each word level sequence of phones into a corresponding sequence of triphones. For example, the training file transcript from section 2.1.2

`sil m ɛ ʁ s i sp m ə s j ø sp`

would become

`sil m+ɛ ɛ-ʁ+s ʁ-s+i s-i sp m+ə ə-s+j s-j+ø j+ø sp`

This form of triphone transcription is known as *word-internal* since the word boundary symbols `sil`, and `sp` block the addition of context. Therefore, some biphones will also be generated as contexts at word boundaries will sometimes only include two phones. This triphone level transcript MLF was then used to construct a new triphone set, to replace the original monophone phone set, and includes every bi- and triphone for which there is at least one example in the training data.

This brings up a short point about *data sparsity.* Data sparsity occurs when there are not enough examples of a phone or triphone in the training data. Given the initial list of 39 phones, excluding non-speech sounds, there are theoretically $59,319$ possible triphones. The list of attested triphones in the training data contains $6,127$ examples, and of those, a good many occur fewer than 3 times. Very poor estimates of parameter vectors would be obtained for these infrequently occurring triphones. Additionally, there are cases where the transition parameters do not vary significantly with acoustic context, but nevertheless need to be estimated accurately. These problems of data insufficiency will affect the output distributions and a solution is required. The solution effected by HTK is to tie model states together by tying only those parameters which have little effect on discrimination.

Tied-state triphones

A tied state is one in which the individual transition matrices have been replaced by a reference to a matrix shared by both models. When re-estimating tied parameters, the data which would have been used for each of the original untied parameters is pooled so that a more reliable estimate can be obtained. The first step in creating tied-state triphones is to indiscriminately clone all states of the original monophone models across all states of their corresponding triphone models. That is, for each existing monophone `p`, copy the transition matrix into every triphone `x-p+y`. The result is a set of triphone HMMs with all triphones in a phone set sharing the same transition matrix. The second step in creating tied-state triphones involves clustering similar acoustic states together so that those state clusters can then be tied.

State clustering is accomplished by constructing a decision tree which attempts to find those contexts which make the largest difference to the acoustics and which should therefore distinguish clusters. A decision tree is a binary tree in which a yes/no question about the right and left contexts of each triphone is attached to each node. The questions should progress from wide, general classifications (such as consonant, vowel, nasal, etc) to specific instances of each phone. It is advantageous for the full set of questions to describe every possible context which can influence the acoustic realization of a phone, including any linguistic or phonetic classification which may be relevant, since any questions determined to be irrelevant to the data will be ignored. The set of questions used to construct the decision tree was created by first generating a *Phone by Feature* matrix with a column for each phone in the phone set and a row for each phonologically based feature determined to be potentially relevant to the data. Cells in the matrix were populated as either a + or − value for the feature, or 0, indicating no value to be specificed for the feature. There were 21 phonologically based features chosen. The complete Phone by Feature matrix is given in Appendix A.1. In order to generate the questions themselves, this feature matrix was iterated over to extract classes of phones that could be described by various combinations of features. The first set of questions, encompassing wide, general classifications, was composed by extracting those sets of phones that could be described by a single feature. For example, for the phone set, the initial 7 questions are similar to the following

```
QS  "R_Consonantal" {*+b,*+d,*+g,*+f,*+k,*+p,*+s,*+t,*+v,...,*+z}
QS  "R_NonConsonantal" {*+a,*+,*+ɥ,*+i,*+j,*+ã,*+o,*+ɔ̃,*+ɛ̃,...,*+œ̃}
QS  "R_Sonorant" {*+ã,*+œ̃,*+ɲ,*+ɑ,*+œ,*+ɔ,*+ə,*+e,*+ɛ,...,*+a}
QS  "R_NonSonorant" {*+ʃ,*+b,*+d,*+g,*+f,*+k,*+p,*+s,*+ʒ,...,*+t}
QS  "R_Syllabic" {*+a,*+ɑ,*+e,*+i,*+ã,*+o,*+ɔ̃,*+ɛ̃,*+œ̃,...,*+œ}
QS  "R_NonSyllabic" {*+ʁ,*+ʃ,*+ʒ,*+ʤ,*+ɲ,*+b,*+ɥ,*+d,*+g,...,*+f}
QS  "R_Nasal" {*+ɲ,*+m,*+n,*+ŋ,*+ɛ̃,*+ã,*+œ̃,*+ɔ̃}
```

and define the question "Is the right context a consonant, sonorant, etc?".

More specific classes were defined by extracting those sets of phones that could be described by combinations of features of length 2,3,4, and 5. This was accomplished iteratively by constructing a list of permutations of the features of length, for example, 2, and then updating the list of questions describing sets of phones with the following constraints:

- Any null set was excluded. For example, the set of phones defined as being simultaneously + and − Consonantal is null and would not be included in the list of questions.

- Any singleton set was excluded. Since the most specific questions refer to individual phones, any combination of features that describe a set of just one member would not be included in the list of questions.

- Any duplicate set was not included. Any set of phones described by n features that was previously described by $n-1$ features would not be included in the list of questions.

At the end of round 5 of these permutations, only 2 unique, non-null, non-singleton sets of phones were described and, therefore, no further rounds of permutations of length 6 or greater were examined. The final list of questions to be used in constructing a decision tree contains 774 questions about the right and left contexts of phones and includes wide, general questions about sets of phones described by a single feature to more specific questions about sets of phones described by up to 5 features, to the most specific questions about individual phones.

The HTK tool `HHEd` uses this list of questions to find those contexts which make the largest difference to the acoustics and which should therefore distinguish clusters which can then be tied so that the data which would have been used for each of the original untied parameters can be pooled and a more reliable estimate of the HMM definition obtained. At the outset, all states for a given phone model are pooled to form a single cluster and are placed at the root node of a tree. Trees are built using a top-down sequential optimisation process. Each question in the question set is used to split this pool into two sets. Splitting any pool into two will increase the log likelihood since it provides twice as many parameters to model the same amount of data. The question is then found which maximizes this increase in log likelihood and is selected for the first branch of the tree. The process is then repeated until the increase in log likelihood achievable by any question at any node is less than a specificed threshold. Depending on each answer, the pool of states is successively split and this continues until the states have trickled down to leaf nodes. All states in the same leaf node are then tied. A simple example of the results of this decision tree-based clustering and state tying is shown in figure 2.2 for the center states of the phone œ̃ in the data. In the tree, the question which forms the first branch of the tree, `R_NonBoundary?`, splits the data into word-final vs word-internal observations of œ̃. This question was chosen since it maximized the increase in log likelihood when it was used to split the single pool of all observations into two separate pools. No further applications of questions resulted in an increase in log likelihood above a minimum threshold for discrimination and therefore all the center states from word-internal observations of œ̃ are clustered and tied (the rightmost leaf at the bottom of the tree). The question which forms the second branch of the tree, `L_ConsonantalVoice?`, splits the pool into observations of word-final œ̃ preceded by a voiced vs voiceless consonant. The question at the third branch of the tree, `L_NonNasal?`, splits the remaining pool into observations of word-final œ̃ preceded by a nasal vs non-nasal phone.

The end result of state clustering and tying is a set of HMM definitions where data from model states that are acoustically similar have been pooled into a single, shared set of parameters that can better estimate the HMM definitions. A further 8 rounds of re-estimation following the steps outlined in 2.2.1

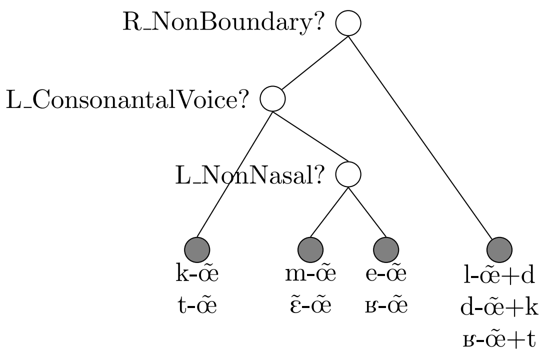

Figure 2.2: Decision tree-based state tying results for œ̃

produced a set of context-dependent, word-internal triphone HMM models. Similar to the monophone models, these models were incrementally mixed from 1 to 8, successively incrementing the number of mixtures by a factor of 1 with 4 invocations of `HERest` to update the model parameters in between each mixture incrementation. With these re-estimated HMM definitions, a system using continuous density 8-mixture Gaussian context-dependent word-internal triphone models was ready for evaluation.

2.2.3 Cross-word triphone models

The context-dependent, word-internal triphone HMMs were expected to be an improvement over the context-independent monophone HMMs constructed at the end of section 2.2.1. This improvement is due to the better estimation of model parameters provided by taking into account the effects of neighbouring phones. However, a further reality of continuous speech is that the effects of context will often spread across the word boundary. This is especially true in this system given the particular interest in the variable pronunciations of word-final consonant clusters which, when followed by a vowel intial word, will often be phonetically realized as an onset. This cross-word effect is also prevalent when a liaison consonant appears. Therefore, the system could be further improved by taking into account these cross-word contextual effects. Constructing a context-dependent, cross-word triphone system is similar to the steps followed in section 2.2.2, except the triphone level transciptions are generated without allowing the `sp` segment to block the addition of context. In this way, the only biphones created are those that appear at the beginning or ending of a training file. The training file example in section 2.1.2 now becomes

```
sil m+ɛ ɛ-ʁ+s ʁ-s+i s-i+m   i-m+ə m-ə+s ə-s+j s-j+ø j-ø+l
```

A new set of cross-word triphone models is created by copying the monophone models and re-estimating. Following this, similar acoustic states of these new triphones can be tied to ensure that all state distri-

Model	*Type*	*Example*	*Trained on*
Mono	Monophones	[t], [i], [d], [ə]	French
Wintri	Word-internal triphones	[t-i+t], [-d+ə]	French
Cross	Cross-word triphones	[t-i+t], [i-t+d], [t-d+ə], [d-ə+]	French
SAT	Speaker adapted Cross	S_1[t-i+t], [i-t+d], [t-d+ə], [d-ə+]	French
		S_2[t-i+t], [i-t+d], [t-d+ə], [d-ə+]	
P2FA	Monophones	[T], [IY], [D], [AH]	English

Table 2.3: Summary of acoustic models used.

butions can be robustly estimated. Finally, these models were incrementally mixed from 1 to 8 and a system using continuous density 8-mixture Gaussian context-dependent cross-word triphone models was ready for evaluation.

2.2.4 Speaker adaptive training

Adaptation techniques are applied to well trained speaker independent model sets to enable them to better model the characteristics of particular speakers. Supervised adaptation, where the true transcription of the data is known, is performed using maximum likelihood linear transformations to estimate a series of transforms that reduces the mismatch between the current model set and the adaptation data. In adaptive training, constrained maximum likelihood linear regression (CMLLR) adaptation transforms are used to represent speaker differences during training so that a "neutral" canonical model can be trained. Speaker adaptive training using CMLLR estimates a set of linear transformations, one for each speaker in the training data, that shift the feature vectors in the initial system so that each state in the HMM system is more likely to generate the adaptation data. This improves performance of forced alignment when there are multiple speakers. Speaker adaptation was performed by generating a CMLLR transform for each speaker in the training data. These transforms were then used to update and re-estimate (through four rounds of training using `HERest`) a neutral canonical set of HMM from the context-dependent cross-word triphone models. In an effort to further improve performance, separate canonical models were estimated for each dialect. Updates of the transforms were interleaved with updates of the canonical model set after every four rounds of training and this process was repeated a total of four times. A final set of speaker dependent, continuous density, 8-mixture Gaussian, context-dependent, cross-word triphone models was then ready for evaluation.

2.3 Measuring agreement

This project is exploring the variable pronunciations of word-final consonant clusters. In this chapter, the research question being asked is: How likely is it that the judgments of a forced aligner will agree with a researcher's judgments on the variable pronunciations of word-final consonant clusters in two dialects of French? This section begins by exploring which set of 5 acoustic models produces a set of pronunciation judgments that most closely agrees with the manually judged pronunciations of word-final consonant clusters in a set of data. Following the identification of the best set of acoustic models to use for forced alignment (if the goal is to select a set of models that best mimics the pronunciation judgments of a human listener) the accuracy of the selected set of acoustic models is evaluated with respect to several of the variables that will be used in an accompanying corpus analysis (presented in Chapter 3) in order to determine if the accuracy of the model set is equal under different conditions.

The term *accuracy* will be used in a very specific way throughout this section. When describing the *accuracy* of a particular instrument of measurement, it is generally assumed to be the case that an event has actually occurred x times and the tool used to detect this event identified y occurrences, thus giving a sense of $y \div x$ as a measure of the accuracy of the tool. However, it is notoriously difficult to systematically and consistently identify the actual sequence of phones in an utterance of recorded speech. Several studies have suggested that even trained experts, working with the same set of recordings may differ by as much as 20% in their opinions of which phones were realized and where the temporal boundaries indicating the start and end of individual phones lie (Saraclar and Khudanpur, 2004; Goldman, 2011). Therefore, lacking any independent evidence to determine whether a particular sequence of phones actually did occur in an utterance, it is difficult to measure the accuracy of a system of forced alignment in this usual sense.

Instead, the approach adopted here interprets the term *accuracy* in terms of reference to a set of manually labelled pronunciations. That is, a pronunciation has been manually identified to have occurred x times and the forced aligner has identified y occurrences, thus giving a sense of $y \div x$ as a measure of agreement between the two methods. Interpreted this way, the *accuracy* of the system of forced alignment described in this chapter is really a measure of how often the pronunciation judgments agree between the two methods of coding.

For this project, a single judge (this researcher) used exclusively perceptual means to identify the variable pronunciations of the word-final consonant clusters in these data. This was accomplished during the data preparation phase while producing a phonetically accurate transcription for training of the acoustic models. Because a strict binary decision was required in order to produce such a transcription, no ambiguous tokens were permitted. For example, in the case of word-final schwa the decision could only be schwa present or absent. In light of the findings from Bürki et al. (2011c), this likely means that some of the perceived occurrences of word-final schwa were much shorter in duration than other schwa vowels and may have contained weaker acoustic cues. It is also likely that many short and/or acoustically weak potential schwa vowels were perceived as absent. In a ternary structure (i.e. present/absent/ambiguous) common to most decision tasks, a significant number of these occurrences might have been identified instead as ambiguous.

It must not be forgotten that the set of manually labelled pronunciations is not infallible. It is certain that included in that set of manually labelled pronunciations are errors of the type where the researcher erroneously indicated a pronunciation different from the one they perceived. It is also likely that when the acoustic information was ambiguous or indeterminate the researcher may have inconsistently selected from the various pronunciations available. It may also have been the case that, since listener's perceptions of pronunciations rely on more than just acoustic information, the manually labelled pronunciations may often have reflected what the listener *expected* to hear, given his exposure to and experience with the language. All of this suggests that no system of forced alignment, no matter the training protocol, could be expected to agree 100% of the time with a set of manually labelled transcriptions. What is more important to detemine is whether it is likely that the results of forced alignment will be sufficiently similar to the results of manual labelling under the conditions expected to be found in the speech data to be segmented.

Test data sets

The testing data is drawn from the same corpus described in section 2.1. The total data set contained 23 hours of recordings of continuous speech, which includes approximately 11.5 hours from each assembly. All the models had been trained on 18 hours of speech data (9 hours per dialect). While in theory it is preferable to train and test an automatic forced aligner on separate sets of data, in practice it is often the case that testing is performed on the training data itself. It was decided to test values of agreement on the variable pronunciations of word final consonant clusters on the complete 18 hour data set it was trained on for several reasons. First, as explained in section 2.1.2, the system was trained on a set of International Phonetic Alphabet (IPA) transcriptions, but tested on an orthographic transcription. The orthographic transcription allowed for multiple candidate pronunciations and as such, the system had not truly "seen" the data it was recognizing. Second, since the goal is to determine how likely it is that the judgments of an automatic forced aligner will agree with a researcher's judgments on the variable pronunciations of word-final consonant clusters and this question is being evaluated with respect to several conditions, it was important to have sufficient data in order to draw meaningful conclusions. A smaller test set of speech data was not likely to generate sufficient data. Third, the final system had several opportunities for further evaluation on novel speech data when the audio recordings obtained from the PFC were incorporated into the corpus for the analysis presented in Chapter 3.

While it was not explored in the current project, the PFC corpus provides an excellent opportunity for further evaluating the accuracy of a system of forced alignment. Several minutes of recorded audio from each participant (approximately 6 minutes of the conversations, 3 minutes per conversation, and all of the read text exercise, typically 2 minutes) has been coded for the presence or absence of both schwa and simplification following a word-final consonant cluster. The complete guidelines describing the coding protocol followed by researchers in the PFC group are available on the PFC site under the publications link in PFC Bulletin 1. In brief, each site where an optional schwa is possible is evaluated and coded according to four criteria which results in a four digit numeric code inserted into the text at the optional schwa site indicating the following:

- **First digit:** Schwa is absent (0), present (1) or uncertain (2);
- **Second digit:** Syllabic position is either a monosyllable (1), or the initial (2), medial (3) or final (4) syllable of a polysyllabic word, or was metathesized (5);
- **Third digit:** The lefthand context is a vowel (1), a consonant (2), a pause (3), an uncertain schwa (4) or a simplified consonant cluster (5);
- **Fourth digit:** The righthand context is a vowel (1), a consonant (2), a strong (3) or a weak (4) pause.

As an example, the PFC codings for the variable pronunciations in (4), ignoring the optional schwa vowel in *"de"*, would be:

(4) *". . . titres de gloire. . . "*
"claims to fame"

a. [titʁ də glwaʁ]→titres[0422] de gloire
b. [tit də glwaʁ]→titres[0452] de gloire
c. [titʁə də glwaʁ]→titres[1422] de gloire

Determining the accuracy of a system of forced alignment using these additional data would give further confidence in the system, however it was not feasible to use these previously coded PFC data as an initial test set. In the first place, the volume of available test material provided by the PFC data was too small. The coding had not yet been completed for all of the investigations representing the Québec dialect and the $6-8$ minutes per participant of data that had been coded (for either dialect) provided too few examples of word-final consonant clusters. In the second place, the tests of the system of forced alignment (described below in Section 2.3.3) sought to determine the accuracy of forced alignment according to several of the variables to be used during the corpus analysis (see Chapter 3). In addition to differences according to dialect, differences according to speech style and following context were to be evaluated. In these coded portions of the PFC corpus, not enough variety of individual consonant clusters was present and the word-final consonant clusters that were present tended to occur in only one of three possible following contexts (preconsonantal, prevocalic or prepausal).

Forced alignment

Forced alignment is performed by the HTK tool `HVite`. Given a recognition network specifying what is allowed to be spoken and how each word can be pronounced, a set of HMMs, and an unknown utterance, the probability of any path through the network can be computed. The task of `HVite` is to find those paths which are most likely. The most likely path is found using the *Token Passing Model*. This model is an extension of the Viterbi algorithm, which computes the maximum likelihood state sequence recursively by

$$\varphi_j(t) = \max_i\{\varphi_i(t-1) + log(a_{ij})\} + log(b_j(o_t)) \tag{2.11}$$

The Viterbi algorithm attempts to find the best path through a matrix where the vertical dimension represents the states of the HMM and the horizontal dimension represent the frames of speech (each 10ms). The log probability of any path through this matrix is computed by summing the log transition probabilities and the log output probabilities along that path. At each time t, each partial path $\varphi_i(t)$ is known for all states i, and equation 2.11 can be used to compute $\varphi_j(t)$, extending the partial paths by one time frame. The Token Passing Model makes this concept of a state alignment *path explicit*. For each state j of a HMM at time t there is a single moveable token which contains the partial log probability $\varphi_j(t)$, representing a partial match between the observation sequence o_1 to o_t and the model, with the constraint that the model is in state j at time t. The token passing algorithm is then executed at each time frame t and updates equation 2.11 such that

- a copy of every token in state i is passed to all connecting states j, incrementing the log probability of the copy by $log[a_{ij}] + log[b_j(o(t))]$,
- the tokens in every state are examined and all but the token with the highest log probability are discarded.

The history of a token's route through the network is recorded. When a token is propagated from the exit state of a model to the entry state of another, that transition represents a potential model boundary. A *Link Record* is generated that stores the identity of the model from which the token has just emerged and the current value of the token's link. Once all the speech has been processed, the Link Record attached to the link with the highest log probability can be traced back to give the matching sequence of models.

During recognition of an unknown utterance with an associated word network that describes all possible sequences of words, the result of these operations is the most likely sequence of words. During forced alignment, the word network is constructed from the given transcription and the sequence of words is known ahead of time. The role of the Viterbi and Token Passing algorithms therefore is constrained to determine, given a dictionary with multiple pronunciations listed for words, the most likely sequence of phones (i.e. the most likely pronunciation of each word). Furthermore, if the dictionary lists but a single pronunciation per word, the process of forced alignment is further constrained to determine just the most likely boundaries between phones (since the sequences of both words and phones is known ahead of time).

The results of forced alignment output by HTK contain

- the start and end times of each phone model[4],
- the name of the phone model and
- the word from which the phone was retrieved.

An example of the result of forced alignment for the example file from section 2.1.2 looks like

```
#!MLF!#
"*/ANQ069-1.lab"
0 1500000 sil silence
1500000 1800000 m merci
1800000 2500000 ɛ
2500000 3000000 ʁ
3000000 4600000 s
4600000 5300000 i
5300000 5300000 sp sp
5300000 5800000 m monsieur
5800000 6400000 ə
6400000 7000000 s
7000000 7300000 j
7300000 7600000 ø
7600000 7600000 sp sp
7600000 7900000 l le
7900000 8200000 ə
8200000 8200000 sp sp
```

[4]HTK outputs times in reference to the frame number where every frame represent 10ms of speech. For historical reasons, because it used to be necessary to express time as an integer, HTK uses 100-nanosecond units. Since our sampling rate was 11025Hz, as identified by Yuan and Liberman (2008), frame references were converted to seconds by $(t \div 10000000.0 + 0.0125) \times (11000.0 \div 11025.0)$ where t is the frame number.

In this example, the start and end times for the 3 `sp` models are identical. This indicates that the optional short pause inserted between words was not realized in this sequence. From these MLFs, the words and pronunciations output by each of the models were retrieved. All words ending in a final consonant cluster were identified and extracted, along with the pronunciations output by each of the models, into a data frame for comparison with the manually judged pronunciations.

In the following section (2.3.1) are described the results of performing forced alignment using each of the models (p2fa, mono, wintri, cross and sat) on the complete 18 hour test set using a pronunciation dictionary containing multiple candidate pronunciations for many words, including at least 3 candidate pronunciations for every word with a final consonant cluster.

2.3.1 Selecting the appropriate model set

Each set of models was used to perform forced alignment on the 18 hour test set of speech recordings. Additionally, in order to obtain a 'baseline' value of the likelihood that a forced aligner will agree with a researcher's judgments, forced alignment was performed with an aligner used previously. The earlier aligner, the Penn Phonetics Lab Forced Aligner (P2FA), was originally constructed to work with the English language (Yuan and Liberman, 2008). It had been modified to work with the French language (Milne, 2011a). By changing the labels of the HMMs to French phones, but leaving intact the actual HMMs themselves, the P2FA, modified to work with French, contains monophone HMM models based on English language training data that have been labeled with French phones to work with French transcriptions and a French pronunciation dictionary. During some preliminary testing of this modified P2FA, Milne (2011b) concluded that this aligner agreed with the judgments of human researchers on the realizations of word final schwa in 74.3% of examined cases; was consistent in both dialects under consideration (Québec and France); and was consistent across three contexts (pre-vocalic, pre-consonantal, pre-pausal). However, the vast majority of errors (96.9%) made by the modified P2FA were of the sort where the aligner "perceived" a schwa to be realized, while the human researchers had not. It is possible that the relatively low accuracy rate, along with the over-representation of schwa realizations, could have been due to either the differences in acoustic cues between the French and English schwa vowels (phonetically, [ø/œ] vs [ə]) or the possibility that elements of word final consonant release could have been misinterpreted by the aligner as vocalic. Therefore it was hypothesized that improvements in performance could be obtained by constructing an automatic forced aligner built with true French acoustic models.

The hypothesis that the accuracy rate of forced alignment could be improved by constructing an

HMM model set trained and tested on French speech data has been confirmed by the results presented below in section 2.3.2. Section 2.2 described the construction of four sets of acoustic HMM models. The first, called Mono, is a set of single mixture, monophone HMM definitions. The second, called Wintri, is a set of single mixture, context-dependent word-internal triphone HMM definitions. The third, called Cross, is a set of 8-mixture, context-dependent cross-word triphone HMM definitions. The fourth, called SAT, is a set of speaker dependent, continuous density, 8-mixture Gaussian, context-dependent, cross-word triphone models. Each set of HMM models was used to perform forced alignment on an 18 hour corpus of speech data drawn from the national assemblies of both Québec and France. Forced alignment was performed in conjunction with a pronunciation dictionary containing at least three candidate pronunciations for every word that ended in a consonant cluster. The three pronunciations included a pronunciation where the word-final consonant cluster was unmodified. That is, the realized pronunciation included all and only the consonants in the cluster. This pronunciation candidate is represented as "Full". A second candidate pronunciation included a pronunciation where the word-final consonant cluster was realized with an epenthetic schwa vowel. This pronunciation candidate is represented as "Schwa". The third pronunciation candidate included a pronunciation where the final consonant of the cluster was deleted, or not phonetically realized. This pronunciation candidate is represented as "Simplify".

(5) $/C_1C_2/ \rightarrow$

a. **Full**: $[C_1C_2]$

b. **Schwa**: $[C_1C_2\text{ə}]$

c. **Simplify**: $[C_1]$

With the word network defined by the transcription provided and a pronunciation dictionary containing multiple candidate pronunciations for words in the network, the role of forced alignment was to determine the most likely pronunciation candidate given the acoustic data. The pronunciation decisions that resulted from forced alignment using each of the HMM model sets were collected into a data frame and a value for a dependent (outcome) variable **agree** was calculated based on whether the most likely pronunciation, as decided by forced alignment, agreed with the pronunciation as perceived by a human researcher.

	Frequencies$_{n=counts}$
Schwa	$0.303_{n=2,419}$
Simplify	$0.282_{n=2,253}$
Full	$0.415_{n=3,311}$

Table 2.4: Frequencies and counts of manually judged pronunciations of word final consonant clusters

	Québec	France	Totals
Schwa	$0.169_{n=639}$	$0.423_{n=1,780}$	$2,419$
Simplify	$0.424_{n=1,602}$	$0.155_{n=651}$	$2,253$
Full	$0.407_{n=1,536}$	$0.422_{n=1,775}$	$3,311$

Table 2.5: Frequencies and counts of manually judged pronunciations by dialect

2.3.2 Agreement on variable pronunciations

The 18 hour test set contains $7,983$ instances of word-final consonant clusters from both dialects of French. Tables 2.4 and 2.5 display the counts and frequencies of each of the 3 possible pronunciations overall and by dialect as judged manually. These frequencies form the results against which each model will be evaluated.

To begin, as shown by Table 2.6, the set of speaker dependent models (SAT) is, in general, the most accurate of the 5 sets of models being evaluated. Nearly 87% of the automatically judged variable pronunciations of word-final consonant clusters by this set of models agreed with the manually judged pronunciations. This is a dramatic improvement over the baseline value of just 62% of correct responses automatically judged by the modified P2FA model set.

	model				
agree	P2FA	Mono	Wintri	Cross	SAT
0	0.376	0.289	0.182	0.245	0.134
1	0.624	0.711	0.818	0.755	0.866

Table 2.6: Proportion of correct responses by model set

Looking closer at the proportion of correct responses according to dialect, the set of speaker dependent models (SAT) also appears to be the most accurate of the 5 sets of models within each dialect. Table 2.7 displays the proportion of automatically judged pronunciations in agreement with manually judged pronunciations according to both model set and dialect. The SAT model set agreed with the manually judged pronunciations of word-final consonant clusters approximately 86% of the time in the dialect from France and 87% of the time in the dialect from Québec. The difference between dialects is not significant ($\chi^2(1) = 2.85, p = 0.09$) while the odds of agreeing with the manually judged pronunciations in each dialect compared with the baseline reference values of agreement obtained by using the modified P2FA model set increase by a factor of 3.38 for the dialect from France, and by a factor of 4.53 for the dialect from Québec.

	model					
	agree	P2FA	Mono	Wintri	Cross	SAT
FR	0	0.356	0.272	0.166	0.215	0.141
	1	0.644	0.728	0.834	0.785	0.859
QC	0	0.399	0.308	0.199	0.278	0.128
	1	0.601	0.692	0.801	0.722	0.872

Table 2.7: Proportion of correct responses by dialect and model set

Verifiying whether any of the 5 model sets are better at identifying a particular pronunciation involved examining the proportion of automatically judged pronunciations in agreement with manually judged pronunciations for each of Schwa, Simplify and Full.

Correctly identifying pronounced schwa: $/C_1C_2/\rightarrow[C_1C_2\text{ə}]$

Table 2.8 displays the proportion of automatically judged schwa pronunciations in agreement with manually judged schwa pronunciations according to both model set and dialect. The model set with highest proportion of correct responses was the SAT model set. The SAT model set agreed with the manually judged schwa pronunciations of word-final consonant clusters approximately 93% of the time in the dialect from France and 88% of the time in the dialect from Québec. The difference between dialects was significant ($\chi^2(1) = 16.60, p < 0.05$). The odds of the SAT model agreeing with the manually judged schwa pronunciations in the dialect from France are 13.48 while the same odds in the dialect from Québec are 7.27. This results in an odds ratio of 1.85 suggesting it is almost twice as likely that the SAT models will agree with the manual judgments with respect to a schwa pronunciation in the dialect from France as compared with the dialect from Québec. The odds of agreeing with the manually judged pronunciations in each dialect compared with the baseline reference values of agreement obtained by using the modified P2FA model set increase by a factor of 3.51 for the dialect from France, and by a factor of 2.09 for the dialect from Québec.

	model					
	agree	P2FA	Mono	Wintri	Cross	SAT
FR	0	0.207	0.212	0.107	0.134	0.069
	1	0.793	0.788	0.893	0.866	0.931
QC	0	0.224	0.155	0.146	0.182	0.121
	1	0.776	0.845	0.854	0.818	0.879

Table 2.8: Proportion of correct schwa responses by dialect and model set

Correctly identifying a simplified consonant cluster: $/C_1C_2/\rightarrow[C_1]$

Table 2.9 displays the proportion of automatically judged simplified pronunciations in agreement with manually judged simplified pronunciations according to both model set and dialect. The model set with highest proportion of correct responses was the SAT model set. The SAT model set agreed with the manually judged schwa pronunciations of word-final consonant clusters 86% of the time in the dialect from France and nearly 90% of the time in the dialect from Québec. The difference between dialects was significant ($\chi^2(1) = 7.27, p < 0.05$). The odds of the SAT model agreeing with the manually judged simplified pronunciations in the dialect from France are 6.13 while the same odds in the dialect from Québec are 8.94. This results in an odds ratio of 1.46 suggesting it is almost one and a half times more likely that the SAT models will agree with the manual judgments with respect to a simplified pronunciation in the dialect from Québec as compared with the dialect from France. The odds of agreeing with the manually judged pronunciations in each dialect compared with the baseline reference values of agreement obtained by using the modified P2FA model set increase by a factor of 1.87 for the dialect from France, and by a factor of 3.01 for the dialect from Québec.

		model				
	agree	P2FA	Mono	Wintri	Cross	SAT
FR	0	0.233	0.181	0.183	0.241	0.140
	1	0.767	0.819	0.817	0.759	0.860
QC	0	0.252	0.204	0.124	0.212	0.101
	1	0.748	0.796	0.876	0.788	0.899

Table 2.9: Proportion of correct simplified responses by dialect and model set

Correctly identifying unmodified consonant clusters: $/C_1C_2/\rightarrow[C_1C_2]$

Table 2.10 displays the proportion of automatically judged unmodifed ('Full') pronunciations in agreement with manually judged unmodified pronunciations according to both model set and dialect. The model set with highest proportion of correct responses was the SAT model set. The SAT model set agreed with the manually judged unmodified pronunciations of word-final consonant clusters approximately 79% of the time in the dialect from France and 84% of the time in the dialect from Québec. The difference between dialects was significant ($\chi^2(1) = 15.70, p < 0.05$). The odds of the SAT model agreeing with the manually judged unmodified pronunciations in the dialect from France are 3.71 while the same odds in the dialect from Québec are 5.31. This results in an odds ratio of 1.43 suggesting it is almost one and a half times more likely that the SAT models will agree with the manual judgments with respect to an unmodified pronunciation in the dialect from Québec as compared with the dialect from France. The odds of agreeing with the manually judged pronunciations in each dialect compared with the baseline reference values of agreement obtained by using the modified P2FA model set increase by a factor of 4.52 for the dialect from France, and by a factor of 8.85 for the dialect from Québec.

		model				
	agree	P2FA	Mono	Wintri	Cross	SAT
FR	0	0.550	0.367	0.220	0.288	0.212
	1	0.450	0.633	0.780	0.712	0.788
QC	0	0.625	0.480	0.299	0.387	0.158
	1	0.375	0.520	0.701	0.613	0.842

Table 2.10: Proportion of correct unmodified responses by dialect and model set

Justification for SAT as the appropriate model set

In all cases that were examined, the set of speaker dependent, continuous density, 8-mixture Gaussian, context-dependent, cross-word triphone models (SAT) proved to be the most accurate at agreeing with the manually judged variable pronunciations of word-final consonant clusters in these data. Depending on the dialect (France or Québec) this set of acoustic models agreed with the manually judged pronunciations of word-final consonant clusters between 93.1% $\sim$ 87.9% for Schwa pronunciations, 86.0% $\sim$ 89.9% for Simplify pronunciations and 78.8% $\sim$ 84.2% for Full pronunciations. There was a significant increase in the odds of automatically judged pronunciations being in agreement with manually judged pronunciations when using the SAT set of acoustic models as compared to the baseline reference values obtained when using the modified P2FA set of models. In general, the odds of being in agreement increased by a factor of 3.88 indicating that it is nearly 4 times more likely that the automatically judged variable pronunciations of word-final consonant clusters will be in agreement with manually judged pronunciations using the SAT set of acoustic models instead of the modified P2FA set of acoustic models.

2.3.3 Agreement of forced alignment under several conditions

It is known that the performance of a speech recognition system is likely to be degraded under certain conditions. For example, words with highly variable pronunciations are difficult to model within automatic speech recognition technologies and cause many errors in text-to-speech (Lanchantin et al., 2008), in word recognition (Adda-Decker, 2007), or in phonetic alignment (de Mareüil, 2007; Bürki et al., 2008). Words in connected speech exhibit considerable variation (Johnson, 2004; Zwicky, 1972). Among the potential sources of variability are phonological processes, which lead to dramatic changes in the phonetic form of the words; segments are added, deleted, or replaced by others (Bürki et al., 2011a). de Mareüil et al. (2005), using 9 hours of French TV show recordings that were both manually transcribed and automatically force aligned observed that it was often difficult to distinguish hesitations, which all corresponded to "euh" or its variants, from the pronunciation of a final schwa. Relevant to the current project, speech style has been shown to have an effect on recognition. Changing from a reading to a spontaneous style of speech is associated with an increase in word error rates (Weintraub et al., 1996). Additionally, speech rate has been shown to have a significant effect on recognition (Siegler and Stern, 1995). When speech rate exceeds a threshold, recognition accuracy drops. Given that the SAT model set has been idenitified as being the most accurate of the 5 models just evaluated, it may be advantageous to examine whether the accuracy of forced alignment using this set of acoustic models suffers under various conditions. The speaker adapted model set will be used to force-align a large corpus of natural language spoken French (presented in Chapter 3) in order to determine the variable pronunciations of word-final consonant clusters. The corpus contains examples of both read and spontaneous styles of speech as well as a varied distribution of speech rates. Therefore, it will be wise to determine whether the accuracy of this set of acoustic models is comparable across the variables that will be used in that analysis. The results of the following tests will help to give confidence to the results of forced alignment on the additional data, beyond this test set, included in the corpus.

The conditions under which levels of agreement between manual and automatic judgment of the variable pronunciations of word-final consonant clusters will be compared are similar to the factors evaluated in the analysis presented in Chapter 3. They include speech style, following context and speech rate. The same 18 hour corpus of speech data drawn from the national assemblies of both Québec and France will be used for testing. Visual inspection of the video of the proceedings provided by the two national assemblies allowed for the identification of when a speaker was actively reading from prepared notes or when the speaker was instead speaking without the aid of a text. This distinction formed the basis for the representation of two styles of speech: Reading or spontaneous. The following context was identified from the transcription and includes whether the word following the word-final consonant cluster began with a consonant (including a realized liaison consonant), a vowel, or whether the consonant cluster was instead followed by a pause. Speech rate was calculated as the number of phones present in an utterance divided by the length in seconds of the utterance to produce a value expressed in phones per second. While a commonly used measure of speech rate has been the number of words divided by the length of the utterance, it has been suggested that calculating speech rate this way may be unsatisfactory "due to the effects of differences in the structure of words and syllables. Words can be mono- or poly-syllabic, and syllables themselves can vary greatly in the number and type of segments occurring in onset, peak and coda positions." (Campbell, 1988, p. 93). Word rate may also be unreliable because of the indeterminacy of any pause durations between words (Siegler and Stern, 1995).

Agreement according to speech style

The test set contains 7983 observations of word-final consonant clusters. Each observation has been identified as being drawn from either a reading (speaking with the use of prepared text) or a spontaneous (speaking without the aid of notes) style of speech. Table 2.11 displays the amount of data available for each speech style for each dialect.

	Dialect		
Style	FR	QC	Sum
Reading	2747	1653	4400
Spontaneous	1459	2124	3583
Sum	4206	3777	7983

Table 2.11: Observations of word-final consonant clusters by speech style and dialect

Examining the proportion of word-final consonant cluster pronunciations that are in agreement between manual and automatic methods of coding, it appears to be the case that the accuracy of forced alignment using the speaker adapted model set does not significantly degrade when changing from a reading to a spontaneous speech style. Table 2.12, which displays the frequency and proportion of agreement between the two methods of coding suggests that levels of agreement are approximately the same between the two styles of speech in both dialects.

Dialect	Agree	Reading	Spontaneous	Sum
QC	0	249 (15.06%)	233 (10.97%)	482
	1	1404 (84.94%)	1891 (89.03%)	3295
FR	0	372 (13.54%)	219 (15.01%)	591
	1	2375 (86.46%)	1240 (84.99%)	3615
Total		4400	3583	7983

Table 2.12: Frequency and proportion of agreement according to style

Looking closer at whether either of the two pronunciations of schwa or simplification appear to be affected by changes in speech style, Tables 2.13 (for the dialect from Québec) and 2.14 (for the dialect from France) suggest that in both dialects the levels of agreement between manual and automatic coding of word-final consonant clusters pronounced with a final schwa vowel are not dependent on speech style. In fact, with obtained values of $\chi^2(1) = 1.36, p > 0.05$ for the dialect from Québec and $\chi^2(1) = 0.53, p > 0.05$ for the dialect from France the null hypothesis cannot be rejected in both cases: Levels of agreement between manual and automatic coding of word-final consonant clusters pronounced with a final schwa vowel do not change according to speech style.

Tables 2.15 (for the dialect from Québec) and 2.16 (for the dialect from France) suggest that in the dialect from Québec the levels of agreement between manual and automatic coding of word-final consonant clusters pronounced with a deleted final consonant (simplified) may be dependent on speech

	Observed			Expected		
	Agree			Agree		
Style	Yes	No	Total	Yes	No	Total
Reading	273	43	316	277.80	38.20	316
Spontaneous	287	34	321	282.20	38.80	321
Total	560	77	637	560	77	637

Table 2.13: Observed and expected frequencies of agreement for schwa as a function of style for Québec

	Observed			Expected		
	Agree			Agree		
Style	Yes	No	Total	Yes	No	Total
Reading	1118	79	1197	1114.33	82.67	1197
Spontaneous	540	44	584	543.67	40.33	584
Total	1658	123	1781	1658	123	1781

Table 2.14: Observed and expected frequencies of agreement for schwa as a function of style for France

style, while in the dialect from France the levels of agreement are not. With a value of $\chi^2(1) = 4.75, p < 0.05$ for the dialect from Québec, these data suggest that levels of agreement were actually higher in the spontaneous style of speech than in the reading style. For the dialect from France, $\chi^2(1) = 0.34, p > 0.05$ indicates no significant difference in levels of agreement between the two styles. For the dialect from France the null hypothesis cannot be rejected: Levels of agreement between manual and automatic coding of word-final consonant clusters pronounced as simplified do not change according to speech style.

	Observed			Expected		
	Agree			Agree		
Style	Yes	No	Total	Yes	No	Total
Reading	552	76	628	564.81	63.19	628
Spontaneous	887	85	972	874.19	97.81	972
Total	1439	161	1600	1439	161	1600

Table 2.15: Observed and expected frequencies of agreement for simplify as a function of style for Québec

Judging from these results, it is safe to assume that, notwithstanding the small effect of speech style in the dialect from Québec when identifying word-final consonant clusters that have been simplified, changing from a reading to a spontaneous style of speech is not likely to have a significant impact on the results of forced alignment.

With levels of agreement ranging from 85% to 89%, these results on the effect of style compare favorably with similar studies conducted on corpora of Dutch (Kuipers and van Donselaar, 1997; Schuppler

	Observed			Expected		
	Agree			Agree		
Style	Yes	No	Total	Yes	No	Total
Reading	348	52	400	343.91	56.09	400
Spontaneous	210	39	249	214.09	34.91	249
Total	558	91	649	558	91	649

Table 2.16: Observed and expected frequencies of agreement for simplify as a function of style for France

et al., 2011), German (Kipp et al., 1997), and English (Cucchiarini and Binnenpoorte, 2002). An additional metric commonly used to asses inter-transcriber reliability (which this is, with one transcriber being electronic) is Cohen's Kappa (κ) (Carletta, 1996).[5] This measure, which is assumed to be more reliable than simple percent agreement calculation, attempts to take into account the agreement occurring by chance. The equation for κ is given by 2.12:

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{2.12}$$

In 2.12, $Pr(a)$ is the relative observed agreement between raters, and $Pr(e)$ is the hypothetical probability of chance agreement. $\kappa = 1$ when the raters are in perfect agreement, and 0 if there is no agreement other than what might be expected due to chance.

Values for κ of 0.79, 0.80, 0.85, and 0.86 were returned as measures of inter-transcriber reliability for pronunciations of schwa in reading and spontaneous data and for pronunciations of simplification in reading and spontaneous data, respectively. According to the frequently used scale developed by Landis and Koch (1977), this suggests almost perfect agreement between automatic and manual methods of coding when evaluated according to different styles of speech.

Agreement according to following context

Each of the 7983 observations of word-final consonant clusters has been identified as being followed by either a consonant initial word ('C'), a vowel initial word ('V') or a pause ('P'). There were a few instances where the following context happened to be a non-speech sound (NS), such as a cough or laughter. These are included in the following table for illustration purposes, but are not evaluated further. Table 2.17 displays the amount of data available for each following context for each dialect.

Examining the proportion of word-final consonant cluster pronunciations that are in agreement between manual and automatic methods of coding, it appears to be the case that the accuracy of forced alignment using the speaker adapted model might degrade according to following context. Table 2.18,

[5] Thanks to one of my anonymous reviewers for suggesting the use of the additional performance metric.

	Dialect		
Context	FR	QC	Sum
C	2541	2247	4788
V	587	789	1376
P	1043	711	1754
NS	35	30	65
Sum	4206	3777	7983

Table 2.17: Observations of word-final consonant clusters by following context and dialect

which displays the frequency and proportion of agreement between the two methods of coding suggests that levels of agreement were highest when the word-final consonant cluster was followed by a consonant initial word and lowest when followed by a pause. This was true of both dialects.

Agree		Consonant	Vowel	Pause	Sum
QC	0	230 (10.24%)	108 (13.69%)	138 (19.41%)	476
	1	2017 (89.76%)	681 (86.31%)	573 (80.59%)	3271
FR	0	281 (11.06%)	90 (15.33%)	209 (20.04%)	580
	1	2260 (88.94%)	497 (84.67%)	834 (79.96%)	3591
Sum		4788	1376	1754	7918

Table 2.18: Frequency and proportion of agreement according to context

Looking closer at whether either of the two pronunciations of schwa or simplification appear to be affected by changes in following context, Tables 2.19 (for the dialect from Québec) and 2.20 (for the dialect from France) suggest that in both dialects the levels of agreement between manual and automatic coding of word-final consonant clusters pronounced with a final schwa vowel are dependent on following context. In fact, with obtained values of $\chi^2(2) = 42.51, p < 0.05$ for the dialect from Québec and $\chi^2(2) = 182.20, p < 0.05$ for the dialect from France the null hypothesis must be rejected in both cases: Levels of agreement between manual and automatic coding of word-final consonant clusters pronounced with a final schwa vowel do change according to following context. In both dialects the proportion of observed schwa pronunciations that were in agreement between the two methods of coding was higher than expected when the consonant cluster was preconsonantal and lower than expected when either prevocalic or prepausal.

Tables 2.21 (for the dialect from Québec) and 2.22 (for the dialect from France) suggest that in both dialects the levels of agreement between manual and automatic coding of simplified word-final consonant clusters are dependent on following context. In fact, with obtained values of $\chi^2(2) = 107.26, p < 0.05$ for the dialect from Québec and $\chi^2(2) = 77.19, p < 0.05$ for the dialect from France the null hypothesis must be rejected in both cases: Levels of agreement between manual and automatic coding of simplified word-final consonant clusters do change according to following context. In both dialects the proportion of observed simplified pronunciations that were in agreement between the two methods of coding was

	Observed			Expected		
	Agree			Agree		
Context	Yes	No	Total	Yes	No	Total
C	430	35	465	408.61	56.38	465
V	25	15	40	35.14	4.85	40
P	103	27	130	114.23	15.76	130
Total	558	77	635	558	77	635

Table 2.19: Observed and expected frequencies for agreement for schwa as a function of context for Québec

	Observed			Expected		
	Agree			Agree		
Context	Yes	No	Total	Yes	No	Total
C	1276	41	1317	1228.50	88.49	1317
V	29	24	53	49.43	3.56	53
P	347	54	401	374.05	26.94	401
Total	1652	119	1771	1652	119	1771

Table 2.20: Observed and expected frequencies for agreement for schwa as a function of context for France

higher than expected when the consonant cluster was preconsonantal and lower than expected when either prevocalic or prepausal.

	Observed			Expected		
	Agree			Agree		
Context	Yes	No	Total	Yes	No	Total
C	1006	49	1055	949.76	105.23	1055
V	171	56	227	204.35	22.64	227
P	249	53	302	271.87	30.12	302
Total	1426	158	1584	1426	158	1584

Table 2.21: Observed and expected frequencies for agreement for simplify as a function of context for Québec

Cohen's Kappa calculated on these data confirm that agreement levels between the two methods of coding significantly differ according to whether or not the consonant cluster is preconsonantal or prevocalic. For example, when followed by a consonant-initial word, $\kappa = 0.90$ for pronunciations of simplified clusters (almost perfect agreement) and 0.78 for pronunciations of schwa (substantial agreement). However, when followed by a vowel-initial word, $\kappa = 0.73$ for simplification (substantial agreement) and just 0.48 for pronunciations of schwa (moderate agreement).

These results examining the effect of following context on the levels of agreement between manual and

	Observed			Expected		
	Agree			Agree		
Context	Yes	No	Total	Yes	No	Total
C	440	34	474	407.13	66.87	474
V	25	20	45	38.65	6.35	45
P	89	37	126	108.22	17.78	126
Total	554	91	645	554	91	645

Table 2.22: Observed and expected frequencies for agreement for simplify as a function of context for France

automatic methods of coding for the variable pronunciations of word-final consonant clusters indicate that for both dialects, the two methods of coding are most likely to agree when the consonant cluster is preconsonantal and least likely to agree when the consonant cluster is prevocalic. Levels of agreement were exceptionally high for the preconsonantal context. When identifying a word-final consonant cluster in a preconsonantal context that had been manually judged as having a final schwa vowel, the automatic forced aligner also judged these same pronunciations as schwa nearly 93% of the time for the dialect from Québec and almost 97% of the time for the dialect from France. When identifying manually judged simplified word-final consonant clusters in a preconsonantal context, the automatic forced aligner agreed approximately 95% of the time for the dialect from Québec and 93% of the time for the dialect from France. The levels of agreement between the two methods of coding were not nearly as high for word-final consonant clusters in either a prepausal or prevocalic context. Prepausally, levels of agreement were approximately 80% and 87%, respectively for the dialects from Québec and France, when correctly identifying manually judged schwa pronunciations, and 83% and 71% when identifying manually judged word-final consonant cluster simplification. In the prevocalic context, the results of the Chi-square tests of independence suggest that the proportion of observations that were in agreement between the two methods of coding were significantly lower than would be expected if the different frequencies of observations between the three contexts were due simply to chance. This was particularly true for the dialect from France where levels of agreement were just 55% for both the schwa and simplify pronunciations.

Agreement according to speech rate

Speech rate is a continuous variable that measured the number of realized phones present during an utterance and expresses this measurement as a value of phones per second. Table 2.23 lists some descriptive statistics for speech rate for each dialect and Figure 2.3 indicates that it is a normally distributed variable.

Dialect	Min.	1^{st} Qu.	Median	Mean	3^{rd} Qu.	Max.
QC	2.786	11.140	12.960	12.900	14.550	30.770
FR	2.897	11.280	12.840	12.810	14.320	23.750

Table 2.23: Descriptive statistics for speech rate

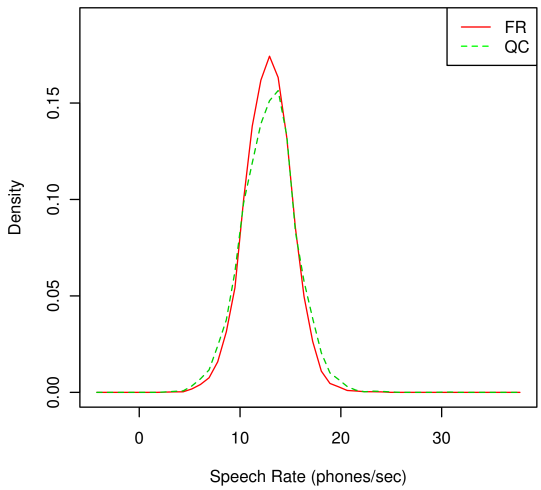

Figure 2.3: Kernel density plot showing normal distribution of values for speech rate

Tables 2.24 (for the dialect from Québec) and 2.25 (for the dialect from France) display the coefficients for two models predicting the response of **agree** as a function of speech rate for word-final consonant clusters manually judged as having a final schwa vowel. In neither dialect does it appear to be the case that speech rate has an effect on the levels of agreement between the two methods of coding, manual or automatic.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.22434	0.56939	2.150	0.0315	*
Rate	0.0647	0.04799	1.348	0.1777	

Table 2.24: Coefficients for model predicting agreement for schwa as a function of speech rate for Québec

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.87402	0.53016	3.535	0.000408	***
Rate	0.05846	0.04237	1.380	0.167648	

Table 2.25: Coefficients for model predicting agreement for schwa as a function of speech rate for France

Tables 2.26 (for the dialect from Québec) and 2.27 (for the dialect from France) display the coefficients for two models predicting the response of **agree** as a function of speech rate for word-final consonant clusters manually judged as being simplified. In neither dialect does it appear to be the case that speech rate has an effect on the levels of agreement between the two methods of coding, manual or automatic.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.80126	0.42075	4.281	$1.86e-05$	***
Rate	0.02925	0.03121	0.937	0.349	

Table 2.26: Coefficients for model predicting agreement for simplify as a function of speech rate for Québec

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.86813	0.65612	4.371	$1.23e-05$	***
Rate	−0.07900	0.04785	−1.651	0.0987	.

Table 2.27: Coefficients for model predicting agreement for simplify as a function of speech rate for France

These results examining the effect of speech rate on the levels of agreement between two methods of coding when judging the variable pronunciations of word-final consonant clusters should be interpreted as meaning that variable rates of speech do not have a significant impact on the levels of agreement between manual and automatic coding. In both dialects (Québec and France) and for both pronunciations (Schwa and Simplify) these models show that the coefficient for `Rate`, which measures the slope of the logistic regression as a function of rates of speech, is not significantly different from 0. These relationships are displayed visually in Figure 2.4. In the figure, the left-hand y-axis plots predicted probability and the right-hand y-axis plots frequency.

2.4 Chapter summary and discussion

This chapter investigated how often, and under what conditions, the process of forced alignment produced judgments on the variable pronunciations of word-final consonant clusters in two dialects of French that were in agreement with the judgments produced by a manual (i.e. human) researcher. The use of forced alignment could be an extremely efficient way for the community of linguists to obtain the data it requires in order to conduct empirical research. With its ability to segment even extremely large amounts of recorded speech in very short periods of time, an accurate forced aligner can shorten the period of time between the inception of a research question and the presentation of results by months or even years. It may even be preferable to manual segmentation because the alignment is applied absolutely consistently by machine. However, just as hand labelling suffers from inconsistency across labellers, forced alignment suffers from inconsistency when inadequately or inappropriately trained acoustic models are

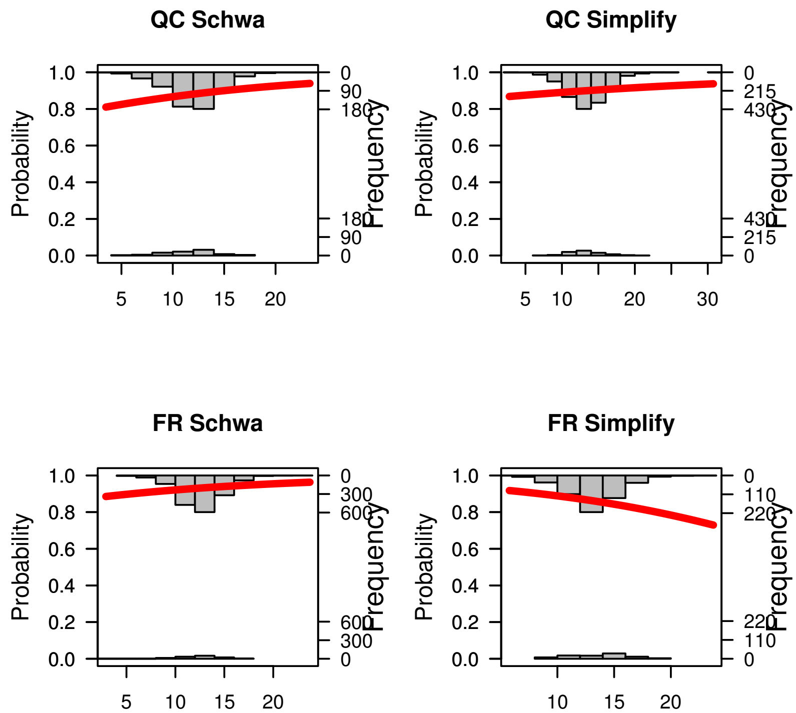

Figure 2.4: Histograms and logistic regressions predicting levels of agreement for schwa and simplify

used. Therefore this chapter sought to answer two specific questions. First, from within a selection of different types of acoustic models, which set was consistently more accurate at producing judgments on the variable pronunciations of word-final consonant clusters that most closely match the judgments of a human listener? Second, under which of a set of several conditions do the levels of agreement between manual and automatic coding significantly differ? The answer to the first question gives a set of acoustic models to be used in forced alignment that will confidently reproduce pronunciation judgments similar to those of a human researcher. The answer to the second question gives a better understanding of the conditions under which the two methods of coding might be expected to disagree, and perhaps why.

The motivation for building and testing a system of automatic forced alignment arose as a solution to the practical problem of how to segment a large amount of previously recorded French speech. Manual segmentation is incredibly time-consuming and for the more than 100 hours of speech to be segmented, it was inconceivable that a single researcher could manually label all the phones. Unfortunately, there did not exist an efficient, reliable, freely available system of forced alignment capable of segmenting recorded French speech. The specific research questions that are investigated in Section 2.3 of this chapter are a natural extension of the idea that, just because the resources required to build a system of forced alignment are freely available and it was possible to build one, how does one know whether or not it is working the way one supposes it is working? Testing of fully fledged automatic speech recognition systems generally involves a metric that describes the system's word error rate, which calculates the number and proportion of missed, inserted or changed words in an utterance. However forced alignment is not concerned with word error rates, since the true transcription of the utterance is provided and therefore words will never be missed, inserted or changed. What can be missed, inserted or changed during the recognition stage of forced alignment are the specific pronunciations of words, specifically words that can have more than one pronunciation. The accuracy of a system of forced alignment can be determined, therefore, by determining how often the actual sequence of phones in an utterance was missed, inserted or changed during forced alignment.

It was mentioned at the outset of this chapter, and will be reinforced here, that the term *accuracy* has been used in a very specific way throughout this chapter. The approach adopted here interprets the term *accuracy* in terms of reference to a set of manually labelled pronunciations. That is, a pronunciation has been manually identified to have occurred x times and the forced aligner has identified y occurrences, thus giving a sense of $y \div x$ as a measure of agreement between the two methods. Interpreted this way, the *accuracy* of the system of forced alignment described in this chapter is really a measure of how often the pronunciation judgments agree between the two methods of coding.

The results presented in Section 2.3 of this chapter suggest that, of the 5 types of acoustic models that were evaluated, the set of speaker dependent, continuous density, 8-mixture Gaussian, context-dependent, cross-word triphone models (SAT) proved to be the most accurate at agreeing with the manually judged variable pronunciations of word-final consonant clusters in these data. With an overall level of agreement (read: *accuracy*) of nearly 87% (85.95% in the dialect from France, 87.24% in the dialect from Québec), this set of acoustic models represents an improvement over the next best performing set of acoustic models (word-internal triphones) by a factor of 1.43, and a significant increase in performance over an initial set of acoustic models, a modified version of the Penn Phonetics Lab Forced Aligner, by a factor of 3.88.

Further examination of the levels of agreement found when using the speaker adapted model set suggest that the accuracy of this system of forced alignment does not suffer from the effects of either speech style or speech rate. It was determined that there was little or no change in the levels of agreement between manual and automatic methods of coding when switching from a read text to a spontaneous style of speech. This was true both for pronunciations of schwa insertion and simplification following a word-final consonant cluster. Furthermore, the levels of agreement within each speech style ranged from 84.94% $\sim$ 89.03%, suggesting that, within each style of speech, this set of acoustic models accurately reflects the set of manually labelled pronunciations.

Neither was there any apparent change in the levels of agreement according to changes in speech rate. Again, this was true both for pronunciations of schwa insertion and simplification following a word-final consonant cluster. The logistic regression models showed that the coefficient for `Rate`, which measured the slope of the logistic regression as a function of rates of speech, was not significantly different from 0 meaning that variable rates of speech did not have a significant impact on the levels of agreement between manual and automatic coding.

When determining whether the level of agreement between the two methods of coding changed according to whether the word-final consonant cluster was in a preconsonantal, prevocalic or prepausal context, some interesting effects were observed. When the consonant cluster was preconsonantal (followed by a consonant intitial word in the absence of a pause), the levels of agreement were exceptionally high. When identifying both schwa insertion and simplification following a word-final consonant cluster, the automatic method of coding agreed with manual judgments between 92.47% $\sim$ 96.89% of the time. When the consonant cluster was prepausal (followed by a non-zero length `sp` phone), the levels of agreement were lower. When identifying either schwa insertion or simplification following a word-final consonant cluster, the results of forced alignment agreed with manual labelling between 70.64% $\sim$ 86.53% of the time. While the extremely high level of accuracy obtained for the preconsonantal context leaves no doubt that this set of models accurately reflects the set of manually labelled pronunciations, the relatively high levels of accuracy obtained for the prepausal context suggest that, more or less, the models accurately reflect the manually labelled pronunciations in this context as well.

This same conclusion was less well supported in the results for the prevocalic context where levels of agreement for schwa insertion or simplification following a word-final consonant cluster varied from a low of 54.72% to a high of 75.33%. In fact, when the word-final consonant cluster was followed by a vowel initial word in the absence of a pause, the proportions of observations that were in agreement for these two pronunciation variants were not much above chance levels in the dialect from France. There are several explanations for why the levels of agreement might have dropped so far just in this context.

The first reason may have something to do with the problem of data sparsity during training of the acoustic models. In the training data there were 1376 examples of a word-final consonant cluster followed by a vowel initial word. Of these, only 320 were judged manually to be either simplified (227) or to have had schwa inserted (93). This may have resulted in poor parameter estimation for some word-final triphones, especially those containing an inserted schwa between the final consonant and the following vowel. Consider the cross-word triphone that would need to be evaluated by the forced aligner when schwa insertion following a word-final consonant cluster occurs in a prevocalic context: C-ə+V.

This is not a frequently occurring word-internal sequence of phones in French. A search of the Lexique database for this sequence returns just 4 words: "re-hygiène", "contre-ut", "re-histoires" and "re-inter". None of these words were present in the training data. Therefore the acoustic parameters re-estimated during training for these triphones are mostly pooled estimates from more generic C-V+V triphones. This is a possible explanation for the drop in the levels of agreement just in this prevocalic context.

A second plausible explanation relies on the fact that it is precisely when a word-final consonant cluster is in a prevocalic context that most previous accounts conclude that both schwa insertion and simplification following a word-final consonant cluster are least likely to occur. Indeed, manual labelling of the training data indicated that more than 3/4 of the word-final consonant clusters in the prevocalic context were pronounced unmodified (i.e. Full). Interestingly, when looking at the observations that were not in agreement (there were 39 and 76 respectively for Schwa and Simplify), 90% of the time the results of forced alignment indicated an unmodified pronunciation (i.e. Full) instead of the inserted schwa or simplified consonant cluster indicated by the manual labelling. This suggests that the results of forced alignment, with respect to the variable pronunciations of word-final consonant clusters in a prevocalic context, might be more in line with what much previous research has suggested is expected to happen.

A third explanation, which may be a contributing factor to inconsistent labelling, involves the possibility that the acoustic information in this context is ambiguous or indeterminate. It could be the case that when a word-final consonant cluster is in a prevocalic context, some elements of the release of the final consonant could be misinterpreted, by either a human listener or an automatic forced aligner, as having elements of a vocalic nature resulting in inconsistently inserted schwa vowels. Alternatively, the onset of the following word-initial vowel may mask or otherwise obscure elements of the final consonant, especially if that consonant is relatively sonorous (as in the case of the frequently occurring consonant clusters ending with a final /ʁ/ or /l/), resulting in inconsistently simplified consonant clusters. In fact, of the observations that were not in agreement when manually judged as being Full in this context (83 of a total of 1011), 60% of the time the forced aligner indicated an inserted schwa and 40% of the time it indicated a simplified consonant cluster.

A fourth possibility exists that concerns the confusion between an inserted schwa vowel and a filled pause. de Mareüil et al. (2005), using 9 hours of French TV show recordings, observed that it was often difficult to distinguish hesitations, which were all transcribed as '*euh*', from the pronunciation of a final schwa. The pronunciation dictionary gave a pronunciation of [ø] for these filled pauses, which shares many acoustic characteristics with [ə]. Given the difficulty in distinguishing between a word-final schwa vowel and a filled pause, it is possible that some of the manually judged final schwa vowels were confused with filled pauses and this contributed to the poor performance of the forced aligner in the prevocalic context.

The corpus used for this investigation (the AssNat corpus) came with transcriptions provided by the national assemblies which were manually adjusted by this researcher to reflect, verbatim, what was said on the recordings. Careful attention was paid to insert a filled pause token '*euh*' whenever it was judged to be appropriate. In these data, there were 57 such occurrences of filled pauses and the accuracy of forced alignment in these cases was quite high at 89.5%. This suggests that the presence of a filled pause did not negatively impact the performance of the forced aligner. However, without knowing whether a word frequency rate of 0.007%(57/7983) for filled pauses is typical for this type of corpus, it is diffcult to estimate how many occurrences of word-final schwa vowels (either manually or automatically identified) might have been misinterpretations of filled pauses.

If any or a combination of these reasons are true, then because the alignment is applied absolutely consistently by machine, the low levels of agreement between the two methods of coding apparent in the prevocalic context might be attributed to inadequate parameter modelling for consonant clusters in a prevocalic context; inconsistent labelling in this context by the manual coder; ambiguous acoustic information; or confusion with filled pauses.

In conclusion, the generally high levels of agreement between the results of manual and automatic identification of the variable pronunciations of word-final consonant clusters should be taken as evidence in favour of the conclusion that the speaker dependent acoustic models built and tested in this chapter produce results that are likely to be similar to the results produced by manual identification. The results of forced alignment are not likely to be affected by changes in speech style or speech rate. While the results of forced alignment may differ from the results of manual labelling according to different following contexts, it may in fact be the case that the results of forced alignment with respect to the variable pronunciations of word-final consonant clusters are actually a better estimate of the actual pronunciation of prevocalic consonant clusters.

Chapter 3

A Corpus Analysis

This chapter examines schwa insertion and simplification following word-final consonant clusters in a large corpus of natural language spoken French containing examples from two dialects and two speech styles. Of interest is knowing the effect of variables, such as speech style, following context, motivation for simplification and speech rate, on the variable pronunciations of word-final consonant clusters in two dialects of French.

This chapter begins with a review of the existing literature on word-final consonant clusters in French, which is divided into two areas of study. On the one hand are reports of consonant cluster simplification. The lexicon of French contains many words that terminate in a sequence of two or more consonants (a consonant cluster) and it is often the case that the final consonant of the cluster is deleted, subject to certain conditions; a process known as consonant cluster simplification. On the other hand are reports of word-final $\emptyset \sim$/ə/ alternation. It has long been noted that, in French, a word-final consonant cluster, especially when followed by a pause or consonant-initial word, will often be pronounced with a final schwa vowel; a process known as schwa insertion. The review of these two bodies of literature will provide information about what previous authors have observed with respect to the distribution of the variable pronunciations of word-final consonant clusters in French. The literature review concludes with a brief description of the pilot study undertaken by Milne and Côté (2009) which was an attempt at comparing both consonant cluster simplification and schwa insertion in a single corpus containing two dialects of French.

Following the literature review, the corpus used for the current study is introduced along with a description of the data set. The set of variables, including following context, motivation for simplification, speech style and speech rate, are then explored with respect to their individual effects on both simplfication and schwa insertion following a word-final consonant cluster. Following this, these variables are combined into a single statistical model for each dialect predicting the responses of either simplification or schwa insertion. A short section examining a possible relationship between schwa insertion and simplification following a word-final consonant cluster precedes a final discussion of the results presented in the chapter.

3.1 Literature Review

In the French language, there are two well known phonological operations that target consonant clusters at word boundaries: cluster simplification and schwa insertion. Both are understood to alter a sequence of consonants in order to either facilitate articulation or enhance perception of the sequence. For example, word-final consonant clusters will often be either simplified, as in (1-a) and (2-a), or have schwa inserted, as in (1-b) and (2-b).

(1) *"titres de gloire"*
"claims to fame"
  a. [tit də glwaʁ]
  b. [titʁə də glwaʁ]

(2) *"manifestent leur colère"*
"express their anger"
  a. [manifɛs lœʁ kɔlɛʁ]
  b. [manifɛstə lœʁ kɔlɛʁ]

The variability of word-final consonant clusters in French has been investigated either from the perspective of consonant cluster simplification or from the perspective of the word-final ∅ ∼/ə/ alternation. Many of these studies have been based on data from introspection; some have used corpora of experimentally collected language; while a few more recent studies have been based on large databases of recorded speech.

For word-final consonant clusters, both consonant cluster simplification and schwa insertion are highly variable in their application. Even when a particular motivation has been proposed for the simplified productions (cf. Côté (2004a)) or for the schwa productions (cf. Tranel (1987)), most authors note that the application of either remains optional in most circumstances. As well, it has been claimed that these variable pronunciatons may be sensitive to various phonological conditions, whether these be perceptual needs (Côté, 2000, 2004b), the demands of co-articulation (Barnes and Kavitskaya, 2002), prosodic requirements (Eychenne, 2005), or even lexical retrieval (Pustka, 2011). Additionally, many studies have suggested that the application of both consonant cluster simplification and schwa insertion following word-final consonant clusters varies according to dialect (Milne and Côté, 2009), speech style (Armstrong, 1998, 2001) and speech rate (Hansen, 1994). Confusing the matter still further, variables such as position of the word in the utterance (Malécot, 1976), word-length (Adda-Decker et al., 2008) and word-frequency (Racine and Grosjean, 2002) have all been demonstrated to have an effect on either consonant cluster simplification or the insertion of schwa at word boundaries. Sociolinguistic factors may also play a role in influencing the variable pronunciations of word-final consonant clusters (Laks, 1977; Armstrong, 2001; Durand and Eychenne, 2004; Boughton, 2008).

3.1.1 Schwa following a word-final consonant cluster

The insertion of schwa at word-boundaries in French has been extensively studied (Grammont, 1961; Dell, 1985; Tranel, 1987; Picard, 1991; Dell, 1995; Ayres-Bennett and Carruthers, 2001; Côté, 2000, 2007; Eychenne, 2005). Noske (1993) identifies a typology of 6 contexts where French schwa alternates with zero, of which his Type D ($\emptyset \sim$/ə/ alternation in the environment CC]_[C) is of present concern. Historically in France, the deletion process of schwa in these environments took place over the course of the 15^{th} to the 17^{th} centuries (Fouché, 1958). Dell (1985) suggests that in standard varieties of French, in general, word-final schwa isn't pronounced anymore, except to break up consonant clusters.

The phonological literature has offered many contributions towards understanding and explaining the variable pronunciations of word-final consonant clusters with respect to when and why schwa is or is not realized. Notwithstanding the obviously wide range of proposals and inevitable disagreements among the authors, almost without fail, researchers agree that, following a word-final consonant cluster, the realization of schwa is optional. The oft-cited work of Grammont (1914, 1961) contains the famous *"loi des trois consonnes"* whereby a sequence of three consonants is a prime and usual context for the realization of word-final schwa. However, this work (along with that of Fouché (1958)), was primarily intended as a pronunciation guide, rather than a formal description of the grammar of French. Early work within the framework of generative phonology often made reference to the notion of syllable well-formedness: Schwa is required whenever the surrounding consonants could not be properly syllabified without it (Pulgram, 1961; Morin, 1974; Tranel, 1987; Noske, 1993). However, while these syllabic approaches attempted to account for cases of *obligatory* schwa, they either failed, or neglected to even attempt to explain, the cases of *optional* schwa following a word-final consonant cluster. Côté (2000), in rebutting the earlier syllabic accounts, made the case instead for a sequential account whereby, even though the realization of schwa following a word-final consonant cluster is optional, the frequency of its realization increases according to several phonological generalizations and the Sonority Sequencing Principle (SSP). The SSP is marginally inviolable at word boundaries and triggers the realization of schwa if the medial consonant in the sequence is not a permissible sonority peak. In other words, if the medial consonant is trapped between two less sonorous consonants, the realization of schwa is more likely. Côté's phonological generalizations concern the role of adjacent vowels, the vulnerability of stop consonants, the desirability of contrast within sequences of consonants, and the effect of the adjacent prosodic boundary. In order to account for the variability of schwa following a word-final consonant cluster, she believes that these segmental factors have "a cumulative effect on the likelihood of schwa insertion and retention: the more such factors are present, the less probable schwa insertion/retention is." (p133).

Eychenne (2005) suggests that some of the observed variability in the realization of word-final schwa could be attributed to the notion that in some dialects schwa is lexical (e.g. in the Languedoc region of Southern France), while in others it emerges to satisfy the prosodic requirement that the head of a prosodic phrase be heavy (e.g. in the Midi-Pyrénées region of Southern France), while in others, it is neither lexical nor prosodic (e.g. in young Parisians' French).

With a view to describing the variability of schwa following a word-final consonant cluster in terms of sociolinguistic or stylistic factors, several authors have taken advantage of the increasing availability of spoken language corpora. Using recorded speech collected as part of the Phonologie du Français Contemporain (Durand et al., 2002, 2005, 2009), several authors note that schwas tend to be present more often in standard than in some regional varieties of French (e.g., Québec French) but are especially frequent in southern French (Durand and Eychenne, 2004; Eychenne, 2006). From conversational interviews recorded in Paris, Hansen (1994) finds that younger speakers tend to produce schwas less often than more elderly speakers. Both Eychenne (2003) and Kemp et al. (1980), while examining schwa at word boundaries in speakers from Québec, find that schwa at word boundaries in the vernacular does not exist, though for a small subset of four speakers who displayed a speaking style favoring high overall rates of cluster conservation, schwa surfaced principally following liquids (/ʁ, l/) and variably following stops and nasals.

Investigating the effects of speech style on the variability of word-final schwa, Adda-Decker et al. (1999) compare the occurrence of pronunciations with and without schwa in large corpora of orthographically transcribed speech: the BREF corpus (Lamel et al., 1991) comprised of 120 hours of read newspaper speech and a portion of the MASK corpus (Lamel et al., 1995) containing 35 hours of spontaneous speech collected via a travel information dialog system. They found that in the BREF corpus, final schwas in polysyllabic words are produced twice as often as in spontaneous speech, while in the MASK corpus, the majority of final schwas are dropped. Using a corpus of 30 one hour recordings of French radio interview speech, Adda-Decker et al. (2002) investigated the variation between the number of syllables present in the citation form of a word versus the number of syllables present in the fluent, spontaneous speech form of the word. Their results showed that the optional schwa vowel contributes to a large amount of pronunciation variation and that a substantial number of word-final syllables may be completely deleted. Among the observed deletions, 40% (i.e. $9,000$ occurrences out of $24,000$) correspond to syllables containing a schwa. Among the large number of omitted syllables, more than half are cross-word syllables: parts of words on word boundaries disappear more easily. From a corpus of 13 hours of broadcast news speech, Nemoto et al. (2010) combined time-aligned phonemic and lexical transcriptions, as well as automatic prosodic and POS annotations to compare average f_0 profiles according to word classes of given syllabic length, word-final schwa, duration and syntagms.[1] They found that average f_0 profiles tended to be raised in the presence of final schwa and interpret this to mean that the presence of word-final schwa may reveal measurable cues contributing to word boundary location.

Concentrating on a single style of speech, Bürki et al. (2011b) used the subset of the ESTER corpus (24 hours of radio-broadcasted news produced by 574 speakers (Galliano et al., 2005)) for which the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) automatic speech alignment system had produced a phonetic transcription aligned with the acoustic signal at the word and phoneme levels. While they restricted themselves in their study to the "schwa-zero alternation" in word-internal position in French, their detailed examination of the variables that might condition the presence versus absence of schwa using a large corpus of radio-broadcasted speech yields important information pertinent to the present study. They note that although many studies have investigated schwa alternation in French, no clear and complete picture has yet emerged of exactly which variables condition the presence of schwa. They attribute this lack of consensus to the fact that most studies consider variables individually (with one exception (Racine and Grosjean, 2002)), despite the fact that these variables often interact or are correlated. Their findings that relate to the current study include significant main effects for speech rate and respect of the sonority sequencing principle. Schwa was more often present at lower speech rates and schwa was more often present if the sequence of consonants that would result from the absence of schwa did not obey the sonority principle according to a six-level scale, that is, a sonority scale that differentiates between fricatives and stops, as opposed to one which groups the obstruents (c.f. Côté (2009)). They also discovered large differences among speakers: The effect of the sonority principle did

[1] Syntagm: A linguistic unit consisting of a set of linguistic forms (phonemes, words, or phrases) that are in a sequential relationship to one another

not affect all speakers' behaviour equally.

3.1.2 Simplification following a word-final consonant cluster

Modern French has a large number of word-final consonant clusters, the majority of which arose from the loss of word-final schwas in the pronunciation of French in the 17^{th} century. Consonant cluster simplification in French, like word-final schwa, has also been the beneficiary of numerous investigations. Similar to the case of word-final schwa, consonant cluster simplification, when it occurs, is optional. Word-final consonant cluster simplification also shares with accounts of word-final schwa the lack of a clear and complete understanding of exactly which variables condition the simplifications. Based on previous work using similar corpora to the one employed here, consonant cluster simplification is more common in the Québec dialect of French (Milne and Côté, 2009). In the Québec dialect, word-final consonant cluster simplification may occur freely in all contexts (preconsonantal, prepausal and prevocalic position) (Côté, 2004a), in contrast with the French spoken in France, where deletion of the final consonant in a cluster is more constrained and is allowed only before a consonant-initial word (Dell, 1985) or is restricted to more colloquial varieties prevocalically (de Cornulier, 1978; Tranel, 1987). In fact, in a 'standard conversation', de Cornulier (1978) is unsurprised that word-final consonant cluster simplification may occur before a consonant or pause, but regards simplification as stigmatized before a vowel in the absence of a pause. In so doing, he recognizes a social or stylistic component to word-final consonant cluster simplification. Several sociolinguistics studies have since shown that, although it is less common in these contexts, consonant cluster simplification (at least for clusters that end in /ʁ|l/) can indeed occur before vowels and pauses (Armstrong, 2001; Laks, 1977; Pooley, 1996). Laks (1977), studying /ʁ/ deletion patterns in word-final consonant clusters in the Parisian neighbourhood of Villejuif found relatively low rates of simplification and that the more formal the conversation and the higher the speaker's social status, the less likely the deletion of /ʁ/.

With respect to speech style, Armstrong (1998, 2001) found similar results. Comparing recordings from a more formal interview with recordings from a less formal conversation of school age girls in the Lorraine region of north-eastern France conducted in 1990, in formal settings, rates of /ʁ/ deletion before a vowel were lower than before consonants and pauses (22.1% vs 55.6%). In prevocalic environments, /ʁ/ deletion was limited to a few frequently occuring lexical items. In the case of /ʁ/ deletion, the effect of style was more important than other extra-linguistic factors, in this case gender and age. Looking at word-final consonant cluster simplification involving only clusters that terminate in /ʁ|l/ in the casual speech recordings from the Aveyrons and Paris regions of the PFC corpus, Pustka (2011) finds that the following context plays an important role: consonant cluster simplification was more common before a consonant, but cautions that prevocalic simplification was not so infrequent as to be dismissed out of hand. It was also found that the preceding context may also play a role: deletion of /ʁ/ or /l/ was more frequent when the first member of the cluster was a plosive than a fricative. These results from casual speech were similar to the analysis of Adda-Decker et al. (2002) who found that, for words ending in /tʁ/ and /dʁ/ preceding a consonant, the pronunciations [t] and [d] are preferred, in an average ratio of 3 to 2: after the elision of the schwa in this context, consonant cluster simplification was observed in 240 occurrences, while the final consonant was maintained together with the plosive in 170 occurrences.

Côté (2004a,b) has presented an account of the simplified productions where consonant clusters that contain either a rising sonority contour, or lack crucial featural contrasts, are susceptible to deletion. While her analysis is restricted to the variety of french spoken in Québec, her proposal makes several predictions regarding what to expect in the current study. She claims two types of clusters: Clusters that simplify and clusters that are stable. She divides the clusters according to their motivation for simplification. Her proposal appeals to either perceptual salience or sonority as violable constraints. Côté's argument for perceptual salience rests on the observations that salience is determined by the degree of contrast within the cluster and by the nature and position of each consonant. For example, /st/ and /ts/ display the same featural contrast, but simplification only affects the first one. In these cases the least salient consonants may delete. Côté's argument for sonority rests on the definition that the "SSP is violated in all clusters whose last consonant is more sonorous than the preceding one" (Côté, 2004a, p. 159) and that clusters that violate the SSP may delete. Côté allows for some lexical effects to influence the likelihood of consonant cluster simplification. For some consonant clusters, such as the frequently occurring stop+approximant combinations, simplification may occur in any of the words in which they occur while in others, such as some fricative+stop combinations (e.g. /sk, sp/), simplification may occur in just a subset of the words in which they occur. For these lexically determined cases of simplification word frequency, register, context of usage and word length may all play a role in determining whether a particular cluster is likely to be simplified. She claims that, in general, "the more frequent and less learned a word, the more likely it is to get simplified." (Côté, 2004a, p. 157). In all cases, when it is allowed, word-final consonant cluster simplification is always optional. Tranel (1987) and Goad (2002), noting the overt release of word-final stops in Continental French, offered a possible explanation for the lower incidence of simplification in the French spoken in Northern France.

Curiously, given the substantial amount of attention devoted to either consonant cluster simplification or schwa insertion at word boundaries, very little mention has been made of a possible relationship between these two phenomena. The literature just reviewed indicates that it is not disputed that both consonant cluster simplification and schwa insertion at word boundaries appear to share similar contexts: a sequence of consonants at a word boundary. It would also appear to be the case that the factors possibly influencing the one are similar to the factors possibly influencing the other: A rising sonority contour, faster rates of speech, the amount of contrast between sequences of consonants, either dialectal or regional variation, the special status of boundaries in lexical access, and a casual conversational (as opposed to formal read or broadcast news) speech style have all been associated with both consonant cluster simplification and schwa insertion at word boundaries. Nevertheless, in the literature reviewed, there are but two instances where both consonant cluster simplification and schwa insertion were mentioned together. Kemp et al. (1980) note that for four of their speakers who displayed a speaking style favoring high overall rates of consonant cluster conservation, schwa surfaced principally following liquids (/ʁ, l/) and variably following stops and nasals, while Pustka (2011) wondered about the difficulty of determining a causal relationship between consonant cluster simplification and schwa insertion: "la chute du schwa entraîne-t-elle la chute de la liquide ou bien les deux variables sont-elles (presque) indépendamment corrélées aux facteurs 'tradition' vs 'modernité'?" (p. 27). Neither of these two passing statements was further elaborated upon by the researchers.

3.1.3 Pilot study: Milne and Côté (2009)

Milne and Côté (2009) investigated the relationship between consonant cluster simplification and schwa insertion at word boundaries in two dialects of French: a variety spoken in and around Paris and Northern France (FR), and a variety spoken in Québec (QC). The object of the study was to better understand the differences between the two dialects in terms of rates and contexts of both consonant cluster simplification and schwa insertion following a word-final consonant cluster. The corpora used was drawn from two sources. For the dialect representing French spoken in and around Paris and Northern France (FR), the discussion libre portion of selected investigations from the Phonologie du Français Contemporain (PFC) was used (Durand et al., 2002, 2009). For the dialect representing Québec, the data came from audio and text files of the debates that occurred in the Assemblée nationale du Québec during the week of June $12 - 16, 2007$. Contained in this data are more than six hours of speech data from 65 speakers (46 males and 19 females). All of them are députés de l'Assemblée nationale du Québec and from different regions of Québec.

When rates of consonant cluster simplification and schwa insertion for individual consonant clusters were calculated, the expectation that these two phonological operations pattern differently in each dialect was confirmed. Rates of consonant cluster simplification were higher in the QC dialect than the

FR dialect (51% vs 18% overall simplification rate) and rates of schwa insertion were lower in the QC dialect than the FR dialect (20% vs 40% overall schwa insertion rate). However, it was determined that the distribution of occurences of consonant cluster simplification and schwa insertion among individual consonant clusters was not the same in each dialect. For FR, even as rates of consonant cluster simplification increased (rates of simplification were highest for those clusters ending with /ʁ,l/), rates of schwa insertion remained constant (50% of occurences of schwa insertion with clusters ending in /ʁ,l/). Whereas for QC, as rates of consonant cluster simplification increased (rates of simplification were highest in clusters ending in /ʁ,l/ as well as in clusters ending in a stop consonant), rates of schwa insertion also increased (95% of occurences of schwa insertion with clusters ending in /ʁ,l/ or a stop consonant). A possible relationship between consonant cluster simplification and schwa insertion, at least for the QC dialect, was proposed: in contexts where consonant cluster simplification did not occur (for whatever reason), schwa insertion was an alternative.

However, the analysis suffered from several deficits. In the first place, the conclusions were not statistically validated. While differences were observed between the two dialects of French, no statistical evidence was included to determine whether these differences were significant. Second, the data were not properly balanced between the two corpora. There was more than double the amount of data contained in the FR corpus as there was in the QC corpus, and not all clusters and lexical items were equally represented in the two corpora. Third, the set of word-final consonant clusters included in the results was not very large. Of the 87 possible word-final consonant clusters attested in French (Dell, 1995), only 21 were included in the results. Finally, the results were based on different speech styles. The data for the QC dialect, being drawn from the political debates in the Assemblée nationale du Québec, was most likely representative of a more formal speech style than the data for the FR dialect, which was drawn from a less formal conversational speech style. It might have been expected that the more formal speech style would have displayed fewer occurrences of word-final consonant cluster simplification and possibly more occurrences of schwa insertion following a word-final consonant cluster than the less formal speech style. In fact, the opposite was observed. The QC dialect, representative of a more formal speech style, showed higher rates of simplification and lower rates of schwa insertion following a word-final consonant cluster than the FR dialect, representative of a less formal style of speech. Therefore, the stylistic differences prevented measuring the real differences between the two dialects.

3.2 The data set and variables

The data used for the current study was obtained by combining two corpora of recorded French speech. The first corpus is the force-aligned portion of the AssNat corpus, described previously in Section 2.1. The AssNat corpus is a corpus of political debates from the national assemblies of Québec and France that occurred in the month of May 2011. This corpus contains approximately 126 hours of audio recordings (66 hours from France and 60 hours from Québec) representing 439 different speakers (105 from Québec, 334 from France). From this, 18 hours of recordings, containing variable amounts of speech data from 148 different speakers (62 from Québec, 86 from France), has been time aligned at both the word and phoneme level using the SPLaligner tool (described in Chapter 2). An important addition to this corpus are labels indicating whether the interval of audio was obtained while a speaker was physically reading from a text. The videos of the proceedings in the two national assemblies were watched and every audio interval was coded as either *"Reading"* or *"Spontaneous"* depending on whether or not the video indicated a speaker was reading from a text or speaking without the aid of notes. This distinction between reading and spontaneous speech was used to represent two styles of speech and mimics the distinction provided by the second corpus.

The second corpus contains the read text exercise and both recorded conversations from selected investigations in France and Québec through the PFC: Projet Phonologie du Français Contemporain (Durand et al., 2002, 2005, 2009). The PFC is an ongoing research project aimed at providing researchers interested in the French language with a database of oral data. One of the main goals is to gather data from as many varieties of French as possible, in all parts of the world, in order to investigate dialectal variation. The PFC uses a standardized interview process which includes two conversations (one free, the other guided) and a read text exercise (texte lu). The goal of the discussions is to obtain as natural a style of speech as possible while minimizing the effects of the observer's paradox. For this reason, the free discussion preferably involves at least two participants with or without the researcher. The texte lu is a standarized text which the participant is allowed to study and become familiar with before being asked to read it aloud. The investigations from France were selected with a view to minimizing some of the noted geographically determined variation in the pronunciation of the French spoken in France. Specifically, the obvious differences in the rates and contexts of schwa realizations observed between speakers from the North of France and speakers from the South of France. For this reason, the surveys historically in the Occitan and Basque regions of France were excluded. Table 3.1 displays the list of investigations chosen to represent a Northern France dialect of French, as well as the list of investigations made available to represent a Québec dialect of French. The table also displays some information about the number of speakers and the amount of audio data available.[2]

Dialect	Investigation(Code)	Speakers	Duration of audio $hh : mm : ss$
France	Dijon(21a)	8	02 : 49 : 54
	Nantes(44a)	11	09 : 10 : 37
	Brécey(50a)	11	03 : 58 : 18
	Domfrontais(61a)	12	05 : 25 : 44
	Ogévillier(54b)	11	08 : 39 : 36
	Vendée(85a)	8	03 : 08 : 53
	Brunoy(91a)	10	01 : 48 : 34
	Puteaux-Courbevoie(92a)	6	05 : 30 : 38
	Paris Centre ville(75c)	12	04 : 06 : 03
	Aveyronnais à Paris(75x)	9	08 : 14 : 44
	Lyon(69a)	11	07 : 40 : 30
	Roanne(42a)	9	04 : 41 : 56
Québec	Hawkesbury	13	08 : 24 : 35
	Université Laval(cqa)	9	03 : 32 : 14
	Chicoutimi	11	08 : 40 : 32
	Trois Rivières	12	10 : 09 : 09
	Vanier	12	09 : 30 : 09
	La Pocatière	12	09 : 09 : 01
	Pointe-Fortune	3	01 : 41 : 16
	Chelsea	11	06 : 28 : 37
Totals	France	118	65 : 15 : 28
	Québec	83	57 : 35 : 33

Table 3.1: Volume of data made available through the corpus

In the table are listed a few investigations that do not conform entirely with the assumption of a uniform Québec or northern France dialect. Hawkesbury is in Ontario, not in Québec. However, the French spoken in Hawkesbury, which is on the border with Québec, is essentially the same as that spoken on the other side of the border (Nadasdy, 2005). Aveyronnais à Paris(75x) is a survey of speakers originally from the south of France, but who have been living, and were recorded while in, Paris. This mirrors a similar situation in the France portion of the AssNat corpus which surely includes some members of the national assembly who originate from the south of France, but who now spend a large amount of time in a Parisian context. As such, it might be expected that both of these groups of speakers have adopted some Northern (i.e. "standard") patterns. So both of the France corpora are mainly "Northern", but include some Southern speakers in a Parisian context.

Using the Speaker Adapted Training tool provided by the SPLAligner, the transcriptions and audio recordings of these PFC investigations were time aligned following the same procedures as those described for the AssNat corpus. Even though portions of the PFC corpus had been previously coded for the presence or absence of both schwa and simplification following a word-final consonant cluster (several

[2]The survey points with a code are available online at www.projet-pfc.net. The others are not yet online but will be at some point.

minutes of recorded audio from each participant containing approximately 6 minutes of the conversations and all of the read text exercise, typically 2 minutes), it was decided to use the automatic force aligned judgments for the entirety of the corpus. This was done because the forced aligner had been determined to perform at a level comparable to human annotators and it was desirable to have consistency in judgments across both corpora.

From both of these force-alignments (the AssNat and PFC corpora) a data set was constructed with an entry for every occurence of a word-final consonant cluster. Excluded from the data were a set of words that do not appear to show any variation in their pronunciations. These include the words "*presque*", "*puisque*", "*lorsque*" and "*jusque*". All of these words are pronounced almost categorically with a final schwa vowel before a consonant-initial word and are rarely simplified when occurring before a vowel-initial word. Also excluded was the frequently occurring word "*quelque(s)*". This word is also pronounced almost categorically with a final schwa vowel before a consonant-initial word, including liaison consonants, and is rarely simplified when occurring before a vowel-initial word. "*quelque*" may also often be pronounced [kɛk] without the [l], in which case no schwa is realized. Also excluded from the data for analysis were examples of extremely fast or extremely slow speech. Speech rate is included as a variable of interest and as a normally distributed continuous variable the upper and lower tails of the distribution (at $\alpha = 0.05$) could be considered extreme. The measured rates of speech centred on a mean of 13.178 phones per second and ranged from a minimum of 2.786 phones per second to a maximum of 30.770 phones per second. Using 95% of the distribution meant the upper limit for speech rate was set at 19.67333 phones per second and the lower limit set at 7.109537 phones per second. Observations whose speech rates were either slower or faster than these limits were removed from the outset in order to avoid any outlier effects. Additionally, since true nasal-stop sequences (e.g. /nd, nt, mp/) are rare in French and mostly appear in English loan words, these consonant clusters were not included. Also removed were any specific consonant clusters with too small frequency counts and occurrences of word-final consonant clusters before a non-speech sound, such as a cough or noise. The resulting data set used for this project contains $28,095$ observations of word-final consonant clusters comprised of 29 individual clusters grouped into 12 types collected from 351 different speakers. Table 3.2 displays the aggregate counts of each type of cluster according to both dialect and speech style in the final data set. In the table, A, N, F, and S stand for approximant, nasal, fricative and stop, respectively.

Type	/Clusters/	Reading QC	Reading FR	Spontaneous QC	Spontaneous FR	Sum
AF	ʁʃ, ʁs, ʁv, ʁz, ʁʒ	225	273	2381	1688	4567
AN	ʁm, ʁn	82	230	292	287	891
AS	ʁd, ʁk, ʁt	71	115	888	445	1519
ASA	ʁdʁ	13	17	38	32	100
FA	fʁ, vʁ	47	86	383	239	755
FN	sm	46	35	106	57	244
FS	ft, sk, st	347	609	1145	471	2572
FSA	stʁ	649	931	149	139	1868
SA	bʁ, bl, dʁ, gl, kl, pʁ, pl, tʁ	1400	2156	6922	4066	14544
SF	ks	18	28	140	27	213
SFS	kst	17	66	44	68	195
SS	kt, pt	106	170	252	99	627
Sum		3021	4716	12740	7618	28095

Table 3.2: Attested word-final consonant clusters in the data set

3.2.1 Variables to be selected for analysis

The variables to be explored include speech style, following context, motivation for simplification and speech rate. These variables were selected primarily because the body of literature on both simplification and schwa insertion following a word-final consonant cluster indicated that they were all likely to have explanatory value in any models predicting either simplification or schwa insertion. Secondarily, while there are undoubtedly other variables that are likely to be important predictors for simplification or schwa insertion following a word-final consonant cluster, it was not possible to control for these other variables while still obtaining a large and robust sample size for analysis. Incorporation of other variables of interest, such as lexical effects, inter- or even intra-speaker variation, produced dramatically unbalanced experimental designs, even in this relatively large corpus.

In this investigation, speech style was defined twice. In the first instance, following the example from several previous studies(c.f. Nemoto et al. (2010); Adda-Decker et al. (2002, 1999)), a distinction was made between read and non-read speech as a way of accessing different speech styles. Defined this way, speech style is a categorical predictor with two levels. The first level, Reading, applies to clusters that occurred in a sample of speech during which the participant was actively reading from a text (either reading from prepared notes in the AssNat portion of the corpus, or the texte lu exercise from the PFC portion of the corpus). The second level, Spontaneous, applies to clusters that occurred in a sample of speech during which the participant was not actively reading from a text (speaking without the use of prepared notes in the AssNat portion of the corpus, or the conversational portion of the PFC corpus).

In the second instance (explored only in Section 3.6), speech style was defined instead according to the different corpora collected for this investigation. In addition to containing examples of both read and non-read speech, these two corpora contain examples of speech that were collected in a more formal speech setting (the national assemblies of both Québec and France) as well as examples of speech collected in a less formal setting (casual conversations). Defined this second way, speech style remains a categorical predictor with two levels corresponding to the corpora from which the speech data came. The AssNat corpus is taken to be representative of a more formal speech setting while the PFC corpus is representative of a less formal speech setting.

Following context is also a categorical predictor with three levels. The three levels, 'C', 'V' and 'P' correspond to whether the consonant cluster occurred in a preconsonantal, prevocalic, or prepausal context. All three levels depend on whether or not a pause was realized in the audio record. The presence or absence of a pause is determined by a purely physical measurement. Whether or not a pause was realized following the target word containing the consonant cluster was determined simply by measuring the length of the `{sp}` ('short pause') or `{sl}` ('silence') phone that follows every word in the forced alignment reference transcription. If the length of this pause phone is greater than $0.0ms$, it indicates that the two words were not co-articulated. If the length of this pause phone is exactly $0.0ms$, it indicates that the two words were co-articulated. This measurement therefore makes no attempt to distinguish between pauses possibly inserted at phrasal boundaries, which could be considered phonological, and pauses that occur simply as a result of an absence of speech signal detected by the aligner. Word-final consonant clusters that were followed by a 0.0 length pause (ie, that were co-articulated with the following word) were then identified as being either preconsonantal or prevocalic. Word-final consonant clusters that were followed by a non-zero length pause were identified as being prepausal and no attempt was made to distinguish whether the word following the pause was consonant- or vowel-initial.

The proposal by Côté (2004a) (which only referred to the dialect from Québec, but will be applied here also to the dialect from France), makes several predictions regarding what to expect in this data and how the variable of motivation for simplification is defined. She claims two types of clusters: Clusters that simplify and clusters that are stable. She divides the clusters according to their motivation for simplification. Her proposal appeals to either perceptual salience or sonority as violable constraints. Her complete analysis will not be duplicated here.[3] In general, the word-final consonant clusters that were investigated here that are predicted to simplify were those that either violated the Sonority Sequencing Principle (SSP) or contained a final consonant, specifically a final stop-consonant, that shared some features with its neighbour. The word-final consonant clusters that are predicted to remain stable were those that either didn't violate the SSP or whose final member, even if it was also a stop consonant, contained crucial featural contrasts with its neighbour. The generalizations to be drawn with respect to

[3] For a complete understanding of the proposal, please read carefully (Côté, 2004a)

the present data are that:

1. Clusters that violate the SSP are predicted to simplify. In this data set they are {/bʁ, bl, dʁ, gl, kl, pl, pʁ, ʁdʁ, tʁ, vʁ, fʁ, stʁ, sm/}.

2. Clusters that respect the SSP but whose members lack crucial featural contrasts are predicted to simplify. In this data set they are {/kst, kt, pt, sk, st/}. Notably, all of them contain a final voiceless stop.

3. Clusters that respect the SSP and whose members contain crucial featural contrasts are predicted to remain stable. In this data set they are {/ks, ʁʃ, ʁm, ʁn, ʁt, ʁd, ʁk, ʁs, ʁv, ʁz, ʁʒ/}

The motivation for simplification, therefore, can actually be decomposed into two separate categorical predictors: One that was used only when examining word-final consonant cluser simplification and one that was used only for schwa insertion, although they both essentially refer to a similar distinction.

When examining simplification, the motivation for simplification is adapted from the proposal in Côté (2004a) and refers to the difference between SSP violating consonant clusters and stop-final consonant clusters. When applied only to those clusters that are predicted to simplify, the variable 'SSP' then has two levels: 0 applies to stop-final consonant clusters and 1 applies to SSP violating consonant clusters.

When examing schwa insertion, the motivation for simplification refers simply to whether or not a consonant cluster also participated in simplification and has two levels: 'Simplify' applies to consonant clusters that did present observations of simplifications (SSP violating and perceptually deficient stop-final consonant clusters) and 'Stable' applies to consonant clusters that did not present observations of simplification (all /ʁ/-initial consonant clusters as well as /ks/).

Speech rate is a continuous variable that quantifies the rate of speech, measured in terms of pronounced phones per second. This measurement was obtained from the results of forced alignment of the data. During forced alignment, an optional short pause `{sp}` or silence phone `{sl}` is inserted between words. When it is realized (that is, when the short pause or silence phone has a duration of $> 0.0ms$), it indicates that the acoustic record reflects an absence of speech signals for a brief period of time and that the two words were not co-articulated. Therefore, the number of realized phones in an interval of non-interrupted speech was counted and this number is divided by the duration in seconds of the interval to give us a value for speech rate measured in phones per second.

The variable for speech rate is included in the analysis because, while exploring the data, it was observed that there were significant differences in the mean speech rates between the two styles: reading and spontaneous. In both dialects, mean rates of speech were higher in the spontaneous style than in the reading style. Figure 3.1 displays this observation. If it is true that changes in style are associated with changes in speech rate, then failing to take into account changes in speech rate renders any analysis of the effects of style unreliable. If differences in the probability of simplification due to rates of speech can be predicted, then the differences in the response terms between reading and spontaneous styles of speech that would be expected due to differences in rates of speech between the two styles can also be predicted. Any difference between the two styles beyond this prediction cannot be put down to differences in rates of speech and can therefore be attributed to differences between the two styles.

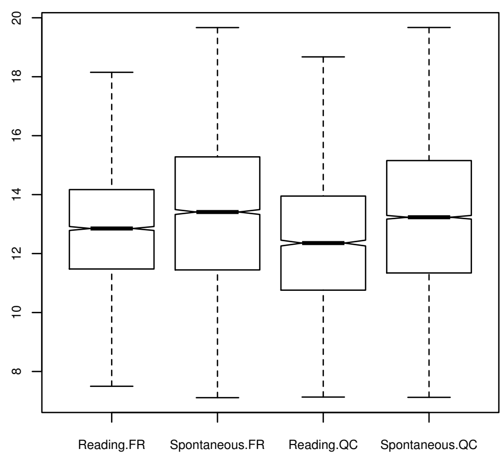

Figure 3.1: Boxplot of SpeechRate by style and dialect

What follows is an examination of the individual effects of each of these variables of interest on the responses of either simplification or schwa insertion following a word-final consonant cluster in each dialect.

3.3 Simplification in the corpus

To begin, it will not be surprising to learn that, in general, word-final consonant cluster simplification is more prevalent in the dialect from Québec than in the dialect from France. Approximately 72% of the occurences of word-final consonant clusters in these data simplified in the data from Québec compared with approximately 38% in the data from France. With an odds ratio of 4.33, word-final consonant cluster simplification is slightly more than 4 times more likely to occur in the data from Québec than in the data from France.

The data indicate that word-final consonant cluster simplification is observed in the same group

Simplify	FR	QC	Sum
0	5764 (62.37%)	3229 (27.68%)	8993
1	3477 (37.63%)	8435 (72.32%)	11912
Sum	9241	11664	20905

Table 3.3: Frequency and proportion of simplification according to dialect

of clusters in both dialects. In this data they are the 13 clusters that violate the Sonority Sequencing Principle and the 5 clusters whose final member is a perceptually deficient stop consonant. The remaining 11 clusters showed either exceptionally low or no occurences of simplification. Table 3.4 shows the overall proportion of observations of word-final consonant cluster simplification for each cluster in each dialect.

		kst	kt	pt	sk	st								
Stop-final	QC	0.85	0.61	0.63	0.39	0.72								
	FR	0.25	0.24	0.30	0.14	0.20								
		bʁ	bl	dʁ	gl	kl	pʁ	pl	tʁ	stʁ	ʁdʁ	fʁ	vʁ	sm
SSP	QC	0.73	0.66	0.76	0.47	0.67	0.54	0.73	0.75	0.75	0.73	0.68	0.78	0.52
	FR	0.37	0.40	0.49	0.33	0.33	0.11	0.54	0.42	0.41	0.43	0.21	0.38	0.14
		ʁʃ	ʁs	ʁv	ʁz	ʁʒ	ʁn	ʁm	ʁd	ʁk	ʁt	ks		
Stable	QC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	FR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

Table 3.4: Proportion of simplification of clusters by Dialect

3.3.1 Differences in simplification according to following context

The literature reviewed indicated that word-final consonant cluster simplification in the dialect from Québec may occur freely in all contexts: Preconsonantal, prevocalic, and prepausal. Left unexplored is the question of whether it is equally likely in all three contexts in this dialect. The literature on word-final consonant cluster simplification in the dialect from France suggests that deletion of the final consonant is more frequently observed when preconsonantal than when either prevocalic or prepausal. Table 3.5, which presents both the frequency and proportion of observations of word-final consonant cluster simplification in each dialect according to following context, suggests that word-final consonant cluster simplification appears to be more likely before a consonant-initial word than before either a vowel-initial word or a pause and that this may be true of both dialects.

For word-final consonant clusters that appeared in a preconsonantal context, simplification of the cluster was observed in 79% of the cases in the data from Québec and 44% of the cases in the data from France. These rates dropped to 68% and 31%, respectively, when the cluster was prepausal, and dropped further to 57% and 23% when prevocalic. Although it is conceivable that these differences

Simplify		Consonant	Vowel	Pause	Sum
QC	0	1470 (20.88%)	1147 (42.62%)	612 (31.64%)	3229
	1	5569 (79.12%)	1544 (57.38%)	1322 (68.36%)	8435
FR	0	3126 (55.77%)	1469 (76.27%)	1169 (68.36%)	5764
	1	2479 (44.23%)	457 (23.73%)	541 (31.64%)	3477
Sum		12644	4617	3644	20905

Table 3.5: Frequency and Proportion of Simplification According to Context

were due to chance, that doesn't seem likely. A chi-square test of the null hypothesis that word-final consonant cluster simplification is independent of the following context will help to answer the question of whether following context plays a role in word-final consonant cluster simplification in either dialect. Table 3.6 displays the observed and expected frequencies for simplification as a function of context for the data from Québec. Table 3.7 displays the similar information for the data from France.

	Observed			Expected		
	Simplification			Simplification		
Context	Yes	No	Total	Yes	No	Total
C	5569	1470	7039	5090.36	1948.64	7039
V	1544	1147	2691	1946.04	744.96	2691
P	1322	612	1934	1398.60	535.40	1934
Total	8435	3229	11664	8435	3229	11664

Table 3.6: Observed and Expected Frequencies for Simplification as a Function of Context for Québec

In these data from Québec, the obtained value of $\chi^2(2) = 462.6$ is significantly larger than than the critical value of $\chi^2(2) = 5.99$ and the null hypothesis must be rejected. There are significantly more observations of word-final consonant cluster simplification when the cluster is preconsonantal and significantly fewer observations when prevocalic than would be expected.

	Observed			Expected		
	Simplification			Simplification		
Context	Yes	No	Total	Yes	No	Total
C	2479	3126	5605	2108.93	3496.07	5605
V	457	1469	1926	724.67	1201.33	1926
P	541	1169	1710	643.40	1066.60	1710
Total	3477	5764	9241	3477	5764	9241

Table 3.7: Observed and Expected Frequencies for Simplification as a Function of Context for France

In these data from France, the obtained value of $\chi^2(2) = 262.63$ is significantly larger than than the critical value of $\chi^2(2) = 5.99$ and the null hypothesis must again be rejected. As was seen in the data from Québec, there are significantly more observations of word-final consonant cluster simplification

when the cluster is preconsonantal and significantly fewer observations when prevocalic than would be expected. Therefore, it would appear that, in these data, word-final consonant cluster simplification may occur in all three contexts (preconsonantal, prevocalic, prepausal) in both dialects, although the frequencies of simplification are higher in the data from Québec for all three, and the effect of following context is similar in both dialects: Observations of word-final consonant cluster simplification are higher than expected when the cluster is preconsonantal, and lower than would be expected when the cluster is either prevocalic or prepausal. In both dialects, word-final consonant cluster simplification is not independent of following context.

3.3.2 Differences in simplification according to motivation for simplification

The motivation for word-final consonant cluster simplification in Québec has been proposed to be driven by two independent factors: The Sonority Sequencing Principle (SSP) and perceptual salience (Côté, 2004a). For clusters whose final consonant is more sonorous than the preceding one, a violation of the SSP has occurred and the final consonant in the cluster may delete. In the data used for the current study, all the SSP violating clusters, except for /sm/, terminate in a final /ʁ/ or /l/. For clusters that end with a stop consonant and whose final consonant lacks crucial featural contrasts with the preceding one, the perceptual saliency of the final consonant is deficient and it may delete. In these data, the consonant clusters that may simplify because of a lack of perceptual salience all terminate in a final stop consonant. There are a few non-SSP violating, non-stop-final consonant clusters that also regularly simplify in French, such as N+N /mn/ or F+F /vz/, but these were not attested in these data. Both the SSP violating and stop-final consonant clusters (at least those stop-final clusters that are not preceded by /ʁ/) in these data are expected to regularly simplify in the data from Québec. The literature reviewed pertaining to word-final consonant cluster simplification in the dialect from France, however, describes patterns of simplification only for SSP violating clusters, specifically word-final consonant clusters terminating in a final /ʁ/ or /l/. The pilot study by Milne and Côté (2009) observed that rates of word-final consonant cluster simplification involving stop-final clusters were indeed lower than the rates observed in SSP violating clusters in their data from France and correspondingly grouped those stop-final clusters with other non-simplifying, or stable, clusters. However, in these current data, Table 3.4 suggests that the frequency of simplification in stop-final clusters in the data from France are significantly higher than the rates of simplification observed with stable clusters and may, in fact, be more similar to the rates of simplification of SSP violating clusters. Table 3.8, which presents both the frequency and proportion of observations of word-final consonant cluster simplification in each dialect according to the motivation for simplification (SSP or Stop-Final), suggests that word-final consonant cluster simplification is more common with clusters that violate the SSP than with the stop-final clusters in the data from France, while this difference may not exist in the data from Québec.

Dialect	Simplify	Stop-final	SSP	Sum
QC	0	612 (32.03%)	2617 (26.83%)	3229
	1	1299 (67.98%)	7136 (73.17%)	8435
FR	0	1172 (79.03%)	4592 (59.19%)	5764
	1	311 (20.97%)	3166 (40.81%)	3477
Total		3394	17511	20905

Table 3.8: Frequency and Proportion of Simplification According to Motivation for Simplification

In the data from Québec, the proportion of simplified word-final consonant clusters is approximately the same in both SSP violating clusters and stop-final clusters (73% $\sim$ 68%). However, in the data from France, the proportion of simplified clusters with SSP violating clusters is double the proportion of simplification with the stop-final clusters (41% $\sim$ 21%). A chi-square test of the null hypothesis that word-final consonant cluster simplification is independent of the motivation for simplification will help to answer the question of whether the motivation for simplification plays a role in word-final consonant cluster simplification in either dialect. Table 3.9 displays the observed and expected frequencies for simplification as a function of motivation for the data from Québec. Table 3.10 displays the similar information for the data from France.

	Observed			Expected		
	Simplification			Simplification		
Motivation	Yes	No	Total	Yes	No	Total
SSP	7136	2617	9753	7053.03	2699.97	9753
Stop-final	1299	612	1911	1381.97	529.03	1911
Total	8435	3229	11664	8435	3229	11664

Table 3.9: Observed and Expected Frequencies for Simplification as a Function of Motivation for Québec

In these data from Québec, the obtained value of $\chi^2(1) = 21.52$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must be rejected. There are significantly more observations of word-final consonant cluster simplification when the motivation for simplification is a violation of the SSP and significantly fewer observations when the motivation for simplification is perceptual salience (stop-final) than would be expected.

In these data from France, the obtained value of $\chi^2(1) = 208.78$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must again be rejected. As was seen in the data from Québec, there are significantly more observations of word-final consonant cluster simplification

	Observed			Expected		
	Simplification			Simplification		
Motivation	Yes	No	Total	Yes	No	Total
SSP	3166	4592	7758	2919.01	4838.99	7758
Stop-final	311	1172	1483	557.99	925.01	1483
Total	3477	5764	9241	3477	5764	9241

Table 3.10: Observed and Expected Frequencies for Simplification as a Function of Motivation for France

when the motivation for simplification is a violation of the SSP and significantly fewer observations when the motivation for simplification is perceptual salience (stop-final) than would be expected. Therefore, it would appear that, in these data, word-final consonant cluster simplification may occur in SSP violating and stop-final clusters in both dialects, although the frequencies of simplification are higher in the data from Québec for both, and the effect of motivation, while similar in both dialects, is much stronger in the data from France: Observations of word-final consonant cluster simplification are higher than expected for SSP violating clusters, and lower than would be expected for stop-final clusters. In both dialects, word-final consonant cluster simplification is not independent of the motivation for simplification.

3.3.3 Differences in simplification according to style: Reading vs Spontaneous

Several researchers have noted that speech style may play a role in word-final consonant cluster simplification (Armstrong, 2001, 1998; Laks, 1977). The general conclusion these authors reach is that word-final consonant cluster simplification is less frequent in more formal speech styles (i.e. reading) than in less formal speech styles (i.e. conversational). However, those authors examined only consonant clusters ending in a final /ʁ/ consonant; a subset of the SSP violating clusters. Milne and Côté (2009) examined a broader sample of word-final consonant clusters, but as has been stated several times, the data used in that study may have been drawn from different styles of speech in each of the two dialects compared and therefore the conclusions may not be reliable unless differences in speech style are taken into account as well as differences in dialect. In the data displayed in Table 3.11, word-final consonant cluster simplification appears to be more frequent in a spontaneous style of speech than in a reading style of speech and this appears to be the case in both dialects.

In the data from Québec, nearly 77% of the occurrences of word-final consonant clusters from the spontaneous style of speech simplified compared with approximately 58% in the data from the reading style of speech. In the data from France, approximately 48% of the occurrences of word-final consonant clusters from the spontaneous style of speech simplified compared with nearly 24% in the data from

Dialect	Simplify	Reading	Spontaneous	Sum
QC	0	1113 (42.4%)	2116 (23.41%)	3229
	1	1512 (57.6%)	6923 (76.59%)	8435
FR	0	3098 (76.12%)	2666 (51.56%)	5764
	1	972 (23.88%)	2505 (48.44%)	3477
Total		6695	14210	20905

Table 3.11: Frequency and Proportion of Simplification According to Style

the reading style of speech. While it is not likely that differences this large will be due to chance, a chi-square test of the null hypothesis that word-final consonant cluster simplification is independent of style will help to answer the question of whether speech style plays a role in word-final consonant cluster simplification in either dialect. Table 3.12 displays the observed and expected frequencies for simplification as a function of style for the data from Québec. Table 3.13 displays the similar information for the data from France.

	Observed			Expected		
	Simplification			Simplification		
Style	Yes	No	Total	Yes	No	Total
Reading	1512	1113	2625	1898.31	726.69	2625
Spontaneous	6923	2116	9039	6536.69	2502.31	9039
Total	8435	3229	11664	8435	3229	11664

Table 3.12: Observed and Expected Frequencies for Simplification as a Function of Style for Québec

In these data from Québec, the obtained value of $\chi^2(1) = 366.45$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must be rejected. There are significantly more observations of word-final consonant cluster simplification in the spontaneous style of speech than would be expected and significantly fewer observations in the reading style of speech than would be expected.

	Observed			Expected		
	Simplification			Simplification		
Style	Yes	No	Total	Yes	No	Total
Reading	972	3098	4070	1531.37	2538.63	4070
Spontaneous	2505	2666	5171	1945.63	3225.37	5171
Total	3477	5764	9241	3477	5764	9241

Table 3.13: Observed and Expected Frequencies for Simplification as a Function of Style for France

In these data from France, the obtained value of $\chi^2(1) = 585.41$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must again be rejected. As was seen in the data from Québec, there are significantly more observations of word-final consonant cluster simplification

in the spontaneous style of speech than would be expected and significantly fewer observations in the reading style of speech than would be expected. Therefore, it would appear that, in these data, word-final consonant cluster simplification is more frequent in both spontaneous and reading styles in the data from Québec, and the effect of style is similar in both dialects: Observations of word-final consonant cluster simplification are more frequent than expected in the spontaneous speech style, and less frequent than would be expected in the reading speech style. In both dialects, word-final consonant cluster simplification is not independent of speech style.

3.3.4 Speech style or speech rate as explanatory variable for simplification

As suggested by the non-overlapping notches in the box plots of Figure 3.1, the mean rates of speech are different when comparing a reading with a spontaneous speech style and this is true of both dialects. Table 3.14 shows the mean and standard deviation of speech rates in each dialect according to those speech styles. An analysis of variance of these data indicates no significant effect of dialect ($F(1, 20901) = 3.344, p = 0.0675$). That is, there does not appear to be a significant difference in the distribution of speech rates between the two dialects. There is a significant effect of style ($F(1, 20901) = 333.106, p < 0.001$). In general, speech rates are faster in the data coded as spontaneous than in the data coded as reading. Finally, there is a significant interaction between dialect and style ($F(1, 20901) = 15.701, p < 0.001$). The difference in speech rates between a reading and spontaneous speech style is larger in the data from Québec than in the data from France.

		Mean	SD
QC	Reading	12.40	2.244
	Spontaneous	13.26	2.668
FR	Reading	12.82	2.021
	Spontaneous	13.37	2.648

Table 3.14: Mean and standard deviation for speech rate

It also appears that speech rates are higher when simplification occurs and lower when simplification does not occur. This is also true of both dialects. Table 3.15 shows the mean and standard deviation of speech rates in each dialect according to whether simplification occurred or not. An analysis of variance of these data again indicates no significant effect of dialect ($F(1, 20901) = 3.332, p = 0.068$). There is an effect of simplify ($F(1, 20901) = 270.247, p < 0.001$). In these data, speech rates are faster when word-final consonant cluster simplification has occurred than when simplification has not occurred. There is no significant interaction between dialect and simplify ($F(1, 20901) = 2.861, p = 0.0907$).

How this investigation of speech rates in these data inform the analysis is as follows: Even while

		Mean	SD
QC	No Simplification	12.58	2.488
	Simplification	13.25	2.622
FR	No Simplification	12.92	2.304
	Simplification	13.47	2.535

Table 3.15: Mean and standard deviation of speech rate by simplification according to dialect

the proportion of simplified word-final consonant clusters is greater in spontaneous than in reading speech, speech rates tend to be higher in spontaneous than in reading speech and there is the suggestion that speech rates tend to be higher when word-final consonant cluster simplification occurs. If these observations are correct, seeking to understand whether there exist differences between speech styles in terms of word-final consonant cluster simplification without taking into account the rate of speech is unreliable, because the apparently higher rates of simplification in spontaneous speech may not reflect a difference between the styles, *per se*, but simply that rates of simplification are higher in speech styles with faster rates of speech and, in this sample, the data from the spontaneous style has higher rates of speech than the data from the reading style. In other words, to what extent are the observed differences due to rate or style? In this instance, an analysis of covariance would try to determine whether there is a difference between the styles in terms of rates of word-final consonant cluster simplification independent of any differences in rates of speech between the styles that may exist.

The fact that speech rate has an effect on word-final consonant cluster simplification is shown in Figure 3.2 which predicts the probability of simplification as a function of speech rate. The relationship described by this regression indicates that for every one unit increase in speech rate, the odds of observing word-final consonant cluster simplification increase by a factor of 1.09.

A logistic regression examines to what extent the probability of word-final consonant cluster simplification is affected by rates of speech between the styles. In particular, whether or not the probabilities differ at all. Of concern is whether the odds ratio for simplification is exactly the same in the data from both reading and spontaneous styles (the hypothesis of coincidence) or whether the odds ratio for simplification for any given value of speech rate in both styles are equal, even if the intercepts are different (the hypothesis of parallelism). If coincidence is accepted, then a single overall regression line can be fit to both reading and spontaneous styles. If the regression lines are parallel, then the effect of speech rate on the probability of simplification is the same in each style, but the 'base-line' values for simplification are different for the two styles. Separate regressions will be performed for each of the two dialects.

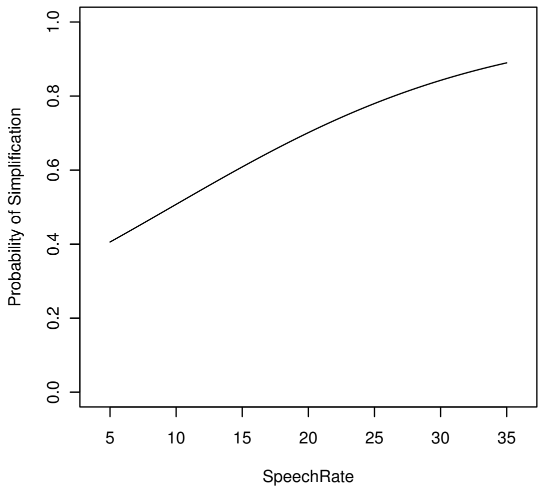

Figure 3.2: Regression Predicting Simplification as a Function of Speech Rate

The complete model is given by (3.1)

$$logit[p(y)] = log\left[\frac{p(y)}{1-p(y)}\right] = \alpha + \beta_1 x + \beta_2 z + \beta_3 x \times z + \varepsilon \tag{3.1}$$

Where $logit[p(y)]$ is the log of the odds ratio for simplification, z is the rate of speech, x is the dummy variable defined by

$$x = \begin{cases} 0 & \text{if style is reading} \\ 1 & \text{if style is spontaneous} \end{cases}$$

and ε accounts for random variation.

The table of coefficients for the model predicting the probability of simplification by style and rates of speech in the data from Québec is given below in Table 3.16. This table indicates that only the interaction term between style and speech rate appears to be highly significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−0.05035	0.22202	−0.227	0.820588	
Spontaneous	−0.09157	0.25497	−0.359	0.719478	
Rate	0.02879	0.01765	1.631	0.102825	
Spontaneous:Rate	0.07279	0.02006	3.629	0.000285	***

Table 3.16: Coefficients for model predicting simplification by style and rate for Québec

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.17, indicates two things. First, the difference between the null deviance and the residual deviance

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			11663	13762	
Style	1	346.28	11662	13416	$< 2.2e-16$
Rate	1	105.51	11661	13310	$< 2.2e-16$
Style:Rate	1	13.11	11660	13297	0.0002934

Table 3.17: Analysis of deviance table for Québec

approximately follows a chi-squared distribution with as degrees of freedom the difference between the degrees of freedom of the two deviances. Thus, with $\chi^2(3) = 465, p < 0.001$, indicates that this model has explanatory value. Second, the small `Pr(>Chi)` values indicates that each term in the model is predictive.

Testing for coincidence, that is, $H_0 : \beta_1 = \beta_3 = 0$, is simply testing significance of the terms x and $x \times z$, simultaneously. This can be done by removing those terms from the model and evaluating whether they add much to the understanding of the data. Comparing the simplified model with just a single term for `Rate` with the initial model returns $\chi^2(2) = 306.58, p < 0.001$. This says that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. Therefore, it seems the two regression lines are not coincident.

Testing for parallelism, that is $H_0 : \beta_3 = 0$, is testing significance of the term $x \times z$. Removing the interaction term and comparing this simplified model with terms only for `Style` and `Rate` with the initial model returns $\chi^2(1) = 13.112, p < 0.001$. This says again that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. The conclusion is that the two regression lines are not parallel.

The table of coefficients for the model predicting the probability of simplification by read or spontaneous style and rates of speech in the data from France is given below in Table 3.18. This table indicates that only the interaction term between style and speech rate appears to be highly significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−1.48984	0.23720	−6.281	$3.37e-10$	***
Spontaneous	0.23018	0.27845	0.827	0.40844	
Rate	0.02574	0.01820	1.414	0.15731	
Spontaneous:Rate	0.06374	0.02111	3.020	0.00253	**

Table 3.18: Coefficients for model predicting simplification by style and rate for France

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.19, indicates two things. First, the difference between the null deviance and the residual deviance approximately follows a chi-squared distribution with as degrees of freedom the difference between the degrees of freedom of the two deviances. Thus, with $\chi^2(3) = 674, p < 0.001$, indicates that this model

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			9240	12239	
Style	1	600.62	9239	11638	$< 2.2e-16$
Rate	1	64.16	9238	11574	$1.146e-15$
Style:Rate	1	9.11	9237	11565	0.002541

Table 3.19: Analysis of deviance table for France

has explanatory value. Second, the small `Pr(>Chi)` values indicates that each term in the model is predictive.

Testing for coincidence, that is, $H_0 : \beta_1 = \beta_3 = 0$, is simply testing significance of the terms x and $x \times z$, simultaneously. This can be done by removing those terms from the model and evaluating whether they add much to the understanding of the data. Comparing the simplified model with just a single term for `Rate` with the initial model returns $\chi^2(2) = 561.3, p < 0.001$. This says that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. Therefore, it seems the two regression lines are not coincident.

Testing for parallelism, that is $H_0 : \beta_3 = 0$, is testing significance of the term $x \times z$. Removing the interaction term and comparing this simplified model with terms only for `Style` and `Rate` with the initial model returns $\chi^2(1) = 9.111, p < 0.05$. This says again that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. The conclusion is that the two regression lines are not parallel.

The very similar models that emerged for each dialect suggests that the interpretation of these models for each dialect is also similar. Because the hypothesis of coincidence was rejected in both dialects it must be accepted that, in both the data from Québec and France, after controlling for rates of speech there exist significant differences between a reading and spontaneous speech style with respect to word-final consonant cluster simplification and it is right to consider speech style as an explanatory variable in both dialects. Since the hypothesis of parallelism was also rejected in both dialects, it must be accepted that the effect of speech rate on word-final consonant cluster simplification is not the same in each style and it is right to consider speech rate as an explanatory variable in both dialects. While reading, there is no effect of speech rate on word-final consonant cluster simplification. This contrasts with the effect of speech rate in the spontaneous speech style. Freed from the demands of reading from a text, the odds of observing word-final consonant cluster simplification increase along with rates of speech. These relationships are presented in Figure 3.3.

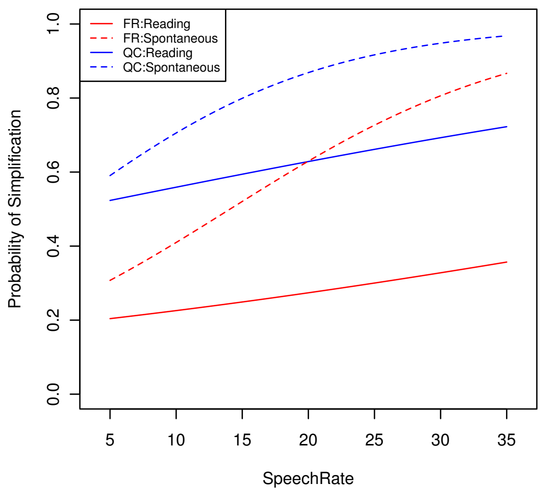

Figure 3.3: Regression Predicting Simplification as a Function of Style and Speech Rate

3.3.5 Summary for simplification

The results presented in this section suggest several interesting things, a few of which have not been previously explored in the literature on word-final consonant cluster simplification. In the first place, in all cases it was observed that the frequency of occurrences of word-final consonant cluster simplification were higher in the data from Québec than in the data from France. The general consensus from previous research is that word-final consonant cluster simplification is more frequently observed in the dialect from Québec than in the dialect from France. Some rates of word-final consonant cluster simplification for the dialect from France (at least for /ʁ/ and /l/ final clusters in a preconsonantal context) have been reported to typically be in the 48% to 58% range (c.f. Pustka (2011); Adda-Decker et al. (2002); Armstrong (1998)). Comparable rates for the dialect from Québec have been reported to be generally higher, perhaps closer to 60% to 70% range (c.f. Milne and Côté (2009); Kemp et al. (1980); Tousignant (1987)). In fact, a direct comparison between the two dialects using strictly comparable corpora and methodologies, such as was done here, has not been previously reported. The findings presented here, that in general word-final consonant cluster simplification is more frequently observed in the dialect from Québec than in the dialect from France contribute to the understanding of the dialectal differences between France and Québec with respect to word-final consonant cluster simplification.

Second, in these data at least, the group of attested word-final consonant clusters appeared to divide into two groups: A group in which simplification of the cluster was observed to regularly occur and a group in which simplification of the cluster was rarely, if ever, observed to occur. The group which did present observations of simplification included all SSP violating clusters as well as the stop-final clusters /kst, kt, pt, sk, st/. The group which did not present observations of simplification included all /ʁ/-initial clusters, as well as the stop+fricative combination /ks/. This division was observed in both dialects and follows closely the predictions from Côté (2004a) of the motivations for word-final consonant cluster simplification. The possibility that word-final consonant cluster simplification occurs more frequently than expected in the stop-final clusters in the dialect from France has not been previously explored. Most research on word-final consonant cluster simplification in the dialect of French from France has concentrated on just the SSP violating clusters, which are generally accepted as being susceptible to simplification, similar to the case of French in Québec.

Third, with respect to those consonant clusters that were observed to participate in simplification (the SSP violating and stop-final clusters) and related to the twin drivers of word-final consonant cluster simplification presented in Côté (2004a), violations of the SSP and perceptual salience, in this corpus it appears to be the case that the frequency of occurrences of simplification are higher for SSP violating clusters and lower for stop-final clusters than would be expected if the differences between the two types of clusters were due solely to chance. This difference between SSP violating and stop-final clusters was present in both the data from Québec as well as in the data from France.

Fourth, in these data it is suggested that the role of following context is also similar in both dialects. In both the data from Québec and in the data from France, the frequency with which a word-final consonant cluster simplifies is higher when the cluster is followed by a consonant-initial word and lower when when the cluster is followed by a vowel-initial word than would be expected if the differences between these two contexts was due only to chance. This result confirms earlier conclusions that, in the dialect from France, word-final consonant cluster simplification is likely to be more frequently observed when the cluster is preconsonantal than when it is prevocalic. However, for the dialect from Québec, while it was expected that word-final consonant cluster simplification would be free to occur in any of the three contexts explored here (preconsonantal, prevocalic, prepausal), it may in fact be more likely to occur preconsonantally than prevocalically. This mirrors the expectations for the dialect from France.

Finally, and perhaps most important to the topic of the current research, the analysis of covariance that tried to determine whether there existed differences between speech styles after taking into account differences in speech rate between the two speech styles resulted in rejecting the hypotheses of coincidence and parallelism for both dialects. The interpretation being that, after taking into account differences in speech rate, there were significant differences between a reading and spontaneous speech style and these differences can be explained as the different sensitivities to speech rate between the two styles: While reading, no effect of speech rate was obtained but while speaking spontaneously, higher rates of speech were associated with higher probabilities of word-final consonant cluster simplification.

Taken together, these results suggest that, even though the global rates of simplification are significantly different in the two dialects, the phonology of word-final consonant cluster simplification may be more similar between these two dialects than has previously been suspected.

3.4 Schwa in the corpus

To begin, the data indicate that schwa insertion following a word-final consonant cluster is observed in different groups of clusters in each dialect. Table 3.20 shows the overall proportion of observations of schwa insertion following a word-final consonant cluster for each cluster in each dialect. In the dialect from France, schwa insertion regularly occurs following almost every word-final consonant cluster in these data. Only the stop+fricative cluster /ks/ was not observed to participate in schwa insertion. This contrasts with the dialect from Québec where schwa insertion following a word-final consonant cluster was observed to regularly occur in all SSP violating clusters; was observed to less frequently occur in stop-final clusters; and was only observed following /ʁm/ and /ʁt/ for those clusters identified as being stable with respect to simplification.

		kst	kt	pt	sk	st								
Stop-final	QC	0.00	0.01	0.05	0.39	0.01								
	FR	0.39	0.23	0.26	0.72	0.36								
		bʁ	bl	dʁ	gl	kl	pʁ	pl	tʁ	stʁ	ʁdʁ	fʁ	vʁ	sm
SSP	QC	0.11	0.12	0.09	0.37	0.14	0.28	0.13	0.11	0.09	0.16	0.09	0.10	0.32
	FR	0.43	0.39	0.28	0.47	0.50	0.68	0.26	0.32	0.19	0.33	0.40	0.34	0.50
		ʁʃ	ʁs	ʁv	ʁz	ʁʒ	ʁn	ʁm	ʁd	ʁk	ʁt	ks		
Stable	QC	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.01	0.00	0.16	0.00		
	FR	0.21	0.20	0.45	0.19	0.23	0.27	0.22	0.29	0.17	0.30	0.00		

Table 3.20: Proportion of schwa insertion for clusters by Dialect

Analogous to the marked difference between the dialects observed with respect to word-final consonant cluster simplification, schwa insertion following a word-final consonant cluster is more common in the dialect from France than it is in the dialect from Québec. Approximately 30% of the occurrences of word-final consonant clusters in these data had schwa inserted in the data from France compared with just 8% in the data from Québec. With an odds ratio of 4.83, schwa insertion following a word-final consonant cluster is nearly 5 times more likely to occur in the data from France than in the data from Québec.

Schwa	FR	QC	Sum
0	8579 (69.56%)	14450 (91.68%)	23029
1	3755 (30.44%)	1311 (8.32%)	5066
Sum	12334	15761	28095

Table 3.21: Frequency and proportion of schwa insertion according to dialect

3.4.1 Differences in schwa insertion according to following context

For schwa insertion following a word-final consonant cluster, with respect to the dialect from France, most of the literature reviewed concludes that the following context should be expected to influence the frequency of occurences of schwa. The historical progression detailing the loss of word-final schwa in France gives rise to the notion expressed by Dell (1985) that word-final schwa exists primarily to break up consonant clusters. Côté (2000) argues that the segmental nature of the particular sequence of sounds in which the word-final consonant cluster occurs produces a cumulative effect on the probability of schwa being inserted or not. In particular, whether the consonant cluster violates the SSP, whether it ends with a final stop consonant and whether it is adjacent to a vowel. Similar to the current study, Bürki et al. (2011b) used forced alignment on a large set of recorded data and concluded that schwa insertion was more common before consonants than before vowels. The Table 3.22, which presents both the frequency and proportion of observations of schwa insertion following a word-final consonant cluster in each dialect according to following context, suggests that schwa insertion following a word-final consonant cluster appears to be more likely before a consonant-initial word or a pause than before a vowel-initial word in the data from France, while schwa insertion appears to be equally (un)likely in all three contexts in the data from Québec.

Schwa		Consonant	Vowel	Pause	Sum
QC	0	9288 (91.58%)	3049 (92.93%)	2113 (90.38%)	14450
	1	854 (8.42%)	232 (7.07%)	225 (9.62%)	1311
FR	0	4957 (63.90%)	2124 (87.48%)	1498 (69.74%)	8579
	1	2801 (36.10%)	304 (12.52%)	650 (30.26%)	3755
Sum		17900	5709	4486	28095

Table 3.22: Frequency and proportion of schwa insertion according to context

For word-final consonant clusters that appeared in a preconsonantal context, schwa insertion was observed in 36% of the cases in the data from France and just 8% of the cases in the data from Québec. In the data from France, these rates dropped to 30% when the cluster was prepausal, and dropped further to 12% when prevocalic. This contrasts with the situation in the data from Québec where the rates of schwa insertion following a word-final consonant cluster appear not to change according to the context following the cluster: 9% when prepausal and 7% when prevocalic. Although it is conceivable that these differences in the dialect from Québec were due to chance, that doesn't seem likely in the dialect from France. A chi-square test of the null hypothesis that schwa insertion following a word-final consonant cluster is independent of the following context will help to answer the question of whether following context plays a role in the variable pronunciations of word-final consonant clusters in either dialect. Table 3.23 displays the observed and expected frequencies for simplification as a function of context for the data from Québec. Table 3.24 displays the similar information for the data from France.

	Observed			Expected		
	Schwa			Schwa		
Context	Yes	No	Total	Yes	No	Total
C	854	9288	10142	843.61	9298.39	10142
V	232	3049	3281	272.91	3008.09	3281
P	225	2113	2338	194.47	2143.53	2338
Total	1311	14450	15761	1311	14450	15761

Table 3.23: Observed and expected frequencies for schwa insertion as a function of context for Québec

In these data from Québec, the obtained value of $\chi^2(2) = 12.06$ is significantly larger than than the critical value of $\chi^2(2) = 5.99$ and the null hypothesis must be rejected. While the observed frequency of schwa insertion when the word-final consonant cluster was preconsonantal is approximately what should be expected, there were fewer observations of schwa insertion when the cluster was prevocalic, and more observations of schwa insertion when the cluster was prepausal than would be expected.

	Observed			Expected		
	Schwa			Schwa		
Context	Yes	No	Total	Yes	No	Total
C	2801	4957	7758	2361.87	5396.13	7758
V	304	2124	2428	739.19	1688.81	2428
P	650	1498	2148	653.94	1494.06	2148
Total	3755	8579	12334	3755	8579	12334

Table 3.24: Observed and expected frequencies for schwa insertion as a function of context for France

In these data from France, the obtained value of $\chi^2(2) = 485.77$ is significantly larger than than the critical value of $\chi^2(2) = 5.99$ and the null hypothesis must again be rejected. While the observed frequency of schwa insertion when the word-final consonant cluster was prepausal is approximately what should be expected, there were fewer observations of schwa insertion when the cluster was prevocalic, and more observations of schwa insertion when the cluster was preconsonantal than would be expected. Therefore, it would appear that, in these data, schwa insertion following a word-final consonant cluster may occur in all three contexts (preconsonantal, prevocalic, prepausal) in both dialects, although the frequencies of schwa insertion are much higher in the data from France for all three. The large difference in obtained χ^2 values indicates that the effect of following context is much larger in the dialect from France than in the dialect from Québec. While in both dialects when a word-final consonant cluster occurred in a prevocalic context the observed frequencies of schwa insertion were lower than would be expected, only in the dialect from France were there significantly more observations of schwa insertion when the cluster was preconsonantal. This, combined with the larger overall effect of following context in the data from France, suggests that the dialect from France may be more sensitive to following context, especially preconsonantal, than the dialect from Québec. However, these results indicate that, in both dialects, schwa insertion following a word-final consonant cluster is not independent of following context.

3.4.2 Differences in schwa according to simplification

One of the interesting observations that arose from the earlier study of Milne and Côté (2009) was that the distributions of schwa insertion following a word-final consonant cluster were very different in the dialect from Québec when compared with the dialect from France. In that study, it was noted that, in the dialect from France, schwa insertion following a word-final consonant cluster was observed to occur following any consonant cluster, while in the dialect from Québec, schwa insertion was observed to occur mainly following just those consonant clusters that also participated in word-final consonant cluster simplification. In their data, nearly 95% of observed cases of schwa insertion in the dialect from Québec occurred in SSP violating or stop-final consonant clusters. In those same consonant clusters in the dialect from France they found approximately 50% of observed cases of schwa insertion. Their hypothesis was that, in the dialect from Québec, schwa insertion following a word-final consonant cluster was motivated as an avoidance of simplification while in the dialect from France, schwa insertion was motivated as a general principle avoiding sequences of consonants. Table 3.20 suggests that the distribution of schwa insertion observed by Milne and Côté (2009) may be replicated in these current data. Almost every consonant cluster investigated in these data, whether they were a stop-final, SSP violating, or stable cluster, regularly occurred with a final schwa vowel in the dialect from France. In the dialect from Québec, of those clusters identified as being stable with respect to simplification, only two of the eleven attested consonant clusters (/ʁt/ and /ʁm/) appear to show any occurrences of schwa insertion. Table 3.25, which presents both the frequency and proportion of observations of schwa insertion following a word-final consonant cluster in each dialect according to whether the consonant cluster participated in simplification (Simplify) or not (Stable), suggests that schwa insertion following a word-final consonant cluster appears to be less frequent following stable consonant clusters in both dialects, though the difference may be greater in the dialect from Québec.

Dialect	Schwa	Simplify	Stable	Sum
QC	0	10493 (89.96%)	3957 (96.58%)	14450
	1	1171 (10.04%)	140 (3.42%)	1311
FR	0	6158 (66.64%)	2421 (78.27%)	8579
	1	3083 (33.36%)	672 (21.73%)	3755
Total		7190	20905	28095

Table 3.25: Frequency and proportion of schwa insertion according to simplification

In the data from Québec, the proportion of schwa insertion following a word-final consonant cluster is higher in the Simplify category (10%) than in the Stable category (3%). This supports the findings of Milne and Côté (2009). In fact, nearly 90% of the observed cases of schwa insertion (1171 of 1311) occurred following SSP violating or stop-final clusters in the dialect from Québec. A proportion very similar to that found earlier. However, in the data from France, where the proportion of schwa insertion following a word-final consonant cluster was expected to be approximately evenly distributed between these two categories (Simplify and Stable), these data show that, similar to the dialect from Québec, schwa insertion is more frequently observed in the Simplify category (33%) than in the Stable category (21%) of consonant clusters. In fact, approximately 82% of the observed cases of schwa insertion (3083 of 3755) occurred following SSP violating or stop-final consonant clusters in the dialect from France. This unexpected result is most likely due to the fact that Milne and Côté (2009) grouped the stop-final consonant clusters with the Stable category while in the current study, since their rates of simplification have been shown to be significantly higher than the other Stable clusters, they have been grouped with the Simplify category. A chi-square test of the null hypothesis that schwa insertion following a word-final consonant cluster is independent of the distinction between Simplify and Stable consonant clusters will help to answer the question of whether schwa insertion following a word-final consonant cluster is an avoidance of simplification in either dialect. Table 3.26 displays the observed and expected frequencies for schwa as a function of simplification for the data from Québec. Table 3.27 displays the similar information for the data from France.

	Observed			Expected		
	Schwa			Schwa		
Simplification	Yes	No	Total	Yes	No	Total
Simplify	1171	10493	11664	970.21	10693.79	11664
Stable	140	3957	4097	340.79	3756.21	4097
Total	1311	14450	15761	1311	14450	15761

Table 3.26: Observed and expected frequencies for schwa insertion as a function of simplification for Québec

In these data from Québec, the obtained value of $\chi^2(1) = 174.36$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must be rejected. There were significantly more observations of schwa insertion following SSP violating and stop-final clusters (Simplify) and significantly fewer observations of schwa insertion following the other clusters (Stable) than would be expected if the differences between these two categories were due simply to chance.

	Observed			Expected		
	Schwa			Schwa		
Simplification	Yes	No	Total	Yes	No	Total
Simplify	3083	6158	9241	2813.36	6427.64	9241
Stable	672	2421	3093	941.64	2151.36	3093
Total	3755	8579	12334	3755	8579	12334

Table 3.27: Observed and expected frequencies for schwa insertion as a function of simplification for France

In these data from France, the obtained value of $\chi^2(1) = 148.16$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must be rejected. Similar to the results obtained when looking at the data from Québec, there were significantly more observations of schwa insertion following SSP violating and stop-final clusters (Simplify) and significantly fewer observations of schwa insertion following the other clusters (Stable) than would be expected if the differences between these two categories were due simply to chance. These results provide evidence for the hypothesis that schwa insertion following a word-final consonant cluster is more likely to occur if the consonant cluster also participates in word-final consonant cluster simplification. The effect, while slightly stronger in the dialect from Québec, is present in both dialects.

3.4.3 Differences in schwa according to motivation for simplification

If schwa insertion following a word-final consonant cluster is more likely to occur if the consonant cluster also participates in word-final consonant cluster simplification, could the motivation for simplification also play a role? In a small sense, this is akin to inquiring about the effect of preceding context, at least for the SSP violating and stop final clusters present in the corpus. In general, it is assumed that for the dialect from France, the preceding context is not expected to contribute much to the understanding of schwa insertion following a word-final consonant cluster. However, Kemp et al. (1980) note that, for the dialect from Québec, in their data when schwa insertion did occur, it was principally following /ʁ/ and /l/ and variably following stops and nasals. This happens to coincide with the SSP violating and stop-final consonant clusters investigated here. It was previously determined that word-final consonant cluster simplification was more likely if the consonant cluster violated the SSP than if the consonant cluster was a stop-final consonant cluster. Looking only at those clusters that participated in word-final consonant cluster simplification, Table 3.28, which presents both the frequency and proportion of observations of schwa insertion following a word-final consonant cluster in each dialect according to the motivation for simplification (SSP or Stop-Final), suggests that schwa insertion is more common with clusters that violate the SSP than with the stop-final clusters in the data from Québec, while this difference may be smaller, or not exist at all, in the data from France.

Dialect	Schwa	Stop-final	SSP	Sum
QC	0	1843 (96.44%)	8650 (88.69%)	10493
	1	68 (3.56%)	1103 (11.31%)	1171
FR	0	916 (61.77%)	5242 (67.57%)	6158
	1	567 (38.23%)	2516 (32.43%)	3083
Total		3394	17511	20905

Table 3.28: Frequency and proportion of schwa insertion according to motivation for simplification

In the data from Québec, the proportion of schwa insertion following a word-final consonant cluster is higher for SSP violating clusters than for stop-final clusters (11% $\sim$ 3%). However, in the data from France, the proportion of schwa insertion for SSP violating clusters is lower than the proportion of schwa insertion for the stop-final clusters (32% $\sim$ 38%). A chi-square test of the null hypothesis that schwa insertion following a word-final consonant cluster is independent of the motivation for simplification will help to answer the question of whether the motivation for simplification plays a role in schwa insertion following a word-final consonant cluster in either dialect. Table 3.29 displays the observed and expected frequencies for schwa as a function of motivation for the data from Québec. Table 3.30 displays the similar information for the data from France.

	Observed			Expected		
	Schwa Insertion			Schwa Insertion		
Motivation	Yes	No	Total	Yes	No	Total
SSP	1103	8650	9753	979.15	8773.85	9753
Stop-final	68	1843	1911	191.85	1719.15	1911
Total	1171	10493	11664	1171	10493	11664

Table 3.29: Observed and expected frequencies for schwa insertion as a function of motivation for Québec

In these data from Québec, the obtained value of $\chi^2(1) = 106.29$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must be rejected. There are significantly more observations of schwa insertion following a word-final consonant cluster when the motivation for simplification is a violation of the SSP and significantly fewer observations when the motivation for simplification is perceptual salience (stop-final) than would be expected.

	Observed			Expected		
	Schwa Insertion			Schwa Insertion		
Motivation	Yes	No	Total	Yes	No	Total
SSP	2516	5242	7758	2588.24	5169.76	7758
Stop-final	567	916	1483	494.76	988.24	1483
Total	3083	6158	9241	3083	6158	9241

Table 3.30: Observed and expected frequencies for schwa insertion as a function of motivation for France

In these data from France, the obtained value of $\chi^2(1) = 18.85$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must again be rejected. Different from what was seen in the data from Québec, there are significantly fewer observations of word-final consonant cluster simplification when the motivation for simplification is a violation of the SSP and significantly more observations when the motivation for simplification is perceptual salience (stop-final) than would be expected. The large difference between the obtained χ^2 values for each dialect indicates that the effect of motivation for simplification on schwa insertion is much stronger in the dialect from Québec than in the dialect from France. Additionally, the direction of the effect is opposite: There were significantly fewer observations of schwa insertion following a stop-final consonant cluster in the dialect from Québec, while there were significantly more observations of schwa insertion in these consonant clusters in the dialect from France.

3.4.4 Differences in schwa according to style: Reading vs Spontaneous

The literature review contains some examples of researchers who have found an effect of speech style on schwa insertion. It is possible that schwa insertion tends to be more common in standard, as compared to regional, varieties of French (Durand and Eychenne, 2004; Eychenne, 2006). Within the dialect from France, it has been observed that the likelihood of word-final schwa insertion in read speech could be twice as high as in spontaneous speech and that in spontaneous speech, the majority of word-final schwas are not pronounced (Adda-Decker et al., 1999). Both Eychenne (2003) and Kemp et al. (1980), while examining schwa at word boundaries in speakers from Québec, found that schwa at word boundaries in the vernacular was extremely unlikely, while in a more formal speech style, including examples of read speech, Milne and Côté (2009) found higher than expected frequencies of schwa insertion following word-final consonant clusters in their data from Québec. In the data displayed in Table 3.31, schwa insertion following a word-final consonant cluster appears to be more frequent in a reading style of speech than in a spontaneous style of speech and this appears to be the case in both dialects.

Dialect	Schwa	Reading	Spontaneous	Sum
QC	0	2489 (82.39%)	11961 (93.89%)	14450
	1	532 (17.61%)	779 (6.11%)	1311
FR	0	2750 (58.31%)	5829 (76.52%)	8579
	1	1966 (41.69%)	1789 (23.48%)	3755
Total		7737	20358	28095

Table 3.31: Frequency and proportion of schwa insertion according to style

In the data from Québec, approximately 17% of the occurrences of word-final consonant clusters from the reading style of speech contained a final schwa vowel compared with just 6% in the data from the spontaneous style of speech. In the data from France, approximately 42% of the occurrences of word-final consonant clusters from the reading style of speech contained a final schwa vowel compared with 23% in the data from the spontaneous style of speech. While it is not likely that differences this large will be due to chance, a chi-square test of the null hypothesis that schwa insertion following a word-final consonant cluster is independent of style will help to answer the question of whether speech style plays a role in schwa insertion in either dialect. Table 3.32 displays the observed and expected frequencies for schwa as a function of style for the data from Québec. Table 3.33 displays the similar information for the data from France.

	Observed			Expected		
	Schwa Insertion			Schwa Insertion		
Style	Yes	No	Total	Yes	No	Total
Reading	532	2489	3021	251.29	2769.71	3021
Spontaneous	779	11961	12740	1059.71	11680.29	12740
Total	1311	14450	15761	1311	14450	15761

Table 3.32: Observed and expected frequencies for schwa insertion as a function of style for Québec

In these data from Québec, the obtained value of $\chi^2(1) = 423.14$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must be rejected. There are significantly more observations of schwa insertion in the reading style of speech than would be expected and significantly fewer observations in the spontaneous style of speech than would be expected.

In these data from France, the obtained value of $\chi^2(1) = 455.83$ is significantly larger than than the critical value of $\chi^2(1) = 3.84$ and the null hypothesis must again be rejected. As was seen in the data from Québec, there are significantly more observations of schwa insertion in the reading style of speech than would be expected and significantly fewer observations in the spontaneous style of speech than would be expected. Therefore, it would appear that, in these data, schwa insertion following a word-final consonant cluster is more frequent in both spontaneous and reading styles in the data from

	Observed			Expected		
	Schwa Insertion			Schwa Insertion		
Style	Yes	No	Total	Yes	No	Total
Reading	1966	2750	4716	1531.37	2538.63	4070
Spontaneous	1789	7618	5171	1945.63	3225.37	5171
Total	3755	8579	12334	3755	8579	12334

Table 3.33: Observed and expected frequencies for schwa insertion as a function of style for France

France, and the effect of style is similar in both dialects: Observations of schwa insertion following a word-final consonant cluster are more frequent than expected in the reading speech style, and less frequent than would be expected in the spontaneous speech style. In both dialects, schwa insertion following a word-final consonant cluster is not independent of speech style.

3.4.5 Speech style or speech rate as explanatory variable for schwa

It has already been explained that rates of speech were different in each style: Speech rates were higher in the spontaneous style of speech and lower in the reading style of speech. Speech rates were also significantly faster when word-final consonant cluster simplification had occurred than when it had not occurred. Table 3.34, which shows the mean and standard deviation of speech rates in each dialect according to whether schwa insertion occurred or not suggests that the opposite may be true with respect to schwa insertion. An analysis of variance of these data indicates no significant effect of dialect ($F(1, 28091) = 0.474, p > 0.05$). There is an effect of schwa ($F(1, 28091) = 240.848, p < 0.001$). In these data, speech rates are slower when schwa insertion has occurred than when it has not occurred. There is no significant interaction between dialect and schwa ($F(1, 28091) = 3.604, p > 0.05$).

		Mean	SD
QC	No Schwa	13.23	2.641
	Schwa	12.48	2.413
FR	No Schwa	13.37	2.526
	Schwa	12.78	2.277

Table 3.34: Mean and standard deviation of speech rate by schwa according to dialect

The fact that speech rate has an effect on schwa insertion following a word-final consonant cluster is shown in Figure 3.4 which predicts the probability of schwa insertion as a function of speech rate. The relationship described by this regression indicates that for every one unit increase in speech rate, the odds of observing schwa insertion decrease by a factor of 0.914.

The table of coefficients for the model predicting the probability of schwa insertion by style and rates of speech in the data from Québec is given below in Table 3.35. This table indicates that only the

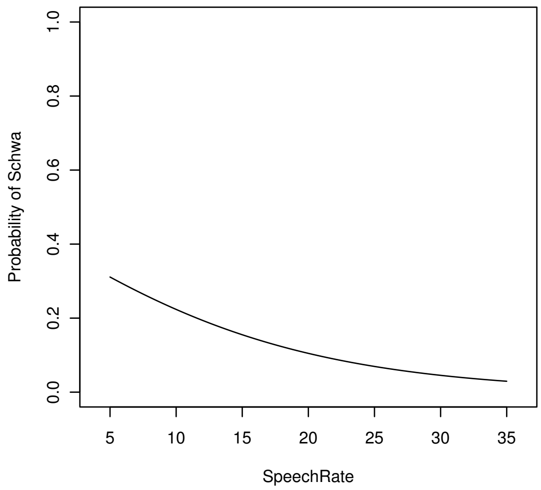

Figure 3.4: Regression predicting schwa insertion as a function of speech rate

interaction term between style and speech rate appears to be highly significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−1.479312	0.263479	−5.615	$1.97e-08$	***
Spontaneous	0.364360	0.320531	1.137	0.256	
Rate	−0.005129	0.020882	−0.246	0.806	
Spontaneous:Rate	−0.119620	0.025258	−4.736	$2.18e-06$	***

Table 3.35: Coefficients for model predicting schwa insertion by style and rate for Québec

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.36, indicates two things. First, the difference between the null deviance and the residual deviance approximately follows a chi-squared distribution with as degrees of freedom the difference between the degrees of freedom of the two deviances. Thus, with $\chi^2(3) = 433.8, p < 0.001$, indicates that this model has explanatory value. Second, the small `Pr(>Chi)` values indicates that each term in the model is predictive.

Testing for coincidence, that is, $H_0 : \beta_1 = \beta_3 = 0$, is simply testing significance of the terms x and $x \times z$, simultaneously. This can be done by removing those terms from the model and evaluating

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			15760	9030.0	
Style	1	354.75	15759	8675.3	$< 2.2e-16$
Rate	1	56.73	15758	8618.6	$5.001e-14$
Style:Rate	1	22.34	15757	8596.2	$2.289e-06$

Table 3.36: Analysis of deviance table for Québec

whether they add much to the understanding of the data. Comparing the simplified model with just a single term for `Rate` with the initial model returns $\chi^2(2) = 333.07, p < 0.001$. This says that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. Therefore, it seems the two regression lines are not coincident.

Testing for parallelism, that is $H_0 : \beta_3 = 0$, is testing significance of the term $x \times z$. Removing the interaction term and comparing this simplified model with terms only for `Style` and `Rate` with the initial model returns $\chi^2(1) = 22.336, p < 0.001$. This says again that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. The conclusion is that the two regression lines are not parallel.

The table of coefficients for the model predicting the probability of schwa insertion by style and rates of speech in the data from France is given below in Table 3.37. This table indicates that the term for speech rate and the interaction term between style and speech rate appear to be highly significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.17650	0.18540	0.952	0.341092	
Spontaneous	-0.02029	0.23143	-0.088	0.930126	
Rate	-0.04022	0.01439	-2.795	0.005192	**
Spontaneous:Rate	-0.06051	0.01774	-3.411	0.000648	***

Table 3.37: Coefficients for model predicting schwa insertion by style and rate for France

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.38, indicates two things. First, the difference between the null deviance and the residual deviance approximately follows a chi-squared distribution with as degrees of freedom the difference between the degrees of freedom of the two deviances. Thus, with $\chi^2(3) = 552, p < 0.001$, indicates that this model has explanatory value. Second, the small `Pr(>Chi)` values indicates that each term in the model is predictive.

Testing for coincidence, that is, $H_0 : \beta_1 = \beta_3 = 0$, is simply testing significance of the terms x and $x \times z$, simultaneously. This can be done by removing those terms from the model and evaluating whether they add much to the understanding of the data. Comparing the simplified model with just a single term for `Rate` with the initial model returns $\chi^2(2) = 403.73, p < 0.001$. This says that this model

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			12333	15160	
Style	1	449.20	12332	14711	$< 2.2e-16$
Rate	1	92.23	12331	14619	$< 2.2e-16$
Style:Rate	1	11.62	12330	14608	0.0006519

Table 3.38: Analysis of deviance table for France

simplification is not justified because it caused a significant reduction in the explanatory power of the model. Therefore, it seems the two regression lines are not coincident.

Testing for parallelism, that is $H_0 : \beta_3 = 0$, is testing significance of the term $x \times z$. Removing the interaction term and comparing this simplified model with terms only for `Style` and `Rate` with the initial model returns $\chi^2(1) = 11.622, p < 0.001$. This says again that this model simplification is not justified because it caused a significant reduction in the explanatory power of the model. The conclusion is that the two regression lines are not parallel.

Very much like the results presented in Section 3.3.4, the very similar models that emerged for each dialect suggests that the interpretation of these models for each dialect is also similar. Because the hypothesis of coincidence was rejected in both dialects it must be accepted that, in both the data from Québec and France, after controlling for rates of speech there exist significant differences between a reading and spontaneous speech style with respect to schwa insertion following a word-final consonant cluster and it is right to consider speech style as an explanatory variable in both dialects. Since the hypothesis of parallelism was also rejected in both dialects, it must be accepted that the effect of speech rate on schwa insertion is not the same in each style which gives further confidence to considering speech rate as an explanatory variable in both dialects. While reading, there is no effect of speech rate on schwa insertion following a word-final consonant cluster in the dialect from Québec. The model suggests there may be an effect of speech rate for the reading style in the dialect from France. The coefficient for `Rate` indicates that for every one unit increase in speech rate, the odds of schwa insertion occurring in this style decrease by a factor of 0.961. This is different from the effect of speech rate in the spontaneous speech style. Freed from the demands of reading from a text, the odds of observing schwa insertion following a word-final consonant cluster decrease as rates of speech increase. This is true of both dialects. These relationships are presented in Figure 3.5.

3.4.6 Summary for schwa

The results presented in this section suggest several interesting things, a few of which have not been previously explored in the literature on schwa insertion following a word-final consonant cluster. In the first place, in all cases it was observed that the frequency of occurrences of schwa insertion were higher in the data from France than in the data from Québec. This confirms the general consensus of most previous research that schwa insertion following a word-final consonant cluster is more frequently observed in the dialect from France than in the dialect from Québec.

Second, in these data from Québec, the group of investigated word-final consonant clusters appeared

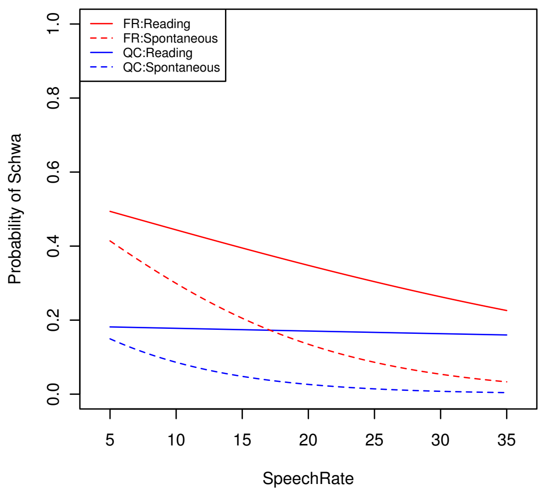

Figure 3.5: Regression predicting schwa insertion as a function of style and speech rate

to divide into two groups: A group in which schwa insertion was observed to occur and a group in which schwa insertion was rarely, if ever, observed to occur. The group which did present observations of schwa insertion included all SSP violating clusters as well as most of the stop-final clusters /kt, pt, sk, st/. The only stop-final consonant cluster which did not present any observations of schwa insertion in the data from Québec was /kst/. The group which did not present observations of schwa insertion included almost all /ʁ/-initial clusters, as well as the stop+fricative combination /ks/. Note that /ks/ corresponds to the single letter 'x'. This is probably not a coincidence. From the group of /ʁ/-initial consonant clusters, just two (/ʁm, ʁt/) were observed with any occurrences of schwa insertion. This division is nearly identical to the one described on page 90 for word-final consonant cluster simplification. This division also provides confirmation for the conclusions expressed in Milne and Côté (2009) that schwa insertion following a word-final consonant cluster in the dialect from Québec almost always occurs in just those consonant clusters that also participate in simplification. In contrast, in these data from France, schwa insertion was observed following almost every word-final consonant cluster. Only the stop+fricative cluster /ks/ was not observed to participate in schwa insertion. This is in line with most other research on word-final schwa insertion in France which does not predict constraints on schwa insertion according to any specific consonant cluster or group of consonant clusters. This result also confirms the findings from Milne and Côté (2009) in which schwa insertion was also observed following all word-final consonant clusters in their data from France. However, while almost all word-final consonant clusters participated in schwa insertion in these data from France, the results from Section 3.4.2 suggest that, even in the dialect from France, schwa insertion is more likely following a word-final consonant cluster that also participated in simplification. This result, while surprising, has not been previously explored in the literature on word-final schwa insertion in the dialect from France and may suggest that, in both dialects, schwa insertion following a word-final consonant cluster may exist as an avoidance of simplification even if it also exists as a general principle to break up consonant clusters only in the dialect from France.

Third, following the observation that, in this corpus, schwa insertion appeared to be more likely following a consonant cluster that also participated in simplification, it was determined that the motivation for simplification (avoidance of sonority violations or avoidance of perceptually deficient final stop consonants) may also be playing a role in the frequency of occurrences of schwa insertion. The results expressed in Section 3.4.3 indicated that, even though the proportion of schwa insertions was higher in both Simplify and Stable consonant clusters in France than in Québec, an effect of the motivation for simplification was observed in both dialects. In the dialect from Québec, there were significantly fewer observations of schwa insertion following a word-final consonant cluster when the consonant cluster contained a final stop consonant than would be expected if the differences between the two types of consonant clusters were attributed solely to chance. In the dialect from France, the opposite was observed: More observations of schwa insertion following a stop-final consonant cluster that would be expected. This result could, perhaps, be interpreted in view of the hypothesis that, in Québec, schwa insertion exists mainly as an avoidance of simplification. The effect of motivation for simplification was observed to be present in both dialects with respect to the occurrence of simplification: The frequency of simplification was lower in stop-final consonant clusters than in SSP violating clusters in both dialects. The fact that lower rates of word-final consonant cluster simplification are associated with lower rates of schwa insertion in stop-final consonant clusters in Québec could be taken as further evidence in support of the schwa insertion as avoidance of simplification hypothesis. The higher rates of schwa insertion in these stop-final consonant clusters in the data from France may be due to the added effect of the avoidance of (any) word-final consonant cluster.

Fourth, regarding the effect of following context on schwa insertion following a word-final consonant cluster, it appears to be the case that, in both dialects, schwa insertion is less likely when the consonant cluster is followed by a vowel-initial word than by either a consonant-initial word or a pause. This result is in agreement with many other researchers who have found that the additional support provided by a following vowel makes schwa insertion less likely to occur. This result also suggests that, when interpreted along with the results of following context on word-final consonant cluster simplification, a preconsonantal environment is more conducive to either schwa insertion or simplification than is a prevocalic environment.

Finally, the analysis of covariance that tried to determine whether there existed differences between speech styles after taking into account differences in speech rate between the two speech styles resulted in rejecting the hypotheses of coincidence and parallelism for both dialects. The interpretation being that, after taking into account differences in speech rate, there were significant differences between a reading and spontaneous speech style and these differences were interpreted as the different sensitivities to speech rate between the two styles: While reading, little or no effect of speech rate was obtained but while speaking spontaneously, higher rates of speech were associated with lower probabilities of schwa insertion following a word-final consonant cluster. This was true of both dialects.

3.5 Generalized linear models

The research questions being investigated in this chapter are concerned with the possible effects of speech style, following context, motivation for simplification and speech rate on the variable pronunciations of word-final consonant clusters in two dialects of French. Of interest is whether the direction and strength of the relationship between these variables and either simplification or schwa insertion following a word-final consonant cluster are different according to different dialects. A related question concerns whether or not schwa insertion following a word-final consonant cluster is primarily used as an avoidance of simplification for speakers of French from Québec while schwa insertion following a word-final consonant cluster is primarily used as an avoidance of any sequence of consonants for speakers of French from France. The results presented thus far in Sections 3.3 on simplification of the word-final consonant cluster and 3.4 on schwa insertion following a word-final consonant cluster have provided an opportunity to reflect on several intuitions about what the answers to these research questions might entail.

In the first place, while some of the observations presented while exploring either simplification or schwa insertion following a word-final consonant cluster have not been previously reported in the literature, none of the results contradict what previous researchers have concluded with respect to these variable pronunciations of word-final consonant clusters.

It was expected that the frequency of occurrences of word-final consonant cluster simplification would be higher in the dialect from Québec than in the dialect from France and that the frequency of occurrences of schwa insertion following a word-final consonant cluster would be higher in the dialect from France than in the dialect from Québec. It was shown that, in all cases in this corpus, this was true.

Given the proposal in Côté (2004a) it was expected that, for the dialect from Québec, the group of word-final consonant clusters investigated in this corpus would divide into two groups: A group that would participate in simplification (SSP violating and stop-final consonant clusters), and a group that would remain stable and not participate in simplification (all /ʁ/-initial consonant clusters, as well as /ks/). This was indeed found to be the situation for the dialect from Québec. The fact that this same division into Simplify or Stable consonant clusters (with respect to simplification) appeared also in the dialect from France might, on the surface, appear surprising. However, the literature review indicated that most previous research on word-final consonant cluster simplification in the dialect from France has restricted itself to just the SSP violating clusters (those ending in /ʁ/ or /l/), which have long been known to participate in simplification for that dialect. Therefore, suggesting that word-final consonant cluster simplification may also occur in stop-final clusters in the dialect from France merely adds new information to the understanding of word-final consonant cluster simplification in that dialect and does not contradict the earlier conclusions concerning the SSP violating consonant clusters.

The hypothesis advanced in Milne and Côté (2009) that schwa insertion would be much more frequently observed in SSP violating and stop-final consonant clusters for the dialect from Québec was confirmed in this corpus. It was expected that schwa insertion following a word-final consonant cluster in the dialect from France would be regularly observed in all the clusters investigated in this corpus. In fact, in this corpus, only the stop+fricative consonant cluster /ks/ did not provide any observations of schwa insertion for this dialect. The new information provided by this investigation suggests that, while schwa insertion does indeed occur following almost any word-final consonant cluster in the dialect from France, it appears that, as is the case in the dialect from Québec, schwa insertion is not independent of whether or not a consonant cluster also participates in simplification. In this corpus it was shown that the frequency of observations of schwa insertion were higher than expected in SSP violating and stop-final clusters (Simplify) than in the other consonant clusters (Stable) if the differences between Simplify and Stable clusters were due to chance. That there may exist a difference in rates of schwa insertion between Simplify and Stable word-final consonant clusters in the dialect from France builds upon, and adds new information to, the existing research on word-final schwa insertion.

With respect to the role of following context, it was expected that the frequency of occurences of both simplification and schwa insertion following a word-final consonant cluster would be higher when the consonant cluster appears in a preconsonantal context than when it appears in a prevocalic context. This has generally been understood to illustrate the supportive effects a prevocalic environment lends to both the production and perception of word-final consonants. In this corpus it has been demonstrated that a preconsonantal context was associated with higher than expected rates of both simplification and schwa insertion following a word-final consonant cluster in both dialects.

Combined, these intuitions should give confidence that the corpus utilized here, since it agrees in many important and previously recognized ways, should be considered as representative of the speech patterns in these two communities of speakers, France and Québec, with respect to the variable pronunciations of word-final consonant clusters. Given that level of confidence, the following investigation of the possible effects of speech style on the variable pronunciations of word-final consonant clusters in two dialects of French should also likely be representative of the speech patterns within each speech style investigated in these two dialects.

The initial conclusions regarding whether both speech style and speech rate should be considered as explanatory variables with respect to both simplification and schwa insertion following a word-final consonant cluster suggested that there may be differences between speech styles, reading and spontaneous, in both dialects but it is unclear whether or not these differences indicate phonological differences. The important point to consider is that, in both dialects, there was little or no effect of speech rate on either simplification or schwa insertion (perhaps a small effect of rate on schwa insertion for the dialect from France) for the data identified as coming from read text (reading style), while the data identified as coming from a spontaneous speech style indicated that faster rates of speech were associated with higher rates of simplification and lower rates of schwa insertion. The interpretation of these results was that the observed differences between the two styles most likely reflects the different sensitivities to speech rate between the two styles and that the higher rates of simplification and lower rates of schwa insertion observed in the spontaneous speech style in both dialects could be due to the tendency to have faster rates of speech while speaking spontaneously without the use of text.

These intuitions suggest that, in agreement with much previous research on simplification or schwa insertion following a word-final consonant cluster, the variables under consideration here (whether or not a consonant cluster participates in simplification, whether a consonant cluster violates the SSP or is stop-final and the context in which the consonant cluster occurs, preconsonantal, prevocalic, or prepausal) all play a role in both explaining and predicting the variable pronunciations of word-final consonant clusters in both dialects. It seems clear that these variables all describe phonological effects. Given that the effects of speech style are inextricably combined with the effects of speech rate, it is unclear whether or not the variable of speech style also might play a phonological role. In order to better understand whether or not there truly exist differences between the speech styles in either dialect with respect to the variable pronunciations of word-final consonant clusters, a logistic regression model predicting the probability of the responses of either simplification or schwa insertion that includes terms for the expected effects of participation in simplification, motivation for simplification, and following context as well as a term for speech style while also including a covariate term for speech rate will determine whether or not differences exist between a reading and spontaneous speech style in either dialect with respect to the variable pronunciations of word-final consonant clusters after taking into account the differences in speech rate that exist between the two styles.

3.5.1 Simplification according to style, context, motivation for simplification and speech rate

The model predicting the response of simplification considers only those consonant clusters that were observed to participate in word-final consonant cluster simplification: SSP violating and stop-final consonant clusters. It includes terms for those variables that were suggested by the analyses in Section 3.3 to have explanatory value. The term Vowel contains two levels: 0 for the preconsonantal context and 1 for the prevocalic context. For simplification and ease of interpretation of the model, this term that categorizes the following context will be reduced to just preconsonantal and prevocalic and will exclude all observations of word-final consonant clusters that occurred in the prepausal context. The term that describes the motivation for simplification, SSP, likewise contains two levels: 0 for stop-final consonant clusters and 1 for SSP violating clusters. The categorical predictor for speech style includes two levels: Reading and Spontaneous. The covariate term for speech rate is a continuous variable. Additionally, since it was observed that the effect of speech rate may likely be constrained to have an effect on just the spontaneous speech style, an interaction term between style and speech rate is included.

In these models, all two-level categorical predictors have been sum coded (e.g. for speech style, read speech = -0.5 and spontaneous speech = 0.5) and the continuous variable of speech rate has been centered to have a mean of 0. This so that the main effects can all be interpreted as "effect of X when all other variables are held at their mean values".[4]

Separate models are built for each dialect and are given by (3.2):

$$logit[p(y)] = log\left[\frac{p(y)}{1-p(y)}\right] = \alpha + \beta_1 style + \beta_2 ssp + \beta_3 vowel + \beta_4 rate + \beta_5 style{\times}rate + \varepsilon \qquad (3.2)$$

Simplification in the data from Québec

The table of coefficients for the model predicting the probability of word-final consonant cluster simplification by style, motivation for simplification, following context and rate in the dialect from Québec is given below in Table 3.39. This table indicates that only the coefficient for simplification (`SSP`) is not

[4]Thanks to one of my anonymous reviewers for pointing out the importance of this. When an interaction between X and Y is in a model, and X is a two-level categorical factor which has been dummy coded and Y is a continuous variable which has not been centered, the interpretation of the main effect of Y is "slope of Y when X is held at level 1" and the interpretation of the main effect of X is "difference in response between level 1 and 2 of X when Y is held constant at 0". Without centering the term for speech rate, this means interpreting the main effect of X when speaking rate is 0, which is non-sensical.

significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.52154	0.03565	14.632	$< 2e-16$	***
Spontaneous	1.07629	0.05688	18.923	$< 2e-16$	***
SSP	0.04859	0.06304	0.771	0.4408	
Vowel	−1.12372	0.05092	−22.070	$< 2e-16$	***
Rate	0.06482	0.01185	5.468	$4.54e-08$	***
Spontaneous:Rate	0.05707	0.02374	2.403	0.0162	*

Table 3.39: Coefficients for combined model predicting simplification in Québec

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.40, confirms that the term SSP is not considered to be predictive and could be discarded. All other terms should be retained in the model.

Simplifying the model by removing the term for SSP and comparing the simplified model with the initial model returns $\chi^2(1) = 0.59182, p = 0.4417$. This means that, in this case, model simplification is justified, since removing the term for SSP from the model did not cause a significant reduction in the explanatory power of the model. Removal of any of the other terms in the model resulted in significant decreases in the explanatory power of the model and they must, therefore, be retained. The table of coefficients for the final model is presented as Table 3.41.

The coefficients displayed in Table 3.41 suggest that not all the intuitions regarding the effects of the various variables explored in Section 3.3 were supported. Even though Table 3.9 on page 82 suggested that there were more observations of word-final consonant cluster simplification when the motivation for simplification is a violation of the SSP and fewer observations when the motivation for simplification is perceptual salience (stop-final) than would be expected, after taking account of the variance explained by the effects of style, following context and speech rate, there does not appear to be enough statistical evidence to conclude that there are differences in rates of simplification between SSP violating and stop-final consonant clusters in the data from Québec. The large negative effect of following context, as given by the coefficient for `Vowel`, indicates that word-final consonant cluster simplification is significantly less likely when the consonant cluster is prevocalic, than when it is preconsonantal. This confirms

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			9729	11330	
Style	1	334.52	9728	10996	$< 2e-16$
SSP	1	1.40	9727	10994	0.2370
Vowel	1	512.27	9726	10482	$< 2e-16$
Rate	1	70.53	9725	10411	$< 2e-16$
Style:Rate	1	5.76	9724	10406	0.0164

Table 3.40: Analysis of deviance table for Québec

```
Coefficients:
                   Estimate  Std. Error  z value   Pr(>|z|)
(Intercept)         0.53703     0.02947   18.222   < 2e-16   ***
Spontaneous         1.07723     0.05686   18.944   < 2e-16   ***
Vowel              -1.12471     0.05090  -22.096   < 2e-16   ***
Rate                0.06507     0.01185    5.492   3.98e-08  ***
Spontaneous:Rate    0.05626     0.02372    2.372     0.0177  *
```

Table 3.41: Coefficients for minimally adequate model predicting simplification in Québec

the intuition that following context, especially a preconsonantal context, is an important predictor for simplified pronunciations of word-final consonant clusters.

Additionally, the significant main effect of speech style suggests that simplification of a word-final consonant cluster is more likely in spontaneous than in a read style of speech when speech rate is held at the mean. Furthermore, there was a signficant interaction between speech style and speech rate. The effect of speech rate is larger in spontaneous than in read speech. These relationships are displayed in Figure 3.6

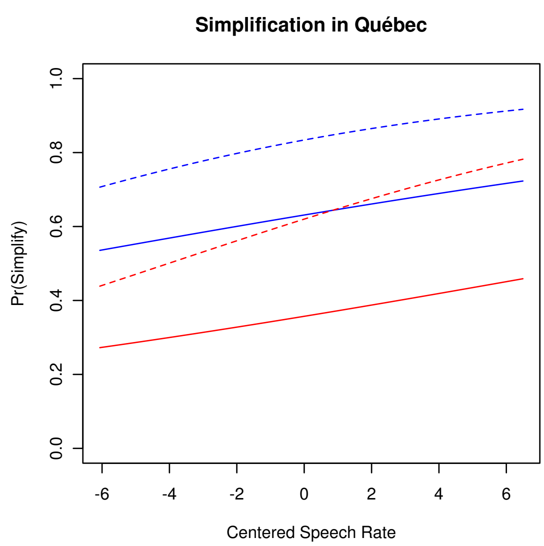

Figure 3.6: Regression predicting simplification in Québec

Simplification in the data from France

The table of coefficients for the model predicting the probability of word-final consonant cluster simplification by style, motivation for simplification, following context and rate in the dialect from France is given below in Table 3.42. This table indicates that all the coefficients are significant.

The analysis of deviance table produced by an ANOVA summary for this model is presented below

Coefficients:					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−1.19695	0.04418	−27.091	$< 2e-16$	***
Spontaneous	1.37755	0.05584	24.670	$< 2e-16$	***
SSP	0.72419	0.07839	9.238	$< 2e-16$	***
Vowel	−1.17668	0.06402	−18.380	$< 2e-16$	***
Rate	0.06200	0.01279	4.847	$1.25e-06$	***
Spontaneous:Rate	0.06138	0.02556	2.402	0.0163	*

Table 3.42: Coefficients for combined model predicting simplification in France

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			7530	10071.7	
Style	1	667.72	7529	9404.0	$< 2.2e-16$
SSP	1	112.45	7528	9291.5	$< 2.2e-16$
Vowel	1	374.69	7527	8916.8	$< 2.2e-16$
Rate	1	54.31	7526	8862.5	$1.71e-13$
Style:Rate	1	5.76	7525	8856.8	0.01639

Table 3.43: Analysis of deviance table for France

as Table 3.43.

Simplifying the model by removing any of the terms in the model resulted in significant decreases in the explanatory power of the model and they must, therefore, be retained. The coefficients displayed in Table 3.42 suggest that all the intuitions regarding the effects of the various variables explored in Section 3.3 were supported. Table 3.10 on page 83 suggested that there were more observations of word-final consonant cluster simplification when the motivation for simplification is a violation of the SSP and fewer observations when the motivation for simplification is perceptual salience (stop-final) than would be expected. After taking account of the variance explained by the effects of style, following context and speech rate, the large, positive effect of the motivation for simplification, as given by the coefficient for `SSP`, indicates that there are significant differences in the rates of simplification between SSP violating and stop-final consonant clusters in the data from France: The odds of observing word-final consonant cluster simplification with SSP violating consonant clusters are higher than with stop-final consonant clusters. The large negative effect of following context, as given by the coefficient for `Vowel`, indicates that, similar to the effect observed in the data from Québec, word-final consonant cluster simplification is significantly less likely when the consonant cluster is prevocalic, than when it is preconsonantal. This further confirms the intuition that following context, especially a preconsonantal context, is an important predictor for simplified pronunciations of word-final consonant clusters. Additionally, the significant main effect of speech style suggests that simplification of a word-final consonant cluster is more likely in spontaneous than in a read style of speech when speech rate is held at the mean. Furthermore, there was a signficant interaction between speech style and speech rate. The effect of speech rate is larger in spontaneous than in read speech. These relationships are displayed in Figure 3.7

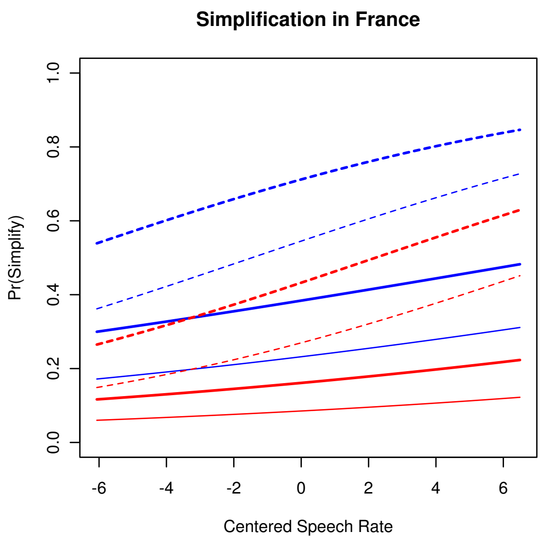

Figure 3.7: Regression predicting simplification in France

3.5.2 Schwa insertion according to style, context, simplification and speech rate

The model predicting the response of schwa insertion considers all the consonant clusters in the data. It includes terms for those variables that were suggested by the analyses in Section 3.4 to have explanatory value. Similar to the model predicting word-final consonant cluster simplifiation, the term Vowel contains just two levels: 0 for the preconsonantal context and 1 for the prevocalic context and excludes all observations of word-final consonant clusters that occurred in the prepausal context. The term that describes whether or not a consonant cluster also participated in word-final consonant cluster simplification, Simplify, likewise contains two levels: 0 for consonant clusters that were not observed to participate in simplification and 1 for consonant clusters that did participate in simplification. In these data, consonant clusters that did present observations of simplification include all SSP violating and stop-final consonant clusters. Consonant clusters that did not present observations of simplifications include all /ʁ/-initial consonant clusters as well as the stop+fricative consonant cluster /ks/. The categorical predictor for speech style includes two levels: Reading and Spontaneous. The covariate term for speech rate is a continuous variable. Additionally, since it was observed that the effect of speech rate may likely be constrained to have an effect on just the spontaneous speech style, an interaction term between style and speech rate is included. The explanatory variable that describes the effect of the motivation for simplification, SSP, will not be considered in this model. SSP applies only to a subset of the consonant clusters considered in this model and therefore its effect, if any, would be more difficult to interpret. The effect that should be more interesting to determine, with respect to schwa insertion following a word-final consonant cluster, is the difference between consonant clusters that may also simplify and those that remain stable. Separate models are built for each dialect and are given by (3.3):

$$logit[p(y)] = log\left[\frac{p(y)}{1-p(y)}\right] = \alpha + \beta_1 style + \beta_2 simplify + \beta_3 vowel + \beta_4 rate + \beta_5 style{\times}rate + \varepsilon \quad (3.3)$$

Schwa insertion in the data from Québec

The table of coefficients for the model predicting the probability of schwa insertion following a word-final consonant cluster by style, simplification, following context and rate in the dialect from Québec is given below in Table 3.44.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-3.17447	0.10344	-30.689	$< 2e-16$	***
Spontaneous	-1.20850	0.07030	-17.190	$< 2e-16$	***
Simplify	1.126227	0.104410	10.787	$< 2e-16$	***
Vowel	-0.306637	0.078647	-3.899	$9.66e-05$	***
Rate	-0.06775	0.01439	-4.706	$2.52e-06$	***
Spontaneous:Rate	-0.091763	0.028805	-3.186	0.00144	**

Table 3.44: Coefficients for combined model predicting schwa insertion in Québec

The analysis of deviance table produced by an ANOVA summary for this model is presented below as Table 3.45.

Simplifying the model by removing any of the terms in the model resulted in significant decreases in the explanatory power of the model. Therefore, the model described by Table 3.44 should be considered as a minimally adequate model. The coefficients displayed in Table 3.44 suggest that all the intuitions regarding the effects of the various variables explored in Section 3.4 were supported. Table 3.26 on page

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			13422	7543.1	
Style	1	372.01	13421	7171.1	$< 2.2e-16$
Simplify	1	148.78	13420	7022.3	$< 2.2e-16$
Vowel	1	12.57	13419	7009.7	0.0003919
Rate	1	41.60	13418	6968.1	$1.122e-10$
Style:Rate	1	10.10	13417	6958.0	0.0014787

Table 3.45: Analysis of deviance table for Québec

96 suggested that there were more observations of schwa insertion following SSP violating and stop-final clusters and fewer observations of schwa insertion following the other clusters than would be expected if the differences between these two categories were due simply to chance. After taking account of the variance explained by the effects of style, following context and speech rate, the large, positive effect of this variable, as given by the coefficient for `Simplify`, indicates that there are significant differences in the rates of schwa insertion between consonant clusters that may also participate in simplification and consonant clusters that remain stable in the data from Québec: The odds of observing schwa insertion following a word-final consonant cluster with SSP violating or stop-final consonant clusters are higher than with other consonant clusters. The negative effect of following context, as given by the coefficient for `Vowel`, indicates that schwa insertion following a word-final consonant cluster is significantly less likely when the consonant cluster is prevocalic, than when it is preconsonantal. While the effect of following context on schwa insertion is not as large as the effect of following context on simplification - consider the coefficient for `Vowel` in Table 3.44: -0.306637 as compared with the same coefficient in Table 3.41: -1.12471 - the significant difference between a preconsonantal and a prevocalic context provides further confirmation of the intuition that following context, especially a preconsonantal context, is an important predictor for the variable pronunciations of word-final consonant clusters.

As was expected given the initial results provided in Section 3.4.5 on page 102, speech style is an important predictor for pronunciations of schwa following a word-final consonant cluster. The odds of observing a response of schwa are significantly lower in spontaneous speech than in read speech, when speech rate is held at the mean. There also exists a significant interaction between speech style and speech rate in that the effect of rate is significantly larger in spontaneous speech than in read speech. These relationships are displayed in Figure 3.8

Schwa insertion in the data from France

The table of coefficients for the model predicting the probability of schwa insertion following a word-final consonant cluster by style, simplification, following context and rate in the dialect from France is given below in Table 3.46.

The analysis of deviance table produced by an ANOVA summary for this model is presented below as Table 3.47.

Simplifying the model by removing any of the terms in the model resulted in significant decreases in the explanatory power of the model. The coefficients displayed in Table 3.46 suggest that all the intuitions regarding the effects of the various variables explored in Section 3.4 were supported. Table 3.27 on page 97 suggested that there were more observations of schwa insertion following SSP violating

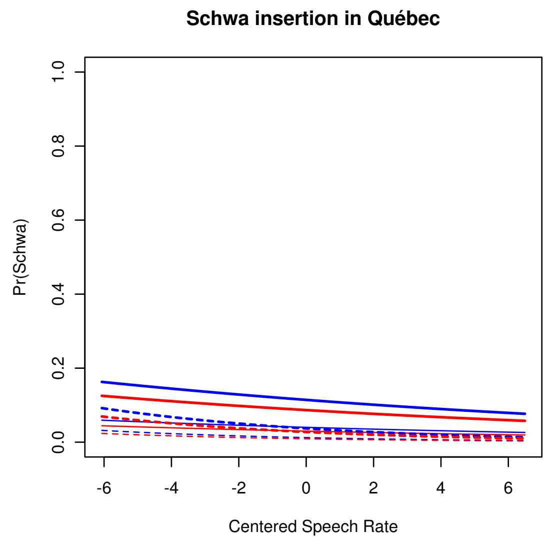

Figure 3.8: Regression predicting schwa insertion in Québec

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.63041	0.05833	-27.952	$< 2e-16$	***
Spontaneous	-0.94377	0.04797	-19.673	$< 2e-16$	***
Simplify	0.577801	0.057746	10.006	$< 2e-16$	***
Vowel	-1.462112	0.067412	-21.689	$< 2e-16$	***
Rate	-0.08614	0.01065	-8.087	$6.11e-16$	***
Spontaneous:Rate	-0.071115	0.021277	-3.342	0.000831	***

Table 3.46: Coefficients for model predicting schwa insertion in France

and stop-final clusters and fewer observations of schwa insertion following the other clusters than would be expected if the frequency of schwa insertion was independent of whether or not a consonant cluster also participates in simplification. After taking account of the variance explained by the effects of style, following context and speech rate, the positive effect of this variable, as given by the coefficient for `Simplify`, indicates that there are significant differences in the rates of schwa insertion between consonant clusters that may also participate in simplification and consonant clusters that remain stable in the data from France: The odds of observing schwa insertion following a word-final consonant cluster

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			10185	12527	
Style	1	585.93	10184	11941	$< 2.2e-16$
Simplify	1	89.32	10183	11852	$< 2.2e-16$
Vowel	1	562.50	10182	11289	$< 2.2e-16$
Rate	1	102.63	10181	11186	$< 2.2e-16$
Style:Rate	1	11.16	10180	11175	0.0008363

Table 3.47: Analysis of deviance table for France

with SSP violating or stop-final consonant clusters are higher than with other consonant clusters. The large, negative effect of following context, as given by the coefficient for `Vowel`, indicates that schwa insertion following a word-final consonant cluster is significantly less likely when the consonant cluster is prevocalic, than when it is preconsonantal. As was expected given the initial results provided in Section 3.4.5 on page 103, speech style is an important predictor for pronunciations of schwa following a word-final consonant cluster. The odds of observing a response of schwa are significantly lower in spontaneous speech than in read speech, when speech rate is held at the mean. There also exists a significant interaction between speech style and speech rate in that the effect of rate is significantly larger in spontaneous speech than in read speech. These relationships are displayed in Figure 3.9

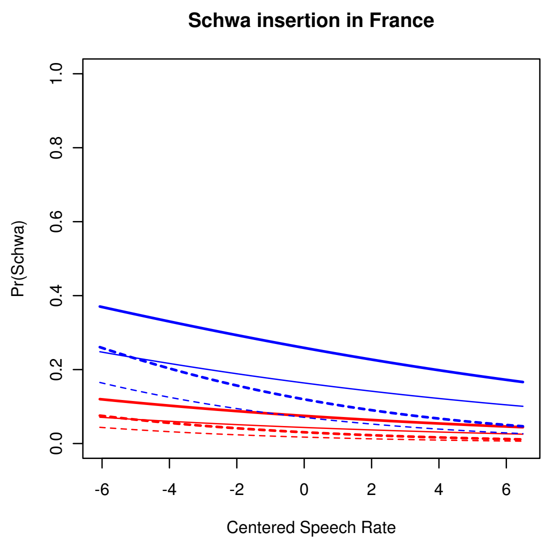

Figure 3.9: Regression predicting schwa insertion in France

3.5.3 Discussion

A meaningful interpretation of the results presented in this section can be best achieved by comparing and contrasting the different effects each of the variables of interest had on the variable pronunciations of word-final consonant clusters in each dialect. While the four models investigated (2 predicting the response of simplification; 2 predicting the response of schwa insertion) may appear, on the surface, to be similar when comparing the models for the data from Québec with the models for the data from France, a closer look at both the strength and direction of the relationships expressed by the coefficients suggests they are, in fact, different in some interesting ways.

In the first place, the effect of speech rate appears to be very similar both between the two dialects

(Québec and France) as well as between the two reponses (simplification and schwa insertion). In the models predicting word-final consonant cluster simplification, the positive coefficients for `Rate` indicate that, in both dialects, as speech rates increased, the odds of observing the response of simplification also increased. The models for both Québec and France suggest this effect is greater in the spontaneous style of speech. In the models predicting schwa insertion following a word-final consonant cluster, the negative coefficients for `Rate` indicate that, in both dialects, as speech rates increased, the odds of observing the response of schwa insertion decreased. The models for both Québec and France suggest this effect is significantly larger in the spontaneous style of speech. In fact, comparing the four coefficients that describe the effect of speech rate for each of the two dialects (Table 3.48) it appears that the effect of speech rate is nearly the same. This should come as no surprise since the effects of speech rate can be assumed to be, for the most part, influenced by the bio-mechanical demands of articulating speech at faster rates. As there is no reason to believe that speakers from Québec are physiologically different from speakers from France, the effects of increasingly rapid articulation should be the same in both populations. The different valences of the coefficients - positive for the models predicting simplification; negative for the models predicting schwa insertion - yields an intuition that as speech rates increase, the physical demands of rapid articulation might predict that simplification of the consonant cluster becomes the more frequent pronunciation variant for these word-final consonant clusters.

Dialect	Simplification		Schwa	
	Reading	Spontaneous	Reading	Spontaneous
QC	0.06507	0.12133	−0.06775	−0.159513
FR	0.06200	0.12338	−0.08614	−0.157255

Table 3.48: Coefficients for rate from models

In the second place, the effect of speech style was also similar in each of these two dialects. For both sets of models - those predicting simplification and those predicting schwa insertion - in spontaneous speech, faster speech rates are associated with higher odds of simplification and lower odds of schwa insertion following a word-final consonant cluster, while in read speech, the effect of speech rate was significantly smaller.

Turning to the effects of following context on either simplification or schwa insertion following a word-final consonant cluster, some interesting differences emerge between the two dialects. The models indicated that, in agreement with much previous research on both simplification and schwa insertion following a word-final consonant cluster, the preconsonantal context is a more likely condition under which simplification or schwa insertion will occur. In both dialects, the responses of simplification and schwa insertion were both more likely to occur when the consonant cluster was in a preconsonantal context than when in a prevocalic context. However, the strength of this relationship, as expressed by the coefficient `Vowel` suggests that there are dialectal differences in the response to this variable. Table 3.49 shows the four coefficients contained in the two models.

Dialect	Simplification	Schwa
QC	−1.12471	−0.306637
FR	−1.17668	−1.462112

Table 3.49: Coefficients for vowel from models

It is easy to see that, in the model predicting simplification, the effect of following context is very similar between the two dialects. In the data from both Québec and France, the odds of observing the response of simplification following a word-final consonant cluster were significantly lower when the consonant cluster was in a prevocalic context than when in a preconsonantal context and the change in these odds is approximately the same. In fact, calculating the odds ratio of observing the response of simplification as opposed to non-simplification indicates that in the dialect from Québec, it is 3.08 times more likely and in the dialect from France it is 3.24 times more likely that simplification will occur in a preconsonantal context than in a prevocalic context. Compare this to the large difference between the two dialects in the values for the coefficients in the model predicting schwa insertion. This suggests that the response to following context, with respect to schwa insertion, is not the same in both dialects. In fact, calculating the odds ratio of observing the response of schwa insertion as opposed to no schwa insertion indicates that in the dialect from Québec, it is just 1.36 times more likely that schwa insertion will occur in a preconsonantal context than in a prevocalic context, while in the dialect from France it is 4.32 times more likely that schwa insertion will occur in a preconsonantal context than in a prevocalic context. This large difference in odds ratios suggests that, even though the effect of following context is statistically significant in both dialects, the phonological role it plays as a "trigger" for schwa insertion as a repair for word-final consonant clusters may be much more pronounced in the dialect from France than in the dialect from Québec.

With respect to the effect of the motivation for simplification, that is, whether the consonant cluster violates the SSP or instead contains a final stop consonant, the results of model simplification for the dialect from Québec suggest that this distinction is not relevant when predicting a response of simplification while it is relevant for the dialect from France. Even though the initial results described in Section 3.3.2 on page 82 suggested that, in both dialects, word-final consonant cluster simplification was not independent of the motivation for simplification, after taking account of the variance explained by the effects of style, following context and speech rate, there does not appear to be enough statistical evidence to conclude that there are differences in rates of simplification between SSP violating and stop-final consonant clusters in the data from Québec. In contrast, the coefficient for `SSP` expressed in the model predicting simplification following a word-final consonant cluster in the dialect from France suggests that it is more than twice as likely (an odds ratio of 2.06) that the response of simplification as opposed to non-simplification of the consonant cluster will be observed with SSP violating consonant clusters as compared with the stop-final consonant clusters.

Tables 3.4 and 3.20 (on pages 79 and 92) described two interesting observations about the distributions of simplification and schwa insertion following a word-final consonant cluster in both dialects. On the one hand, the list of 29 individual consonant clusters examined in these data appeared to divide into two groups with respect to word-final consonant cluster simplification. In the group identified as "Simplify" were all SSP violating and stop-final consonant clusters (excluding all /ʁ/-initial clusters). These consonant clusters all had a significant proportion of simplified consonant clusters in their observed pronunciation variants. In the group identified as "Stable" were all /ʁ/-initial consonant clusters as well as the stop+fricative consonant cluster /ks/. These consonant clusters all had no significant proportion of simplified consonant clusters in their observed pronunciation variants. This division into "Simplify" and "Stable" consonant clusters, with respect to consonant cluster simplification, follows from the predictions described in Côté (2004a) and was the same in both dialects. On the other hand, the list of 29 individual consonant clusters examined in these data did not appear to divide into two groups with respect to schwa insertion following a word-final consonant cluster. In the dialect from France, schwa insertion regularly occurred following almost every word-final consonant cluster in these data. Only the stop+fricative cluster /ks/ was not observed to participate in schwa insertion. This contrasts with the dialect from Québec where schwa insertion following a word-final consonant cluster was observed to regularly occur in all SSP violating clusters; was observed to less frequently occur in stop-final clusters; and was only observed following /ʁm/ and /ʁt/ for those consonant clusters identified as being stable with respect to simplification. The results of the main effect of this division between "Simplify" and "Stable" consonant clusters, discussed on page 95, suggested that there were significantly more observations of schwa insertion following SSP violating and stop-final clusters (Simplify) and significantly fewer observations of schwa insertion following the other clusters (Stable) than would be expected if the differences between these two categories were due simply to chance and that the effect, while present in both dialects, was only slightly stronger in the dialect from Québec. In fact, in the models predicting schwa insertion following a word-final consonant cluster, the coefficient that tests for this effect, `Simplify`, suggests that the strength of the reponse to this variable is very different in each dialect. In the dialect from France, it is only 1.78 times more likely that schwa insertion will occur following "Simplify" type consonant clusters than when following "Stable" type consonant clusters. In the dialect from Québec, it is more than 3 times more likely (an odds ratio of 3.08).

This result, that the odds of observing schwa insertion following SSP violating or stop-final consonant clusters when compared with "Stable" consonant clusters in the dialect from Québec are nearly double the odds of schwa insertion in these clusters in the dialect from France, broadly conforms to the interpretations of Milne and Côté (2009) that, in the dialect from Québec, schwa insertion following a word-final consonant cluster is mainly observed with consonant clusters that may also simplify. When combined with the fact that the strength of the relationship between following context and schwa insertion is significantly less strong in the dialect from Québec than in the dialect from France, it gives rise to the hypothesis that schwa insertion may be motivated as an avoidance of simplification in the dialect from Québec, while it is motivated by a more general principle that seeks to avoid sequences of any consonants in the dialect from France.

3.5.4 Schwa as an alternative to simplification

Since schwa insertion following "Simplify" consonant clusters is much more likely than when following "Stable" consonant clusters in the dialect from Québec and the effect of following context on schwa insertion following a word-final consonant cluster is much weaker, could it be the case that simplification itself, specifically the frequency with which a word-final consonant cluster is simplified, helps to predict schwa insertion in the dialect from Québec. That is, does knowing something about the probability of simplification for a given consonant cluster help to explain the probability of schwa insertion? Since the probability of simplification, at least in the spontaneous style of speech, is associated with the speech rate with which the consonant cluster is articulated, in order to answer the question a data set will need to be constructed that does not merely take account of differences in rates of speech, as has been done up to now, but actually controls for speech rate. In order to control for speech rate, a sample of the current data can be taken in which there exists no difference in speech rate between the two styles and in which no effect of speech rate on either the response of simplification or schwa insertion is present.

Such a sample was identified as containing speech rates that varied from 12.42767 phones per second to 13.92767 phones per second for the data from Québec and from 12.92767 to 13.42767 phones per second for the data from France. Both samples are centred on the mean rate of speech for the entire population (13.17767 phones per second) and contain a portion of the distribution of the entire data set that includes ± 0.75 standard deviations for the data from Québec and ± 0.25 standard deviations for the data from France. Analyses of variance of these two samples indicate no significant difference in the rates of speech between the two styles: $F(1, 3337) = 3.427, p = 0.0642$ for the data from Québec and $F(1, 1014) = 3.281, p = 0.0704$ for the data from France. Logistic regressions of these two samples indicate no significant effect of rate on the response of either simplification ($p = 0.05183$ for the dialect from Québec and $p = 0.844$ for the dialect from France) or schwa insertion ($p = 0.690$ for the dialect from Québec and $p = 0.996$ for the dialect from France). In terms of speech rate, this sample can be considered as containing observations of variable pronunciations of word-final consonant clusters drawn from two different styles of speech but a homogeneous rate of speech.

This example is interested in how variables, such as speech style, following context and rate of simplification, effect the variable pronunciations of word-final consonant clusters. The data set has a binary response variable called **schwa**. There are three predictor variables: **simplification** is a continuous variable expressing the mean rate of simplification; **style** has values for reading and spontaneous; **vowel** takes on the values of ‘V’ if the consonant cluster is prevocalic and ‘C’ if the consonant cluster is preconsonantal. Table 3.50 shows the amount of data available in this sample and, importantly, shows there are no 0 cells.

dialect	**style**	**vowel**	**no_schwa**	**schwa**
QC	Reading	C	321	117
		V	115	14
	Spontaneous	C	1659	96
		V	528	39
FR	Reading	C	152	154
		V	62	7
	Spontaneous	C	271	99
		V	103	14

Table 3.50: Summary of sub-sample

Separate logistic regression models predicting the outcome of schwa according to speech style, following context, and rate of simplification were built for each of the two dialects. The hypothesis to be tested is whether the coefficient for the slope (**simplification**) is significantly different from 0 and specifically whether or not it is positive. Table 3.51 displays the results for the dialect from Québec while Table 3.52 displays the results for the dialect from France.

`Coefficients:`

	`Estimate`	`Std. Error`	`z value`	`Pr(>|z|)`	
`(Intercept)`	−1.4156	0.1531	−9.244	$< 2e-16$	***
`styleSpontaneous`	−1.6387	0.1366	−11.994	$< 2e-16$	***
`vowel`	−0.1947	0.1662	−1.171	0.2414	
`simplification`	0.5374	0.2188	2.456	0.0141	*

Table 3.51: Coefficients for model predicting schwa insertion in Québec

The results suggest that the effects of **simplification** (the frequency with which a consonant cluster

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.08323	0.13188	-0.631	0.528	
styleSpontaneous	-0.91503	0.16235	-5.636	$1.74e-08$	***
vowel	-1.51290	0.24913	-6.073	$1.26e-09$	***
simplification	0.15549	0.33261	0.467	0.640	

Table 3.52: Coefficients for model predicting schwa insertion in France

simplifies) and **vowel** (a preconsonantal or prevocalic context) are not the same in both dialects. For the dialect from Québec, no significant effect of **vowel** emerges from the model. As was suggested earlier, the effect of following context on the presence of schwa insertion following a word-final consonant cluster in the dialect from Québec is not predictive. In support of the hypothesis, the coefficient for the slope in this model (**simplification**) is both positive and significantly different from 0. For every one unit increase in the frequency of simplification, the odds of schwa insertion (versus no schwa insertion) increase by a factor of 1.72. In the dialect from Québec, knowing something about a consonant cluster's frequency of simplification helps to know whether or not schwa insertion may also occur following that consonant cluster.

For the dialect from France, there was a significant effect of **vowel**. As was previously determined, schwa insertion following a word-final consonant cluster in a prevocalic environment was significantly less likely than when the consonant cluster was in a preconsonantal environment. The odds of schwa insertion in a prevocalic context decrease by a factor of 0.22 as compared to the odds of schwa insertion in a preconsonantal context. In the dialect from France, knowing something about the following context helps to know whether or not schwa insertion will occur. Failing to support the hypothesis, the coefficient for the slope in this model (**simplification**) is not significantly different from 0. For every one unit increase in the frequency of simplification, the odds of schwa insertion (versus no schwa insertion) increase by a factor of just 1.16. While it is slightly positive, there is not enough statistical evidence to conclude that knowing something about a consonant cluster's frequency of simplification helps to know whether or not schwa insertion may also occur following that consonant cluster for the dialect from France. These effects are presented in Figure 3.10.

One may wonder whether the significant effect of **style** that emerged in both dialects in these models supports or contradicts the earlier conclusions that the differences between a read and spontaneous style of speech are mainly attributed to the different sensitivites to rate for each style. In the models for both dialects, the coefficient for **style** is significantly negative: The odds of schwa insertion following a word-final consonant cluster in the spontaneous style of speech decrease by a factor of 0.19 in the dialect from Québec and by a factor of 0.40 in the dialect from France when compared to the odds

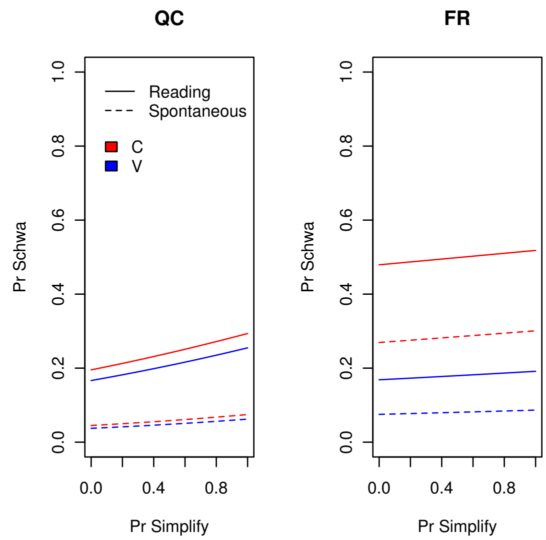

Figure 3.10: Regressions predicting schwa insertion as a function of simplification

in a reading style of speech. It is difficult to determine whether the correct interpretation should be that, after controlling for speech rate, there does exist a significant effect of speech style or whether the correct interpretation should instead remain that apparent differences between the two speech styles is best attributed to differences in sensitivity to speech rate. Figure 3.11 shows the logistic curves that emerged from the models predicting a response of schwa obtained from the analysis in Section 3.4.5. The two solid vertical lines added to the figure indicate the range of speech rates contained in the current sample. It is easy to see that, even though the sample displayed no difference in speech rates between the two styles, within the range of speech rates covered by this sample, the effect of speech rate on schwa insertion for the spontaneous style has already produced a significant gap between the individual regression lines for each style. If a sample had been taken from a slower rate of speech, indicated by the two dashed vertical lines in the figure, it is likely that no significant effect of style would be present in the models. However, since the distribution of speech rate in the complete data set indicated that rates of speech slower than approximately 7 phones per second would be extremely unlikely to occur in a normally distributed population ($p < 0.025$), obtaining such a sample would be next to impossible. This demonstrates the difficulty, even in a relatively large data set, of disentangling the effects of speech rate and speech style.

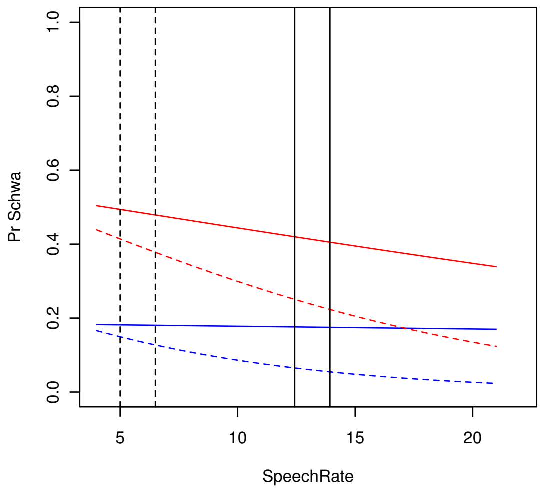

Figure 3.11: Actual and hypothetical samples controlling for rate of speech

3.6 Corpus effects

Thus far, the effect of speech style, when defined as the differences between read and non-read speech, has been a challenge to interpret. The effects of other variables, such as following context and motivation for simplification (expressed either as the difference between SSP violating and stop-final consonant clusters or as the difference between Simplify and Stable consonant clusters) were relatively straightforwardly interpreted. These effects, on the whole, conformed with previous expectations and offered some valuable new insights into some of the possible dialectal differences in the variable pronunciations of word-final consonant clusters in French. The tangle of speech style and speech rate, on the other hand, made it consistently difficult to determine which of the two was the more appropriate explanatory variable. When examining the differences between speech styles, whether as an individual variable in Sections 3.3.4 and 3.4.5, as part of a mixed model in Sections 3.5.1 and 3.5.2 or as part of a smaller sample that tried to control for the effects of speech rate in Section 3.5.4, the best interpretation of the differences between speech styles in these data has remained that it appears to describe the different sensitivities to speech rate: In spontaneous speech, faster rates of speech are positively associated with simplification and negatively associated with schwa insertion, while in read speech, increasing rates of speech had little or no effect on the response of either simplification or schwa insertion following a word-final consonant cluster. This was due, at least in part, to the assumption that the distinction between read and non-read speech would serve as a reliable proxy for the identification of two distinct styles of speech. The fact that, in these data, the distinction between read and non-read speech was accompanied by a statistically significant difference in rates of speech introduced the confound. However the nature of these data, collected from two different sources that each contained examples of read and non-read speech from many speakers, provided an additional opportunity to investigate the possible effect of speech style by exploring differences between the corpora themselves.

Intuitively, differences between the two corpora should be expected. It is likely that the speech patterns observed in a casual conversation will be different from those in a political debate, even if the utterances in the political forum were delivered without the aid of notes. It is also likely that professional speakers, such as members of a legislative assembly, will perform differently from non-professional speakers, such as the participants in the PFC surveys. Several earlier attempts at exploiting stylistic differences between corpora with respect to the variability of the French schwa vowel (c.f. Nemoto et al. (2010); Adda-Decker et al. (2002, 1999)) used corpora of exclusively read text to represent a more formal speech style. The inclusion of the AssNat corpus in this project followed the example of the Phonologie du Français Contemporain (Durand et al., 2002, 2005, 2009), which attempts to include examples of both read and non-read speech from the same speaker in order to assess differences in speech patterns according to stylistic differences. This was seen as an advantage over selecting a corpus of, for example, read newspaper speech (Lamel et al., 1991) or a travel dialog system (Lamel et al., 1995) since the inclusion of the political debates provided a stylistically comparable corpus to the PFC.

This section continues to examine the research question of how variables, such as speech style, following context, motivation for simplification and speech rate, affect the variable pronunciations of word-final consonant clusters in two dialects of French. Different from the previous sections though, speech styles are distinguished by corpus: The data provided by the PFC corpus, which includes non-professional speakers in a less formal setting, and the data provided by the AssNat corpus, which includes professional speakers in a more formal setting. The distinction between read and non-read speech will be maintained and separate analyses will be performed for each dialect. In Section 3.6.1 possible differences between the two corpora for the responses of simplification and schwa insertion following a word-final consonant cluster in read text are explored. Due to the nature of the artificial TexteLu exercise of the PFC corpus, only a few consonant clusters can be included and all of them are in a preconsonantal context. Following this, differences between the two corpora for spontaneous, non-read speech are explored in terms of word-final consonant cluster simplification and schwa insertion following a word-final consonant cluster in Section 3.6.2. The larger amount of data available from the spontaneous speech portion of the two corpora allowed for analyses similar to those previously performed in Sections 3.5.1 for simplification and 3.5.2 for schwa insertion. This section concludes with a summary of the findings and how they help inform the earlier conclusions.

Motivation	ClusterType	Cluster	Text
SSP	FSA	stʁ	"... *premier <u>ministre</u> ne cesse*..."
	FSA	stʁ	"... *premier <u>ministre</u> pour vérifier*..."
	SA	tʁ	"... *seuls <u>titres</u> de gloire*..."
	SA	tʁ	"... *d' un <u>autre</u> côté*..."
	SA	tʁ	"... *de <u>notre</u> liberté*..."
	SA	tʁ	"... *au <u>centre</u> d' une bataille*..."
	SA	bʁ	"... *jeune <u>membre</u> de l' opposition*..."
	SA	kl	"... *quelques <u>articles</u> parus*..."
Stop-final	FS	sk	"... *d' identité <u>risquent</u> de provoquer*..."
	FS	st	"... *l' express, <u>ouest</u> liberté*..."
	FS	st	"... *des <u>activistes</u> des communes*..."
	FS	st	"... *les bords <u>manifestent</u> leur colère*..."

Table 3.53: Word-final consonant clusters in TexteLu exercise of PFC

	FR		QC		Sum
Motivation	AssNat	PFC	AssNat	PFC	
Stop-final	74	391	34	247	746
SSP	770	870	424	585	2649
Sum	844	1261	458	832	3395

Table 3.54: Frequency of word-final consonant clusters in read speech by corpus and dialect

3.6.1 Corpus effects in read speech

The investigation into the possible differences between the read text portions of the two corpora is limited by the word-final consonant clusters represented in the TexteLu exercise of the PFC investigations. In the TexteLu exercise, there were 21 occurrences of a word containing a word-final consonant cluster. Not all of the word-final consonant clusters were represented in both a preconsonantal and prevocalic context, therefore the effect of following context was not able to be evaluated for the read text portions of the corpora. Additionally, there was only a single non-SSP violating, non-stop-final consonant cluster (the type idenitified as being Stable) in a preconsonantal context in the TexteLu exercise. Therefore neither could the differences between Simplify and Stable type consonant clusters on the response of schwa insertion be evaluated. Table 3.53 displays the 6 individual word-final consonant clusters selected from the TexteLu exercise of the PFC corpus along with the snippet of text in which they occurred. In the table, it can be seen that all of the word-final consonant clusters were either SSP violating or stop-final consonant clusters that occurred in a preconsonantal context. These 6 word-final consonant clusters were identified in the read speech portion of the AssNat corpus in order to compare the differences between the two corpora in each dialect. Table 3.54 displays the amount of data made available for the investigation comparing the read speech portions of the two corpora.

Simplification according to corpus: Read speech

Table 3.55 shows the observed and expected frequencies of simplified word-final consonant clusters according to corpus in the read text portion of the data from Québec. In these data from Québec, the obtained value of $\chi^2(1) = 9.59$ is significantly larger than the critical value of $\chi^2(1) = 3.84, p = 0.002$ and the null hypothesis must be rejected. There are significantly more observations of word-final consonant cluster simplification in the PFC corpus and significantly fewer observations in the AssNat corpus than would be expected if the differences between these two corpora were due to chance.

	Observed			Expected		
	Simplification			Simplification		
Corpus	Yes	No	Total	Yes	No	Total
AssNat	252	206	458	278.00	180.00	458
PFC	531	301	832	505.00	327.00	832
Total	783	507	1290	783	507	1290

Table 3.55: Observed and expected frequencies for simplification as a function of corpus for Québec

Table 3.56 shows the observed and expected frequencies of simplified word-final consonant clusters according to corpus in the read text portion of the data from France. The obtained value of $\chi^2(1) = 0.36$ is not significantly larger than than the critical value of $\chi^2(1) = 3.84, p = 0.547$ and the null hypothesis must be accepted. There is not enough statistical evidence to conclude that the frequency of observations of word-final consonant cluster simplification differs according to corpus for the read text portion of the data from France.

	Observed			Expected		
	Simplification			Simplification		
Corpus	Yes	No	Total	Yes	No	Total
AssNat	203	641	844	197.27	646.73	844
PFC	289	972	1261	294.73	966.27	1261
Total	492	1613	2105	492	1613	2105

Table 3.56: Observed and expected frequencies for simplification as a function of corpus for France

Schwa insertion according to corpus: Read speech

Table 3.57 shows the observed and expected frequencies of schwa insertion following a word-final consonant cluster according to corpus in the read text portion of the data from Québec. In these data from Québec, the obtained value of $\chi^2(1) = 23.39$ is significantly larger than than the critical value of

$\chi^2(1) = 3.84, p < 0.001$ and the null hypothesis must be rejected. There are significantly more observations of schwa insertion following a word-final consonant cluster in the AssNat corpus and significantly fewer observations in the PFC corpus than would be expected if the differences between these two corpora were due to chance.

	Observed			Expected		
	Schwa			Schwa		
Corpus	Yes	No	Total	Yes	No	Total
AssNat	144	314	458	108.64	349.36	458
PFC	162	670	832	197.36	634.64	832
Total	306	984	1290	306	984	1290

Table 3.57: Observed and expected frequencies for schwa insertion as a function of corpus for Québec

Table 3.58 shows the observed and expected frequencies of schwa insertion following a word-final consonant cluster according to corpus in the read text portion of the data from France. The obtained value of $\chi^2(1) = 23.63$ is significantly larger than than the critical value of $\chi^2(1) = 3.84, p < 0.001$ and the null hypothesis must be rejected. There are significantly more observations of schwa insertion following a word-final consonant cluster in the AssNat corpus and significantly fewer observations in the PFC corpus than would be expected if the differences between these two corpora were due to chance.

	Observed			Expected		
	Schwa			Schwa		
Corpus	Yes	No	Total	Yes	No	Total
AssNat	549	295	844	495.17	348.83	844
PFC	686	575	1261	739.83	521.17	1261
Total	1235	870	2105	1235	870	2105

Table 3.58: Observed and expected frequencies for schwa insertion as a function of corpus for France

These tables are interesting, but should be interpreted cautiously. These data contain only 6 of the 29 word-final consonant clusters that were examined previously. All of them are either SSP violating or stop-final (i.e. Simplify) consonant clusters and all of them are in a preconsonantal context. Even though a main effect of speech style appeared to be present when examining the data set as a whole, in Section 3.5 it was shown that for both of the responses of either simplification or schwa insertion, after taking account of the variance explained by the variables following context, motivation for simplification and speech rate, the effect of speech style was only relevant when describing the different sensitivities to speech rate between a read and non-read style of speech. The limited data made available from just the TexteLu exercise of the PFC corpus makes such an investigation impossible. Therefore it would be difficult to conclude from these tables alone whether differences truly exist between the read text portion of the corpora or whether these differences might also become less relevant when examined in combination with other variables. Finally, it is unclear whether these tables describe differences in *speech* style or differences in *speaker* style. In addition to speaking in an ostensibly more formal environment (the national assembly), the speakers represented by the AssNat corpus are professionals who rely on their ability to speak both clearly and effectively. The participants in the PFC investigations were recruited from a variety of backgrounds, not all of which required such a high level of speaking proficiency. Therefore, the differences may also reflect a distinction between professional and non-professional speakers.

Nevertheless, whether due to differences according to speech style or some other factor(s), there may be differences between the read text portions of these two corpora. In the data from Québec, the frequency of observations of word-final consonant cluster simplification was lower in the AssNat corpus than in the PFC corpus and observations of schwa insertion following a word-final consonant cluster were higher in the AssNat corpus than in the PFC corpus. In the data from France, the same pattern was observed with respect to schwa insertion though there did not appear to be any difference between the two corpora in terms of simplification.

3.6.2 Corpus effects in spontaneous speech

The spontaneous, non-read portions of the corpora offered a better opportunity to examine possible corpus effects. All 29 of the word-final consonant clusters were attested in both a preconsonantal and prevocalic following context. The larger volume of data also allowed for the effect of speech rate to be included, which is important as the continuous variable of speech rate was shown to have explanatory value in this spontaneous speech data. Table 3.59 displays the counts of word-final consonant clusters from the spontaneous portions of the two corpora according to motivation for simplification and following context.

			FR		QC		Sum
			AssNat	PFC	AssNat	PFC	
C	Simplify	Stop-final	44	247	68	802	1161
		SSP	576	2223	757	3827	7383
	Stable		163	1667	249	2647	4726
V	Simplify	Stop-final	24	199	37	365	625
		SSP	136	941	249	1565	2891
	Stable		25	314	77	381	797
Sum			968	5591	1437	9587	17583

Table 3.59: Frequency of word-final consonant clusters in spontaneous, non-read speech

To begin, Table 3.60 displays the proportion of simplification and schwa insertion following a word-final consonant cluster according to corpus and dialect for the spontaneous, non-read portion of the two corpora. Judging from the figures expressed in Table 3.60, it is likely that differences between the two corpora for the spontaneous speech data exist. In both dialects, the proportion of word-final consonant cluster simplification is higher for the PFC corpus than for the AssNat corpus, while the proportion of schwa insertion is lower. The strength of this corpus effect can be determined through a logistic regression, similar to the ones performed earlier, that also take into account the variance explained by following context, motivation for simplification and speech rate.

	FR		QC	
Response	AssNat	PFC	AssNat	PFC
Simplification	20.97%	36.45%	49.97%	54.52%
Schwa	44.63%	18.30%	12.94%	4.66%

Table 3.60: Proportion of simplification and schwa insertion by corpus

Simplification according to corpus: Spontaneous speech

The table of coefficients for the model predicting the probability of word-final consonant cluster simplification by corpus, motivation for simplification, following context and rate in the dialect from Québec is given below in Table 3.61. This table indicates that the coefficient for the motivation for simplification (SSP) is not significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−1.90984	0.38105	−5.012	$5.39e-07$	***
CorpusPFC	2.84135	0.40991	6.932	$4.16e-12$	***
SSP	0.07933	0.07715	1.028	0.304	
Vowel	−1.30033	0.05890	−22.076	$< 2e-16$	***
Rate	0.21810	0.02905	7.507	$6.07e-14$	***
CorpusPFC:Rate	−0.15305	0.03154	−4.853	$1.22e-06$	***

Table 3.61: Coefficients for model predicting simplification in spontaneous speech for Québec

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.62, indicates that all terms, with the exception of SSP, are on their own significant.

Simplifying the model by removing the term for SSP and comparing the simplified model with the initial model returns $\chi^2(1) = 1.0491, p = 0.3057$. This means that, in this case, model simplification is justified, since removing the term for SSP from the model did not cause a significant reduction in the explanatory power of the model. Removal of any of the other terms in the model resulted in significant decreases in the explanatory power of the model and they must, therefore, be retained. The table of

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			7669	8176.9	
Corpus	1	113.28	7668	8063.6	$< 2.2e-16$
SSP	1	2.43	7667	8061.2	0.1194
Vowel	1	512.49	7666	7548.7	$< 2.2e-16$
Rate	1	64.84	7665	7483.8	$8.116e-16$
Corpus:Rate	1	24.52	7664	7459.3	$7.347e-07$

Table 3.62: Analysis of deviance table for Québec

coefficients for the final model is presented as Table 3.63.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−1.84167	0.37520	−4.908	$9.18e-07$	***
CorpusPFC	2.84258	0.40990	6.935	$4.07e-12$	***
Vowel	−1.30194	0.05888	−22.112	$< 2e-16$	***
Rate	0.21839	0.02905	7.517	$5.60e-14$	***
CorpusPFC:Rate	−0.15364	0.03153	−4.872	$1.10e-06$	***

Table 3.63: Coefficients for minimally adequate model predicting simplification in spontaneous speech for Québec

The coefficients displayed in Table 3.63 suggest several interesting things. First, as was suggested by the higher proportion of occurrences of word-final consonant cluster simplification in the PFC corpus, the coefficient in the model that expresses the difference in the log odds of simplification indicates that word-final consonant cluster simplification is significantly more likely to occur in the PFC corpus than in the AssNat corpus. Second, similar to the findings from the model that examined these two corpora together, there does not appear to be enough statistical evidence to conclude that there are differences in rates of simplification between SSP violating and stop-final consonant clusters in the spontaneous speech data from Québec. Third, again similar to the earlier findings, the large negative effect of following context, as given by the coefficient for `Vowel`, indicates that word-final consonant cluster simplification is significantly less likely when the consonant cluster is prevocalic, than when it is preconsonantal. Fourth, the effect of speech rate, which was shown to have explanatory value in all the spontaneous speech data in general, also had a significant effect in the spontaneous speech data for each of these two corpora: Faster rates of speech were associated with higher probabilities of word-final consonant cluster simplification. An interesting observation that emerges from this model is that the effect of speech rate appears to have less of an effect in the data from the PFC corpus than in the data from the AssNat corpus. These relationships are expressed visually in Figure 3.12.

The table of coefficients for the model predicting the probability of word-final consonant cluster simplification by corpus, motivation for simplification, following context and rate in the dialect from France is given below in Table 3.64. This table indicates that the interaction term between corpus and

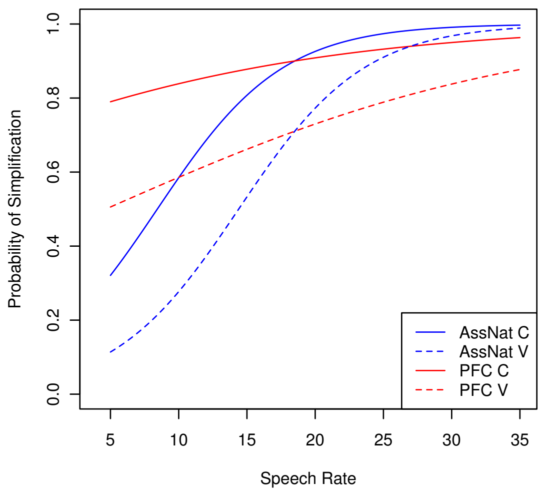

Figure 3.12: Regression predicting simplification in spontaneous speech data for Québec

speech rate (`CorpusPFC:Rate`) is not significant.

`Coefficients:`

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-3.37854	0.53315	-6.337	$2.34e-10$	***
CorpusPFC	2.01729	0.55387	3.642	0.00027	***
SSP	1.10989	0.11123	9.978	$< 2e-16$	***
Vowel	-1.58299	0.07665	-20.652	$< 2e-16$	***
Rate	0.11470	0.03860	2.972	0.00296	**
CorpusPFC:Rate	-0.02866	0.04095	-0.700	0.48394	

Table 3.64: Coefficients for model predicting simplification in spontaneous speech for France

The analysis of deviance table produced by an ANOVA summary for this model, presented below as Table 3.65, indicates that all terms, with the exception of the interaction term `Corpus:Rate`, are on their own significant.

Simplifying the model by removing the interaction term and comparing the simplified model with the initial model returns $\chi^2(1) = 0.49192, p = 0.4831$. Removal of any of the other terms in the model resulted in significant decreases in the explanatory power of the model. The table of coefficients for the final model is presented as Table 3.66.

Very similar to what was found for the spontaneous speech data from Québec, the coefficients in Table 3.66 suggest that the effects of these variables on the probablity of word-final consonant cluster simplification in the spontaneous speech data of the two corpora from France are comparable to the effects discovered during the comparison of read versus non-read speech: Word-final consonant cluster simplification is more likely to be observed following an SSP violating than before a stop-final consonant

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			4389	6083.9	
Corpus	1	245.32	4388	5838.6	$< 2.2e-16$
SSP	1	139.07	4387	5699.5	$< 2.2e-16$
Vowel	1	474.21	4386	5225.3	$< 2.2e-16$
Rate	1	48.33	4385	5177.0	$3.611e-12$
Corpus:Rate	1	0.49	4384	5176.5	0.4831

Table 3.65: Analysis of deviance table for France

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−3.04058	0.22280	−13.647	$< 2e-16$	***
CorpusPFC	1.63586	0.09379	17.441	$< 2e-16$	***
SSP	1.11013	0.11129	9.975	$< 2e-16$	***
Vowel	−1.58293	0.07669	−20.639	$< 2e-16$	***
Rate	0.08927	0.01294	6.900	$5.19e-12$	***

Table 3.66: Coefficients for minimally adequate model predicting simplification in spontaneous speech for France

cluster; is more likely to occur in a preconsonantal context than a prevocalic context; and is positively associated with faster rates of speech. When looking at just these spontaneous, non-read speech data, there is not enough evidence to suggest a difference between the two corpora for the effect of speech rate. In support of the observation that the proportion of occurrences of word-final consonant cluster simplification was higher in the PFC corpus, the coefficient `CorpusPFC` indicates that simplification of the consonant cluster is significantly more likely to occur in this corpus than in the AssNat corpus. These relationships are displayed in Figure 3.13.

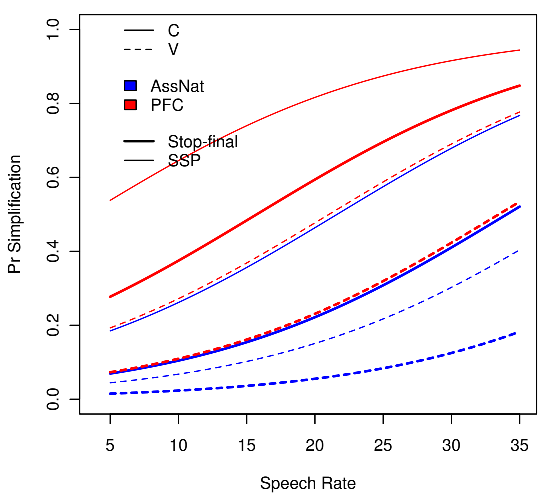

Figure 3.13: Regression predicting simplification in spontaneous speech data for France

Schwa insertion according to corpus: Spontaneous speech

The table of coefficients for the model predicting the probability of schwa insertion following a word-final consonant cluster by corpus, motivation for simplification, following context and rate in the dialect from Québec is given below in Table 3.67. This table indicates that the coefficient for following context (`Vowel`) is not significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.28954	0.43931	-0.659	0.50985	
CorpusPFC	-2.38616	0.49501	-4.820	$1.43e-06$	***
PredictSimplify1	0.84089	0.11402	7.375	$1.64e-13$	***
Vowel	0.14614	0.09211	1.587	0.11262	
Rate	-0.18564	0.03444	-5.390	$7.05e-08$	***
CorpusPFC:Rate	0.10809	0.03914	2.762	0.00575	**

Table 3.67: Coefficients for model predicting schwa insertion in spontaneous speech for Québec

The analysis of deviance table produced by an ANOVA summary for this model is presented below as Table 3.68. This table suggests that the factor of following context (`Vowel`) on its own may be important ($p = 0.042755$) even though the variance explained by the other factors renders its effect null in the model ($p = 0.11262$). In fact, simplifying the model by removing this term resulted in a $\chi^2(1) = 2.4763, p = 0.1156$ which indicates that this term did not significantly add to the explanatory power of the model as a whole.

The table of coefficients for the final model is presented as Table 3.69. These coefficients indicate that in the spontaneous speech data from Québec, schwa insertion following a word-final consonant cluster is less likely to occur in the PFC corpus than in the AssNat corpus; is more likely to occur following consonant clusters that also may simplify; and is negatively associated with speech rate. Similar to what was found for the response of simplification in this dialect, the effect of speech rate was less strongly observed in the PFC corpus. These results, including the lack of an effect of following context, mirror the results that were obtained when the two corpora were combined as a single spontaneous speech style. A plot of the regression equation predicting schwa insertion following a word-final consonant cluster in

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			11023	4846.4	
Corpus	1	125.485	11022	4720.9	$< 2.2e-16$
PredictSimplify	1	69.353	11021	4651.5	$< 2.2e-16$
Vowel	1	4.105	11020	4647.4	0.042755
Rate	1	40.287	11019	4607.1	$2.193e-10$
Corpus:Rate	1	7.775	11018	4599.3	0.005298

Table 3.68: Analysis of deviance table for Québec

the spontaneous, non-read portion of the data from Québec is provided as Figure 3.14.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−0.25210	0.43860	−0.575	0.56544	
CorpusPFC	−2.37068	0.49483	−4.791	$1.66e-06$	***
PredictSimplify1	0.86105	0.11326	7.603	$2.90e-14$	***
Rate	−0.18689	0.03443	−5.429	$5.68e-08$	***
CorpusPFC:Rate	0.10717	0.03913	2.739	0.00616	**

Table 3.69: Coefficients for minimally adequate model predicting schwa insertion in spontaneous speech for Québec

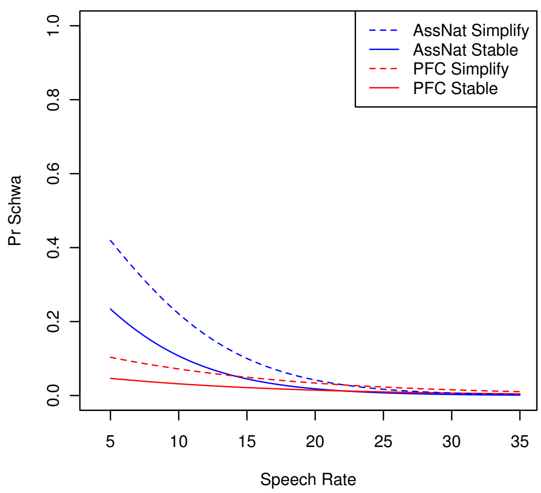

Figure 3.14: Regression predicting schwa insertion in spontaneous speech data for Québec

The table of coefficients for the model predicting the probability of schwa insertion following a word-final consonant cluster by corpus, motivation for simplification, following context and rate in the dialect from France is given below in Table 3.70. This table indicates that the coefficients for motivation for simplification (`PredictSimplify`) as well as the interaction between corpus and speech rate (`Corpus:Rate`) are not significant.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.04232	0.40040	2.603	0.00924	**
CorpusPFC	−0.83264	0.43441	−1.917	0.05527	.
PredictSimplify1	0.06289	0.06757	0.931	0.35198	
Vowel	−0.76198	0.08180	−9.315	$< 2e-16$	***
Rate	−0.08924	0.02984	−2.991	0.00278	**
CorpusPFC:Rate	−0.02929	0.03267	−0.896	0.37000	

Table 3.70: Coefficients for model predicting schwa insertion in spontaneous speech for France

	df	Deviance	Resid. df	Resid. Dev	Pr(>Chi)
NULL			6558	6942.3	
Corpus	1	290.347	6557	6651.9	$< 2e-16$
PredictSimplify	1	0.001	6556	6651.9	0.9805
Vowel	1	86.021	6555	6565.9	$< 2e-16$
Rate	1	89.135	6554	6476.8	$< 2e-16$
Corpus:Rate	1	0.801	6553	6476.0	0.3709

Table 3.71: Analysis of deviance table for France

The analysis of deviance table produced by an ANOVA summary for this model is presented below as Table 3.71. This table shows that the factors of motivation for simplification (`PredictSimplify`) and the interaction term (`Corpus:Rate`) are not significant. Simplifying the model by first removing the interaction term resulted in a $\chi^2(1) = 0.80073, p = 0.3709$. Further simplifying the model by removing the term for motivation for simplification gave a $\chi^2(1) = 0.89364, p = 0.3445$. Given the non-significant decrease in the explanatory power of the model by either of these removals means that the final, minimally adequate model contains terms only for corpus, following context and speech rate, all of which are significant as is shown by Table 3.72.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.41835	0.17359	8.171	$3.07e-16$	***
CorpusPFC	−1.22677	0.07455	−16.456	$< 2e-16$	***
Vowel	−0.75036	0.08116	−9.245	$< 2e-16$	***
Rate	−0.11424	0.01215	−9.401	$< 2e-16$	***

Table 3.72: Coefficients for minimally adequate model predicting schwa insertion in spontaneous speech for France

The model predicting a reponse of schwa insertion following a word-final consonant cluster for the spontaneous speech data from France contains similar terms with similar effects to the model predicting this response for the read versus non-read data from France examined earlier. In just these spontaneous data, the effect of following context and speech rate are both negatively associated with a response of schwa insertion: Schwa insertion following a word-final consonant cluster in the spontaneous speech data from France is less likely when the consonant cluster is prevocalic than when preconsonantal and less likely as speech rates increase. The likelihood of schwa insertion is also lower for the PFC corpus than it is for the AssNat corpus. No difference was found between consonant clusters that also participated in simplification (Simplify) and those that did not (Stable). The negative effect of speech rate was determined to be the same in both corpora. These findings are presented in Figure 3.15.

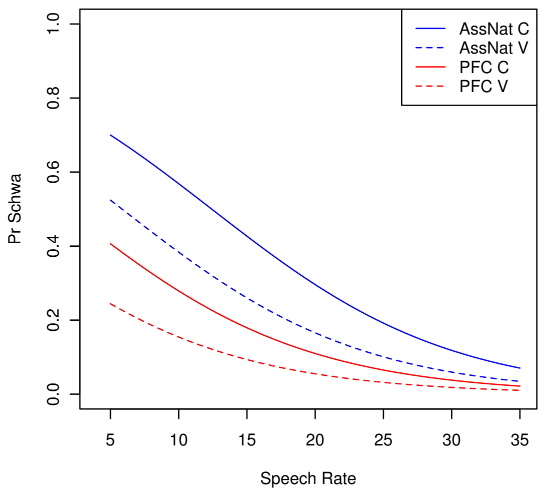

Figure 3.15: Regression predicting schwa insertion in spontaneous speech data for France

3.6.3 Summary of corpus effects

The results of exploring the differences between the two corpora for both the read and non-read portions of the data are interesting, particularly in light of the earlier findings that investigated the differences between read and non-read speech. Whether examining the differences between reading and spontaneous speech or between corpora, the effects of following context and motivation for simplification did not change. For the dialect from Québec, whether or not a word-final consonant cluster was preconsonantal or prevocalic was predictive for a response of simplification, but not for schwa insertion. For the dialect from France, following context was shown to be predictive for both reponses of simplification and schwa insertion. For the response of simplification, the motivation for simplification (SSP violating or stop-final) was predictive in the dialect from France, though not in the dialect from Québec. For the response of schwa insertion, the motivation for simplification (Simplify or Stable) was predictive in the dialect from Québec, but not from France. Similarly, the effect of speech rate, which was determined to have an effect only in the spontaneous style of speech, emerged as predictive in both corpora: For the spontaneous speech data from both the PFC and the AssNat corpora, faster rates of speech were associated with higher probabilities of simplification and lower probabilities of schwa insertion, though this effect was perhaps not as strong in the PFC corpus representing the Québec dialect.

The most interesting result is certainly the fact that differences between the two corpora were discovered and these differences were statistically maintained even after taking into account the variance explained by the other factors. In both dialects, the proportion of occurrences of word-final consonant cluster simplification was higher in the PFC corpus while the proportion of occurences of schwa insertion following a word-final consonant cluster was lower and the differences between the two corpora were statistically significant. Table 3.73 shows the proportion of either the response of simplification or schwa insertion according to dialect and corpus with the direction of change between the two corpora indicating the magnitude of the differences.

	FR			QC		
Response	AssNat		PFC	AssNat		PFC
Reading						
Simplification	24.05%	≈	22.92%	55.02	<	63.82%
Schwa	65.05%	>	54.40%	31.44%	>	19.47%
Spontaneous						
Simplification	20.97%	<	36.45%	49.97%	<	54.52%
Schwa	44.63%	>	18.30%	12.94%	>	4.66%

Table 3.73: Proportion of responses of simplification or schwa insertion by corpus

While the results from just the read portion of the data may not be representative, since they include only a small subset of the available word-final consonant clusters made available by the data set as a whole, the general pattern expressed in Table 3.73 is in line with the assumption that a more formal setting, as opposed to a less formal setting, is likely to be associated with higher probabilities of observing a response of schwa insertion and lower probabilities of observing a response of simplification following a word-final consonant cluster. Since it was true that the effects of following context and motivation for simplification did not change when switching from an analysis of reading versus spontaneous speech styles to one of professional speakers in a more formal setting versus non-professional speakers in a less formal setting, these results might lead to the conclusion that they better indicate an effect of speech style: With respect to the variable pronunciations of word-final consonant clusters, more formal speech styles, such as political speeches delivered in a national assembly, tend to show higher rates of schwa insertion and lower rates of simplification. Less formal speech styles, such as casual conversations among friends, tend to show instead higher rates of simplification and lower rates of schwa insertion. This conclusion can only be preliminary though. These results equally support a hypothesis that professional speakers, such as members of a national assembly, tend to show higher rates of schwa insertion and lower rates of simplification while non-professional speakers, such as the participants in the PFC surveys, tend to show lower rates of schwa insertion and higher rates of simplification.

3.7 Chapter summary and discussion

This chapter was an investigation into the variable pronunciations of word-final consonant clusters in two dialects of French: A dialect representative of the French spoken in Québec and a dialect representative of the French spoken in the north of France. The investigation used a corpus of more than 100 hours of recorded, natural language speech. The corpus was collected from publicly available material provided by the national assemblies of both Québec and France, as well as recorded interviews made available by researchers associated with the Projet Phonologie du Français Contemporain. The complete corpus was forced aligned at the word and phoneme level by the SPLaligner tool, created expressly for this investigation and described in Chapter 2. The large volume of data afforded by this approach (more than $28,000$ observations of word-final consonant clusters) allowed the investigation to determine the effect of variables, such as speech style, following context, motivation for simplification and speech rate, on the variable pronunciations of word-final consonant clusters.

The variables that were investigated were selected primarily because the body of literature on both simplification and schwa insertion following a word-final consonant cluster indicated that they were all likely to have explanatory value in any models predicting either simplification or schwa insertion. Secondarily, while there are undoubtedly other variables that are likely to be important predictors for simplification or schwa insertion following a word-final consonant cluster, it was possible to control for these selected variables while still obtaining a large and robust sample size for analysis. Incorporation of other variables of interest, such as lexical effects, inter- or even intra-speaker variation, produced dramatically unbalanced experimental designs, even in this relatively large corpus.

The collection of a large amount of data combined with the systematic identification of the different pronunciations of word-final consonant clusters made it possible to investigate the dialectal differences in the effects of the variables of interest. While many of the variables selected for this investigation had been previously investigated, either with respect to simplification or schwa insertion following a word-final consonant cluster, in one or the other of the two dialects represented here, a systematic comparison between the two dialects had not been fully explored in the literature to date. The relatively large amount of data collected from each dialect in this corpus made such dialectal comparisons possible. The corpus approach adopted here also made it possible to investigate the combined effects of the selected variables. It has often been acknowledged that explaining the complicated patterns of distribution of the variable pronunciations of word-final consonant clusters in French would most likely involve a number of explanatory factors. Examining the combined effects of several predictor variables can be a challenge with smaller sample sizes collected experimentally or through introspection. The relatively large sample extracted from this corpus allowed for adequate controls in order to quantify the direction and strength of the relationship between the variables of interest and the responses of simplification or schwa insertion following a word-final consonant cluster. Finally, the structure of the corpus, containing numerous observations drawn not just from different dialects, but also from different styles of speech, allowed the investigation to examine whether or not a relationship exists between the two responses of simplification and schwa insertion following a word-final consonant cluster. Since the two share a similar target (a word-final consonant cluster) and may share a set of explanatory variables (eg. style, context and speech rate), could the occurence of one depend in some way on the occurence of the other? A very few number of researchers had previously speculated about the possible existence of such a relationship. As well, a prior pilot study suggested that schwa insertion may depend upon simplification, at least for the dialect from Québec. This corpus provided an opportunity to more fully explore this question.

The main findings from the various analyses performed in this chapter are, on the one hand, in general agreement with much previous research into both simplification and schwa insertion following a word-final consonant cluster. None of the results of the exploration of the effects of speech style, following context, motivation for simplification or speech rate, either individually or in combination, contradicted the conclusions summarized in the literature review. On the other hand, several findings contained in this chapter do provide new information regarding the different strengths of the effects of these variables in each dialect and point to an intriguing possibility that the relationship between schwa insertion and simplification is different within each dialect.

When exploring the frequency of occurences of simplification according to individual consonant clusters, it was expected to be the case that word-final consonant cluster simplification would be restricted to a subset of the 29 consonant clusters examined in the corpus. The division into a group of consonant clusters that would regularly participate in simplification and a group of consonant clusters that would infrequently or rarely participate in simplification has been predicted for both dialects. In both the dialect from Québec and the dialect from France, word-final consonant clusters whose final consonant is more sonorous than the preceding one, thus violating the Sonority Sequency Principle which states that sonority should fall from the nucelus of the syllable to its edges, were expected to regularly participate in simplification. In the data presented in this chapter it was shown that this was, indeed, the case. Furthermore, for the dialect from Québec, it was expected that the group of consonant clusters containing a perceptually deficient final stop consonant would also regularly participate in simplification. This was also shown to be the case. Surprisingly, the same group of stop-final consonant clusters in the data from France showed higher than expected occurrences of simplification. The result was that in both dialects the division between 'Simplify' and 'Stable' consonant clusters was the same. Consonant clusters that either violated the SSP or contained a final stop consonant regularly participated in simplification, while other consonant clusters (in these data they were all /ʁ/-initial or /ks/) did not participate in simplification. It was further determined that, after taking account of the effects of speech style, following context and speech rate, the odds of observing word-final consonant cluster simplification in SSP violating consonant clusters was signficantly higher than in stop-final consonant clusters in the dialect from France, while this difference did not exist in the dialect from Québec: Both SSP violating and stop-final consonant clusters were equally likely to participate in simplification for that dialect. This finding was the same whether the investigation looked at differences between read and non-read speech or at differences between formal and less formal corpora. These results suggest that while it was right for previous research into word-final consonant cluster simplification in the dialect from France to concentrate on SSP violating consonant clusters (mainly /ʁ/ and /l/ final consonant clusters) since they do simplify more frequently, word-final consonant cluster simplification following a stop-final consonant cluster in this dialect is not so uncommon as to be dismissed out of hand.

Exploring the frequency of occurrences of schwa insertion according to individual consonant clusters, the literature reviewed suggested that schwa insertion should regularly occur following all consonant clusters in the dialect from France. In fact, in these data 28 of the 29 consonant clusters displayed regular occurrences of schwa insertion (only /ks/ did not). Since schwa insertion following a word-final consonant cluster occurs infrequently in the dialect from Québec, the literature reviewed suggested it might mostly occur following just SSP violating clusters. The pilot study upon which this project was based suggested further that schwa insertion in the dialect from Québec might be restricted to the same group of consonant clusters identified as being eligible for simplification. In fact, in these data nearly 90% of the observed cases of schwa insertion (1171 of 1311) occurred following SSP violating or perceptually deficient stop-final clusters in the dialect from Québec. Looking further at whether schwa insertion depends, or is affected by, these differences between 'Simplify' and 'Stable' consonant clusters it was determined that, after taking account of the effects of speech style, following context and speech rate, while schwa insertion in the dialect from France was slightly more likely following SSP violating and stop-final ('Simplify') consonant clusters than when following /ʁ/-intial ('Stable') consonant clusters, schwa insertion following 'Simplify' consonant clusters in the dialect from Québec was more than 3 times more likely than when following 'Stable' consonant clusters. This result was also unchanged between the two types of style definitions: read as opposed to non-read or national assembly corpus as opposed to PFC corpus. These results are taken as evidence that, in the dialect from Québec, schwa insertion, although infrequent, is most likely to occur in just those consonant clusters that also participate in simplification.

The effect of following context showed the most interesting dialectal differences. It was expected that the following context, specifically a preconsonantal context, would play a role in predicting the variable pronunciations of word-final consonant clusters. A word-final consonant cluster followed by a consonant-initial word, thus producing a sequence of three or more consonants, has long been understood as the context in which both simplification and schwa insertion most frequently occur. In these data it was shown that this was the case. In both dialects, both simplification and schwa insertion more frequently occurred when the consonant cluster was preconsonantal than when prevocalic. However, after taking into account the combined effects of the other variables of interest, it was determined that the effect of following context on the response of simplification was approximately the same in both dialects, but the effect on the response of schwa insertion was very different in the data from Québec than in the data from France. This dialectal difference emerged both when looking at the corpus with speech style classified as either reading or spontaneous, as well as when style was classified according to corpus: AssNat as an example of more formal and PFC as an example of less formal. The effect of following context in the dialect from France was approximately the same on both the responses of simplification and schwa. This suggests that the motivation for both simplification and schwa insertion in this dialect may be the same: An avoidance of sequences of consonants. However, the effect of following context in the dialect from Québec was much smaller for the response of schwa insertion than for the response of simplification. In fact, in the smaller sub-sample that controlled for speech rate, no effect of following context was found. This suggests that the motivations for simplification and schwa insertion are different in this dialect. While simplification of the consonant cluster was more likely in a preconsonantal context, thus suggesting a similar avoidance of sequences of consonants, schwa insertion following the consonant cluster is equally (un)likely whether in a preconsonantal or prevocalic context, thus suggesting some other motivation.

When considering the effect of following context, it must be remembered that the accuracy of forced alignment was negatively impacted by this variable. The levels of agreement between manual and automatic methods of coding were significantly lower when the word-final consonant cluster was in a prevocalic context. This suggests that the data produced by forced alignment in this prevocalic environment may be significantly different from the same data as produced by a human researcher. It bears repeating though, that the accuracy of the forced aligner was not determined on the basis of whether a particular pronunciation *actually* occurred but on whether the pronunciation *agreed* with the perception of a human judge. A system of forced alignment, as a statistical pattern classifier, given similar intervals of acoustic data will return similar results and will do so repeatedly. The human perception system is not the same thing. Several authors have reported that different judges will report different perceptions even when all presented with the same stimuli (Saraclar and Khudanpur, 2004; Goldman, 2011; Bürki et al., 2011c) and even when the same judge is repeatedly presented with the same stimuli (Bürki et al., 2011c).

In any case, if schwa following a word-final consonant cluster is truly unexpected in a prevocalic context, then the fact that the results of forced alignment, which determined a pronunciation of 'Full' (i.e. neither schwa nor simplification) 90% of the time in a prevocalic context when it disagreed with the results of manual coding suggests that, at least, the acoustic information in this context may be less like the acoustic information of unambiguous schwa. This could be better determined by, for example, a similar experiment to the one performed by Bürki et al. (2011c) which compared the duration of the segment along with the formant structure to conclude that many instances of shorter, weaker schwa vowels were likely to be inconsistently identified.

It might be fair to suggest that the surprising and unexpected occurrences of schwa insertion following a word-final consonant cluster when in a prevocalic context warrant further investigation, perhaps most importantly for the conversational portion of the PFC corpus pertaining to the dialect from Québec, where schwa insertion in this context was expected to be extremely rare. Because the results of forced alignment were generally less accurate in this context, a closer look at these cases might indicate that some of the tokens of realized schwa are predictable, perhaps some are outright errors and perhaps some might be considered as ambiguous as determined either perceptually or empirically.

A plausible other motivation for schwa insertion in the dialect from Québec might have been speech style. Several authors have noted that schwa insertion in the dialect from Québec may be more frequent in a more formal speaking style. The analyses presented in this chapter do provide evidence that speech style may have an effect on the variable pronunciations of word-final consonant clusters, but it was difficult to draw a meaningful conclusion regarding the effect of speech style. In both dialects it appeared to be the case that schwa insertion was more frequently observed in the reading speech style, while simplification was more frequently observed in the spontaneous speech style. However, it also appeared to be the case that rates of speech were faster in the spontaneous speech style and slower in the reading speech style and further that speech rate was positively associated with a response of simplification and negatively associated with a reponse of schwa insertion. Disentangling this potential confound proved difficult. As well, at least for the spontaneous, non-read portions of the speech data, in both dialects schwa insertion was more frequently observed in the more formal national assembly corpus while simplification was more frequently observed in the less formal PFC corpus, although an effect of speech rate was observed in both cases. The interpretation of the results presented in this chapter was that the effect of speech style, when it was defined as reading versus spontaneous, was best interpreted as the different sensitivity to speech rate that existed between the two styles: While little or no effect of speech rate was observed in the reading speech style for either simplification or schwa insertion, faster speech rates were positively associated with simplification and negatively associated with schwa insertion in the spontaneous speech style. When speech style was determined according to corpus, for the spontaneous portion of the speech data faster rates of speech were associated with higher probabilities of simplification and lower probabilities of schwa insertion in both the more formal AssNat corpus as well as the less formal PFC corpus. This was true of both dialects. Therefore, it remains difficult to conclude whether the variable pronunciations of word-final consonant clusters in either dialect are more affected by changes in speech style or speech rate.

The results discussed above are interesting because they confirm many conclusions of previous work on these two processes, which were largely made based on introspection or laboratory experiments, and demonstrate how these earlier claims are borne out as stasticially significant patterns in real world speech. The effects of speech rate, speech style and following context all suggest a cross-linguistically common pattern of reduction. The analyses conducted here conclude that simplification of word-final consonant clusters is more frequent at faster rates of speech, in less formal speaking styles, and in contexts where a sequence of consonants can be avoided. The effects of speech rate and speech style suggest that schwa insertion might be better considered as a marker of careful speech.

More interesting is what these results seem to indicate about the relationship of schwa insertion and simplification to each other within each dialect. The main hypothesis is that the relationship between these two processes is different within in each dialect. In the dialect from France, the analyses presented here provide evidence that both schwa insertion and simplification work together to avoid sequences of consonants. In effect, in this dialect these two phenomena may be conspiring (Kisseberth, 1970) to avoid sequences of consonants. In the dialect from Québec, while simplification exists to avoid specific sequences of consonants (similar to the dialect from France), schwa insertion serves to avoid simplification.

The evidence in support of this view can be found primarily in three places. First, the fact that schwa insertion was observed to regularly occur following all but one of the investigated clusters in the dialect from France, while schwa insertion occurred almost exclusively following only those clusters that also participated in simplification in the dialect from Québec. Second, the much larger effect of following context on the response of schwa in the dialect from France than in the dialect from Québec, and the fact that the effect of following context was very similar for both simplification and schwa in the dialect from France. Third, the final section of this chapter which examined whether the frequency with which a consonant cluster simplifies helps to predict the frequency with which schwa insertion occurs.

The results presented indicate this is not the case for the dialect from France, but the hypothesis was supported in the dialect from Québec: After controlling for rate of speech, the probability of observing schwa insertion was positively associated with the frequency of simplification in this dialect. In the dialect from Québec, knowing something about a consonant cluster's frequency of simplification helps to know whether or not schwa insertion may also occur following that consonant cluster. In the dialect from France, knowing something instead about the following context helps to know whether or not schwa insertion will occur.

In conclusion, the results presented in this large sample corpus analysis suggest that the dialectal differences between Québec and France, with respect to the variable pronunciations of word-final consonant clusters, are not as simple as "simplification in Québec, schwa insertion in France". While the results presented here suggest that the process of simplification following a word-final consonant cluster is similar in both dialects, the process of schwa insertion is likely to be different in each dialect. Word-final consonant cluster simplification appears to be explained equally well in both dialects by the proposal presented in Côté (2004a), although the constraints protesting against a perceptually deficient final stop consonant may be less strongly observed in the dialect from France. In both dialects the direction and strength of the relationship between following context and simplification is similar: Word-final consonant cluster simplification is more likely in a preconsonantal context. In both dialects the direction and strength of the effects of the interaction between speech style and speech rate are similar: Word-final consonant cluster simplification is more likely in a spontaneous speech style and in that speech style is positively associated with speech rate. Schwa insertion following a word-final consonant cluster displays much stronger dialectal differences. Schwa insertion in the dialect from France is more strongly affected by following context and the process of simplification has little explanatory value. Schwa insertion in the dialect from Québec is not affected by following context and the process of simplification has strong explanatory value.

Chapter 4

Conclusion

This thesis should be of interest to researchers in linguistics, French description and dialectology, sociolinguistics, and automatic speech recognition. It contributes to our knowledge of French word-final consonant clusters in both a descriptive and theoretical sense by providing new information about the phenomena of word-final consonant cluster simplification and schwa insertion in these two dialects of French which have not yet been investigated from this perspective. Using an innovative research methodology of the application of speech recognition tools to linguistic data analysis, this thesis helps to clarify the role of many phonological and stylistic factors in the variable pronunciations of word-final consonant clusters in French. The analyses presented in this thesis demonstrate that the effects of most factors on simplification and schwa insertion are qualitatively the same in each dialect, but sometimes differ substantially in quantitative strength. The differences in the effects of different factors between the dialects points towards an interesting finding that the two phenomena are related in each dialect, but in different ways: In Québec, schwa insertion serves to avoid cluster reduction, in France schwa serves to avoid consonant clusters.

This thesis also provides a practical contribution to the field. The development and evaluation of several systems of automatic forced alignment will be useful to others interested in using a forced aligner to label large corpora of speech in future studies. The identification of the best performing acoustic model (speaker-adapted cross-word triphones), while not surprising from an automatic speech recognition perspective, is nevertheless important from the perspective of linguistic research by demonstrating how varying acoustic models affects error in automatically transcribing a variable of linguistic interest using a forced aligner. This thesis also provides the first instance of a forced aligner trained to maximize performance on annotating a particular case of phonological variation and determines that the types of disagreement were similar to those between human transcribers for the presence or absence of /ə/.

This chapter begins with a restatement of the nature of the problem posed by the variable pronunciations of word-final consonant clusters in the French language. It then describes a relatively new approach of using the tool of forced alignment to obtain the data required in order to gain a better understanding of the effect of variables, such as speech style, following context, motivation for simplification and speech rate, on the variable pronunciations of word-final consonant clusters in two dialects of French. Next, the main findings that were presented in Chapters 2, on the accuracy of forced alignment and 3, on the distributions of the variable pronunciations of word-final consonant clusters, are summarized along with a discussion of their implications. Finally, this chapter concludes with some thoughts on interesting directions for future research.

4.1 Restatement of research objectives

In the French language, there are two well known phonological operations that target consonant clusters at word boundaries: cluster simplification and schwa insertion. Consonant cluster simplification involves, in most cases, the deletion of the final consonant of the cluster. Schwa insertion has occurred when a final schwa vowel is realized following the consonant cluster. Both are understood to alter a sequence of (phonological) consonants in order to either facilitate articulation or enhance perception of the sequence. The variability of word-final consonant clusters in French has been investigated either from the perspective of consonant cluster simplification or from the perspective of the word-final $\emptyset \sim$/ə/ alternation. Many of these studies have been based on data from introspection; some have used corpora of natural language; while a few more recent studies have been based on large databases of recorded speech. For word-final consonant clusters, both simplification and schwa insertion are highly variable in their application. Even when a particular motivation has been proposed for either schwa insertion or simplification following a word-final consonant cluster, most authors have acknowledged that the application of either remains optional in most circumstances. This optionality has made a concise description of the distribution of the variable pronunciations of word-final consonant clusters elusive. Notwithstanding the long history of investigation into schwa insertion and simplification following a word-final consonant cluster, no clear and complete picture has yet emerged of exactly which variables condition the choice of pronunciation: simplification or schwa? It has been claimed that these variable pronunciatons may be sensitive to various phonological factors (e.g. perceptual needs, co-articulation, prosodic requiements). As well, the different pronunciations of word-final consonant clusters may vary according to dialect, speech style or speech rate. Additionally, variables such as position of the word in the utterance, word-length and word-frequency have all been demonstrated to have an effect on either consonant cluster simplification or the insertion of schwa at word boundaries. Sociolinguistic factors may also play a role in influencing the variable pronunciations of word-final consonant clusters. This lack of consensus may be due to the fact that most studies consider variables, and pronunciations, individually, despite the fact that they often interact or are correlated.

This thesis project examined both schwa insertion and simplification following word-final consonant clusters in a large corpus of natural language spoken French containing examples from two dialects and two speech styles. The project was interested in determing what effect variables, such as speech style, following context, motivation for simplification and speech rate, have on the variable pronunciations of word-final consonant clusters in two dialects of French. The investigation selected these variables because they had all previously been shown to have explanatory value in predicting occurrences of both schwa insertion and simplification following word-final consonant clusters. A comparison between two dialects of French was done because, even though both consonant cluster simplification and schwa insertion at word boundaries appear to share similar contexts (a post-consonantal word boundary) and possibly similar motivations (avoidance of sequential consonants across a word boundary), it has long been observed that schwa insertion is more common in the one dialect (from France) while simplification is more common in the other (from Québec). A corpus analysis was adopted since it offered the greatest opportunity of obtaining a representative sample of the variable pronunciations of word-final consonant clusters distributed across several explanatory predictors.

The choice of a corpus analysis, as opposed to an introspective or experimentally elicited set of data, presented the investigation with both a practical problem and an opportunity. The value in working with a natural language corpus is the ability to collect large volumes of emprical data with which to test research hypotheses. The corpus that was assembled from the national assemblies of both Québec and France as well as investigations made available by researchers from the Projet Phonologie du Français Contemporain contained many hours of recorded speech of which only a small proportion had been previously analyzed to determine which one of the several possible pronunciations of the word-final consonant cluster was actually realized. Since manual segmentation is an incredibly time-consuming task and for the more than 100 hours of speech to be segmented it was inconceivable that a single researcher could manually label all the phones, the practical problem was how to quickly and accurately, with some degree of objectivity, identify the variable pronunciations of the word-final consonant clusters that occurred in the corpus. The proposed solution was to deploy an automatic forced aligner.

Forced alignment is typically a tool used in the process of constructing an automatic speech recognition system. In speech recognition, acoustic models are used to classify continuous speech into a set of subword units such as phones, context-dependent subphones etc. During forced alignment, orthographic transcripts of the utterances are used to constrain an optimal alignment between existing speech models and the new speech data. A word network is constructed from the given transcription and the role of forced alignment therefore is constrained to determine, given a dictionary with (often multiple) pronunciations listed for words, the most likely sequence of phones (i.e. the most likely pronunciation of each word). Unfortunately, there did not exist an efficient, reliable, freely available system of forced alignment capable of segmenting recorded French speech. Thankfully, the resources required to build a system of forced alignment are freely available and it was possible to build one. With the construction of a novel system of forced alignment capable of segmenting recorded French speech came the opportunity to determine how likely it was that a system of automatic forced alignment will produce judgments on the variable pronunciations of word-final consonant clusters that agree with manually identified pronunciations.

Therefore, the goals of this thesis project were twofold. The first aim of this thesis project involved evaluating a novel system of forced alignment capable of segmenting recorded French speech. On the one hand, from within a selection of different types of acoustic models, which set was consistently more accurate at producing judgments on the variable pronunciations of word-final consonant clusters that most closely match the judgments of a human listener? On the other hand, under which of a set of several conditions do the levels of agreement between manual and automatic coding significantly differ? The answer to the former gives a set of acoustic models to be used in forced alignment that will confidently reproduce pronunciation judgments similar to those of a human researcher. The answer to the latter gives a better understanding of the conditions under which the two methods of coding might be expected to disagree, and perhaps why.

The second aim of this thesis project was to determine how several variables effect the variable pronunciations of word-final consonant clusters in two dialects of French. This was accomplished by performing forced alignment on a large corpus of natural language spoken French containing examples of recordings from two dialects and two speech styles. In the first place, from the data provided by the use of forced alignment, how do variables such as speech style, following context, motivation for simplification and speech rate affect the probability of observing either schwa insertion or simplification following a word-final consonant cluster? In the second place, how do the strengths of these effects compare between two dialects of French, one of which seems to prefer consonant cluster simplification while the other seems to prefer schwa insertion? The answer to the first question helps to better describe the combined effects of speech style, following context, motivation for simplification and speech rate on the variable pronunciations of word-final consonant clusters in French. The answer to the second question gives a better understanding of the dialectal differences with respect to the variable pronunciations of word-final consonant clusters in French.

In Section 4.2 are described the main findings that were presented in Chapter 2. Section 4.3 contains a summary of the main findings that were presented in Chapter 3. Some implications of these findings are also discussed.

4.2 Main findings of forced alignment

The system that was eventually used to perform forced alignment on the complete corpus of recorded speech was built in four stages. At the completion of each stage, a completely functional system of forced alignment was produced. This resulted in four different sets of acoustic models being available to test which set was consistently more accurate at producing judgments on the variable pronunciations of word-final consonant clusters that most closely match the judgments of a human listener. The four sets included monophone, word-internal triphone, cross-word triphone and speaker adapted cross-word triphone acoustic models. All the models were trained and tested on 18 hours of speech data (9 hours per dialect) drawn from a corpus of debates from the national assemblies of both Québec and France. Additionally, in order to obtain a 'baseline' value of the likelihood that a forced aligner will agree with a researcher's judgments, forced alignment was performed with an aligner used previously: The Penn Phonetics Lab Forced Aligner (P2FA), modified to work with the French language by changing the labels of the HMMs to French phones, but leaving intact the actual HMMs themselves, contains monophone HMM models based on English language training data that have been labelled with French phones to work with French transcriptions.

The results presented in Section 2.3 suggest that, of the 5 types of acoustic models that were evaluated, the set of speaker dependent, continuous density, 8-mixture Gaussian, context-dependent, cross-word triphone models (SAT) proved to be the most accurate at agreeing with the manually judged variable pronunciations of word-final consonant clusters in these data. With an overall level of agreement of nearly 87% (85.95% in the dialect from France, 87.24% in the dialect from Québec), this set of acoustic models represented an improvement over the next best performing set of acoustic models (word-internal triphones) by a factor of 1.43, and a significant increase in performance over an initial set of acoustic models, a modified version of the Penn Phonetics Lab Forced Aligner, by a factor of 3.88.

The accuracy of the SAT model set did not suffer from the effects of speech style. The levels of agreement between manual and automatic methods of coding did not significantly change when switching from a read text to a spontaneous style of speech. Furthermore, the levels of agreement within each speech style ranged from 84.94% $\sim$ 89.03%, suggesting that, within each style of speech, this set of acoustic models accurately reflected the set of manually labelled pronunciations.

Neither was there any apparent change in the levels of agreement according to changes in speech rate. The logistic regression models showed that variable rates of speech did not have a significant impact on the levels of agreement between manual and automatic coding.

When a consonant cluster was preconsonantal (followed by a consonant intitial word in the absence of a pause), the levels of agreement were exceptionally high for the SAT models. When identifying both schwa insertion and simplification following a word-final consonant cluster, the automatic method of coding agreed with manual judgments between 93% $\sim$ 97% of the time. When the consonant cluster was prepausal (followed by a non-zero length `sp` phone), the levels of agreement were lower (levels of agreement between 71% $\sim$ 87%), but were still considered acceptable.

Of all the variables investigated, only the prevocalic context, where a word-final consonant cluster was followed by a vowel-intitial word in the absence of a pause, showed a significant decrease in the levels of agreement between manual and automatic methods of coding. Levels of agreement in this context ranged from a low of 54.72% to a high of 75.33%. Three possible explanations for this were suggested. They can summarized as either faulty models or faulty data, both of which may be contributory factors that gave rise to faulty recognition/perception.

Faulty models, or poor parameter estimation, may have been the result of sparse C-ə+V triphone examples in the training data. Owing to the fact that this is not a frequently occurring cross-word sequence of phones in French and it is an even less frequently occurring word-internal sequence of phones, it would be difficult to correct this possible problem with further training data. Faulty data, or ambiguous/indeterminate acoustic information, may have been the result of conflicting acoustic information in just this prevocalic context. A more detailed examination of the acoustic properties of the phones involved in these cases may help to determine whether the thresholds for perceiving a schwa vowel in a prevocalic context are different from those in a preconsonantal context. Faulty recognition/perception, or inconsistent labelling of pronunciations, may have been due to the notion that human listeners are more apt to perceive sounds they expect to hear, and ignore sounds they do not expect to hear, irrespective of the actual acoustic evidence, while a machine is apt to recognize a similar sound given similar acoustic evidence. A closer look at the acoustic record while comparing manual and automatically segmented sequences of phones may help to determine whether this is the case.

4.2.1 Implications

These results have implications both for the current study, in which the SAT models are used to perform forced alignment on a larger corpus, as well as for future investigations seeking to use forced alignment to harvest linguistic data from corpora of recorded speech. First of all, the generally high levels of agreement between the results of manual and automatic identification of the variable pronunciations of word-final consonant clusters should be taken as evidence in favour of the conclusion that, when used with a set of well trained acoustic models (in this case, speaker dependent, cross-word triphone models), the results of forced alignment are likely to be similar to the results produced by manual identification. The results of forced alignment are not likely to be affected by changes in speech style or speech rate. The results of forced alignment should strongly agree with the manual identification of the variable pronunciations of word-final consonant clusters when the consonant cluster is preconsonantal or prepausal. However, the results of forced alignment may differ significantly from manual identification when a consonant cluster is prevocalic.

This final point has special implications for the current study which concluded that the following context was a good predictor for the occurrences of simplification in both dialects, and was a good predictor for the occurrences of schwa insertion only in the dialect from France. However, since 90% of the time the results of forced alignment indicated an unmodified pronunciation (i.e. Full) instead of the inserted schwa or simplified consonant cluster indicated by the manual labelling when the two methods of coding disagreed in the prevocalic context, had the corpus analysis been instead based on completely manually coded pronunciations the magnitude of the effect of following context would be rendered smaller than what was discovered using the results of forced alignment. This would have the result of strengthening the conclusion that following context is not explanatory when predicting observations of schwa insertion in the dialect from Québec while the large size of the effect of following context in the dialect from France, combined with the infrequent observations of schwa insertion in a prevocalic context, make it likely that neither would the conclusions for that dialect have changed significantly.

The more pertinent question to be asked is: Are the levels of agreement between the two methods of coding that were obtained in this investigation *good enough* to warrant the conclusion that the results of forced alignment are likely to confidently reproduce pronunciation judgments similar to those of a human researcher? Based on the current level of understanding, in combination with the results of the multiple tests of agreement under various conditions described in Section 2.3, it is asserted that the results of forced alignment, when performed with the speaker adapted set of acoustic models, are likely to reproduce pronunciation judgments on the variable pronunciations of word-final consonant clusters that are similar to the judgments that would be obtained by manual identification. The justification for this conclusion are as follows.

First, the levels of agreement between manual and automatic methods of coding are likely to differ owing to the inevitable presence of human error; the likelihood of at least some inconsistently identified pronunciations by a human researcher; and the fact that human listeners rely on more than simple acoustic evidence when perceiving pronunciations. Segmentation by machine does not err; is entirely consistent; and the nuances of human perception cannot yet be modelled by a statistical pattern classifier. Therefore, a 100% level of agreement between manual and automatic methods of coding is, for all intents and purposes, unobtainable.

Second, the literature indicates that, at least according to some measures of agreement, the overall reliability of automatic labelling using forced alignment can be considered to be close to the one acheived by human annotators. These studies have usually measured levels of agreement according to the temporal placement of phone boundaries and are, therefore, not directly comparable to the results presented in the current study which measured levels of agreement according to selected pronunciations. However, these findings at least give an expectation that acceptable levels of agreement should be close to the 80% $\sim$ 90% range typically reported. With the exception of the low levels of agreement found for the prevocalic context, the reported overall level of agreement between manual and automatic methods of coding of 87% suggests that the results presented here are squarely in line with expectations.

Third, inquiring about the levels of agreement between manual and automatic methods of coding is akin to inquiring about levels of agreement between human labellers. While the literature on levels of agreement between human labellers is thin, the few reports that do exist suggest that levels of inter-annotator agreement may be lower than expected. Regarding the placement of temporal boundaries and, more relevant to the current study, levels of agreement regarding pronunciation variants, several authors reported levels of agreement between at least two trained experts of approximately 80%. If this level is accepted as being typical, then the results reported in this study support the argument that the levels of agreement between manual and automatic coding are consistent with expected levels of agreement between several human annotators.

4.3 Main findings of corpus analysis

The corpus analysis was an investigation into the variable pronunciations of word-final consonant clusters in two dialects of French: A dialect representative of the French spoken in Québec and a dialect representative of the French spoken in the north of France. The investigation used a corpus of more than 100 hours of recorded, natural language speech. The corpus was collected from publicly available material provided by the national assemblies of both Québec and France, as well as recorded interviews made available by researchers associated with the Projet Phonologie du Français Contemporain. The complete corpus was forced aligned at the word and phoneme level using the set of speaker adapted acoustic models described in Chapter 2. The large volume of data afforded by this approach (more than $28,000$ observations of word-final consonant clusters) allowed the investigation to determine how variables, such as speech style, following context, motivation for simplification and speech rate, effect the variable pronunciations of word-final consonant clusters.

The corpus analysis proceeded by first examining the individual effects of each of the variables of interest on the responses of either schwa insertion or simplification following a word-final consonant cluster. Next, the combined effects of all of the predictors were examined on the individual responses of schwa insertion or simplification. Third, a smaller sub-sample that controlled for the effect of speech rate was drawn from the corpus. This sample was used to determine whether the frequency of simplification of word-final consonant clusters was a good predictor of schwa insertion. Finally, the corpus was divided into read versus non-read portions and differences between the two corpora were explored. For each of these examinations, separate results were reported for each of the two dialects.

When exploring the individual effects of each of the explanatory variables on the responses of schwa insertion or simplification following a word-final consonant cluster, it was determined that, in all cases, consonant cluster simplification was more frequently observed in the dialect from Québec while schwa insertion following a word-final consonant cluster was more frequently observed in the dialect from France.

The results of forced alignment suggested that word-final consonant cluster simplification may occur in all three contexts (preconsonantal, prevocalic and prepausal) in both dialects and the effect of following context was similar in both dialects: Observations of word-final consonant cluster simplification were higher than expected when the cluster was preconsonantal, and lower than would be expected when the cluster was either prevocalic or prepausal. Similarly, schwa insertion following a word-final consonant cluster may occur in all three contexts (preconsonantal, prevocalic and prepausal) in both dialects, although the effect of following context was much larger in the dialect from France than in the dialect from Québec. While in both dialects when a word-final consonant cluster occurred in a prevocalic context the observed frequencies of schwa insertion were lower than would be expected, only in the dialect from France were there significantly more observations of schwa insertion when the cluster was preconsonantal. This, combined with the larger overall effect of following context in the data from

France, suggested that the dialect from France may be more sensitive to following context, especially preconsonantal, than the dialect from Québec.

The variable pronunciations of word-final consonant clusters as recognized by the forced aligner indicated that word-final consonant cluster simplification may occur in SSP violating and stop-final clusters in both dialects, although the effect of motivation for simplification, while similar in both dialects, was much stronger in the data from France: Observations of word-final consonant cluster simplification were higher than expected for SSP violating clusters, and lower than would be expected for stop-final clusters. With respect to schwa insertion following a word-final consonant cluster, the large difference in the size of the effect for each dialect indicated that the effect of motivation for simplification on schwa insertion was much stronger in the dialect from Québec than in the dialect from France. Additionally, the direction of the effect is opposite: There were significantly fewer observations of schwa insertion following a stop-final consonant cluster in the dialect from Québec, while there were significantly more observations of schwa insertion in these consonant clusters in the dialect from France.

The potentially confounded effects of speech style and speech rate were a challenge to interpret. On the one hand, regardless of whether speech style was defined by reading versus spontaneous speech or by corpus, the effect of style was similar in both dialects: It appeared to be the case that schwa insertion was more frequently observed in the reading speech style, while simplification was more frequently observed in the spontaneous speech style. It also appeared to be the case that schwa insertion was more frequently observed in the AssNat corpus (a more formal speech style), while simplification was more frequently observed in the PFC corpus (a less formal speech style). However, rates of speech were faster in the spontaneous speech style and slower in the reading speech style and speech rate was positively associated with a response of simplification and negatively associated with a reponse of schwa insertion. An analysis of covariance was used to determine whether there were differences between the styles in terms of the responses of either simplification or schwa insertion independent of any differences in rates of speech between the styles that may exist. The interpretation of the models that were presented was that the effect of speech style was best interpreted as the different sensitivity to speech rate that existed between the two styles: While little or no effect of speech rate was observed in the reading speech style for either simplification or schwa insertion, faster speech rates were positively associated with simplification and negatively associated with schwa insertion in the spontaneous speech style. This was true of both dialects and of both corpora.

In order to gain a better understanding of how the variables of speech style, following context, motivation for simplification and speech rate affect the variable pronunciations of word-final consonant clusters in French, several logistic regression models were constructed that evaluated the combined effects of these variables on the responses of either schwa insertion or simplification following a word-final consonant cluster. The models that emerged from the data strengthened most of the intuitions that arose from the exploration of the individual effects of the variables of interest. The models further suggested some interesting dialectal differences in the relative strengths of the effects of some of the predictors, specifically the marked dialectal difference in the effects of following context.

In the first place, the understanding of the effect of speech style remained best interpreted as the different sensitivity to speech rate that existed between the two styles. After taking into account the variance accounted for by the other predictors, while little or no effect of speech rate was observed in the reading speech style for either simplification or schwa insertion, faster speech rates were positively associated with simplification and negatively associated with schwa insertion in the spontaneous speech style. This was true of both dialects.

Using the force aligned data, the logistic regression models predicting an outcome of word-final consonant cluster simplification determined that, after taking account of the effects of speech style, following context and speech rate, the odds of observing word-final consonant cluster simplification in SSP violating consonant clusters was signficantly higher than in stop-final consonant clusters in the dialect from France, while this difference did not exist in the dialect from Québec: Both SSP violating and stop-final consonant clusters were equally likely to participate in simplification for that dialect. The models predicting a reponse of schwa insertion following a word-final consonant cluster indicated that after taking account of the effects of speech style, following context and speech rate, while schwa insertion in the dialect from France was slightly more likely following SSP violating and stop-final ('Simplify') consonant clusters than when following /ʁ/-intial ('Stable') consonant clusters, schwa insertion following 'Simplify' consonant clusters in the dialect from Québec was more than 3 times more likely than when following 'Stable' consonant clusters.

The effect of following context showed the most interesting dialectal differences. After taking into account the combined effects of the other variables of interest, it was determined that the effect of following context on the response of simplification was approximately the same in both dialects. In the data from both Québec and France, the odds of observing a response of simplification following a word-final consonant cluster were significantly lower when the consonant cluster was in a prevocalic context than when in a preconsonantal context and the coefficient expressing the change in these odds was approximately the same in both dialects. While the overall frequency of word-final consonant cluster simplification remains higher in the dialect from Québec, the decrease in the probability of observing word-final consonant cluster simplification in a prevocalic context (as compared to a preconsonantal context) is approximately the same in both dialects. That is, the effect of following context on word-final consonant cluster simplification was the same in both dialects.

The effect of following context on the response of schwa insertion, however, was very different in the data from Québec than in the data from France. After taking into account the combined effects of the other predictors, the coefficient expressing the change in probability of observing a response of schwa insertion in the model for the dialect from Québec indicated that schwa insertion was only slightly less likely when the consonant cluster was prevocalic than when it was preconsonantal. The same coefficient in the model for the dialect from France indicated that schwa insertion was significantly less likely when the consonant cluster was prevocalic than when it was preconsonantal. In fact, the odds ratio of observing the response of schwa insertion as opposed to no schwa insertion indicated that in the dialect from Québec, it was just 1.36 times more likely that schwa insertion will occur in a preconsonantal context than in a prevocalic context, while in the dialect from France it was 4.32 times more likely that schwa insertion will occur in a preconsonantal context than in a prevocalic context.

Putting these observations together, the data provided by the results of forced alignment suggested that the effect of following context on a response of simplification was approximately the same in both dialects, while the effect of following context on a response of schwa insertion was different in each dialect. In both dialects, word-final consonant cluster simplification was significantly more likely when the consonant cluster was preconsonantal. In the dialect from France, schwa insertion following a word-final consonant cluster was also significantly more likely when the consonant cluster was preconsonantal. Further, the strength of the effect of following context on both of the responses of simplification and schwa insertion, as expressed by the coefficients in both models, was approximately the same in that dialect. Compare this with the dialect from Québec where schwa insertion following a word-final consonant cluster was only slightly more likely when the consonant cluster was preconsonantal. Further, the strengths of the effect of following context on both of the responses of simplification and schwa insertion in that dialect were different. Put simply, following context was a good predictor of both schwa insertion and simplification following a word-final consonant cluster in the dialect from France. Following context was a good predictor only for word-final consonant cluster simplification in the dialect from Québec.

An attempt was made to control for the effects of speech rate while still allowing speech style to vary by drawing a sample from the corpus in which no effect of speech rate was present. Additionally, it was asked whether a consonant cluster's frequency of simplification was also a good predictor of schwa insertion. The results indicated this was not the case for the dialect from France, but the hypothesis was supported in the dialect from Québec: After controlling for rate of speech, the probability of observing schwa insertion was positively associated with the frequency of simplification in this dialect. In the dialect from Québec, knowing something about a consonant cluster's frequency of simplification helped to know whether or not schwa insertion may also occur following that consonant cluster. In the dialect from France, knowing something instead about the following context helped to know whether or not schwa insertion will occur.

4.3.1 Implications

These results have implications both for a discussion of the dialectal differences in the phonologies required to predict and explain the variable pronunciations of word-final consonant clusters in French, as well as for future investigations seeking to examine either schwa insertion or simplification following word-final consonant clusters.

The results presented in this project may have implications for other, and future, research into either schwa insertion or word-final consonant cluster simplification. The simplest place to begin is with a word of caution about controlling for the effects of speech rate in any corpus of experimentally or naturally collected speech data. It was consistently demonstrated in all the analyses presented in this investigation that speech rate was positively associated with word-final consonant cluster simplification and negatively associated with schwa insertion following a word-final consonant cluster, especially (possibly exclusively) for spontaneous, connected speech. In both dialects, the faster the sample of speech, the more likely it was that simplification would be observed and the less likely it was that schwa insertion would be observed. If the data collected and analyzed during the course of this investigation are representative, then any analysis that seeks to use a corpus of spontaneous, fluent recorded speech should attempt to take account of differences attributed to speech rate lest the analysis be rendered unreliable.

The next comment concerns the use of multiple predictors when investigating variable pronunciations. It was suggested in more than one of the sources reviewed for this project that an accurate description of the variable pronunciations of word-final consonant clusters in French will most certainly be required to appeal to many, possibly interacting or correlating, explanatory variables. Furthermore, it has also been suggested that the presence of additional factors may produce a cumulative effect on the probability of either simplification or schwa insertion following a word-final consonant cluster. This exact notion was not explicitly tested here, but a good example of the efficacy of a multivariate analysis was provided in the example of the disappearing effect of following context on the response of schwa insertion in the dialect from Québec.[1] When examined on its own, following context appeared to be a significant predictor of schwa insertion in that dialect. When examined along with several other predictors, the effect of following context was diminished to the point of disappearing entirely in the model that controlled for

[1] See also Racine and Grosjean (2002) who found support for the use of multiple predictors when investigating schwa vowels in the initial syllable of a word.

speech rates. With the adoption of accurate forced alignment to obtain larger samples of data it will be interesting to see whether some variables that have individually been demonstrated to have explanatory value with respect to the variable pronunciations of word-final consonant clusters retain their predictive qualities when examined in combination.

The implications the results presented in this investigation have on a discussion of the dialectal differences in the phonologies required to predict and explain the variable pronunciations of word-final consonant clusters in French are several. In the first place, it should be noted that none of the results presented here contradicted or refuted previous findings available in the body of research investigating the variable pronunciations of word-final consonant clusters in French.

For example, the long held fact that word-final consonant cluster simplification is more frequently observed in the dialect from Québec, while schwa insertion following a word-final consonant cluster is more frequently observed in the dialect from France, was evident throughout these data. As well, the prediction that not all consonant clusters would be eligible for simplification was also borne out. Further, in line with much previous research it was also the case that, at least for the dialect from France, all consonant clusters would be eligible for schwa insertion. Other commonly observed patterns of the variable pronunciations of word-final consonant clusters in French that were reazlied in these data include higher frequencies of word-final consonant cluster simplification following consonant clusters that violate the Sonority Sequencing Principle in the dialect from France; a preconsonantal context being a prime and usual candidate for both schwa insertion and simplification of the consonant cluster; more careful speech being associated with higher probabilities of schwa insertion and spontaneous speech being associated with higher probabilities of word-final consonant cluster simplification.

Given that the data presented in this investigation share many of the main charateristics of much previously published research into the variable pronunciations of word-final consonant clusters, the use of the results of forced alignment on this relatively large corpus of recorded French speech to gain access to these pronunciations should be justified. As well, these similarities should provide an additional measure of confidence, beyond the investigation of the accuracy of the system of forced alignment, that the overall reliability of automatic labelling using forced alignment can be considered to be close to the one acheived by human annotators. If both of these statements are accepted, then a realistic expectation that the new information discovered during the course of this investigation may also be taken as reliable should be warranted. The new information as suggested by the results of the analyses presented here that may help to inform a discussion of the phonologies of these two dialects are as follows.

With respect to word-final consonant cluster simplification, the division of the consonant clusters into a group of consonant clusters that would regularly participate in simplification and a group of consonant clusters that would infrequently or rarely participate in simplification was shown to be the same in both dialects. In both the dialect from Québec and the dialect from France, word-final consonant clusters whose final consonant was more sonorous than the preceding one, thus violating the Sonority Sequency Principle which states that sonority should fall from the nucelus of the syllable to its edges, as well as consonant clusters containing a final stop consonant, thus incurring a violation of perceptual saliency, regularly participated in simplification. This division into 'Simplify' and 'Stable' type consonant clusters was not unexpected for the dialect from Québec (given the proposal in Côté (2004a)), but judging from the lack of attention paid to these consonant clusters in the literature on simplification pertaining to the dialect from France, it was unexpected for that dialect. While it was shown that SSP violating consonant clusters (mainly /ʁ/ and /l/ final consonant clusters) did simplify most frequently, word-final consonant cluster simplification following a stop-final consonant cluster in the dialect from France was not so uncommon as to be dismissed out of hand.

Related to the first point, these data determined that, in the dialect from Québec, all 'Simplify' consonant clusters (SSP violating and stop-final) shared an equal chance of being simplified. This was different for the dialect from France where word-final consonant cluster simplification was significantly more likely to occur following SSP violating consonant clusters than stop-final consonant clusters. This suggests that, in this dialect, it may either be the case that violations of perceptual saliency are less strongly marked or that salient information is preserved, possibly in elements of the release of the final stop consonant (Côté, 2000; Ginsberg, 1968, Ch. 3), that can be interpreted phonologically in order to help prevent deletion of the consonant.

Additionally, the effect of following context on word-final consonant cluster simplification, specifically a preconsonantal context, was similar in both dialects. Also similar in both dialects were the strength and direction of the effects of both speech style and speech rate on word-final consonant cluster simplification. These observations had the effect of producing very similar patterns of word-final consonant cluster simplification in both dialects. While the overall frequency of word-final consonant cluster simplification was higher for the dialect from Québec, there may be less of a dialectal difference between the phonologies that predict and explain word-final consonant cluster simplification than previously suspected.

With respect to schwa insertion, the data produced by forced alignment suggested that real dialectal differences exist between the phonologies that predict and explain schwa insertion following a word-final consonant cluster. The most important of these differences might be the lack of an effect of following context in the dialect from Québec compared with a strong effect of following context in the dialect from France; the observation that schwa insertion was restricted to just those consonant clusters that also simplifed in the dialect from Québec compared to unrestricted schwa insertion for all consonant clusters in the dialect from France; and the conclusion that, only for the dialect from Québec, the more frequently a consonant cluster simplified, the more likely it was that schwa insertion may occur. These findings indicate that schwa insertion may be motivated by a general avoidance of sequences of consonants for the dialect from France, while in the dialect from Québec it might instead by motivated by a general avoidance of simplification.

The final implication to be drawn from the results concerns the confusing effect of speech style, particularly the lack of an effect of speech rate only during read speech. It was assumed that identifying samples of speech that occurred while a speaker or participant was reading from a text as opposed to speaking spontaneously, either without the aid of notes or during a casual conversation, would be a good indicator of two different styles of speech. It was hoped that this would help to understand whether the variable pronunciations of word-final consonant clusters were associated with, or effected by, changes in speech style. However, the confounded effects of speech rate and speech style made this determination difficult. In fact, the results presented in this investigation may indicate that the act of reading might introduce more than just a change in speech style. Perhaps some aspects of reading influence the variable pronunciations of word-final consonant clusters in ways that were not investigated or controlled for in this study.

4.4 Limitations and directions for future research

There are several limitations that can be described pertaining to the current study. These limitations introduce both a degree of caution in accepting the concusions that have been reached as well as opportunities for future research to verify and expand upon the findings presented in this paper. The main limitation must surely be that not enough variables were explored. This investigation examined the main and combined effects of speech style, following context, motivation for simplification and speech rate on the variable pronunciations of word-final consonant clusters in two dialects of French. While significant and interesting dialectal differences were discovered, the findings could be further strengthened and extended by addressing the following shortcomings.

In the first place, of the 87 possible word-final consonant clusters that occur in the French language, only 29 were included in this study. Because of the uncontrolled nature of the data collection and the need for a sufficient number of examples of a specific consonant cluster, nearly half of the 55 attested word-final consonant clusters contained in the corpus could not be included. This had the unfortunate effect of reducing the scope of the investigation to approximately 1/3 of the 87 possible word-final consonant clusters. Even though 12 of the 16 types of phonological consonant clusters were represented, it would be interesting to see whether the remaining unexamined consonant clusters follow the pronunciation patterns described.

A second consequence of not enough variables was the inability to examine in more detail the possible inter- or intra-speaker variation in the variable pronunciations of word-final consonant clusters. This corpus contained examples of recorded speech from 351 different speakers. While it would have been exciting to have investigated individual variation in such a large group, the very fact that so many speakers were represented meant that not enough data (or not enough comparable data) from each speaker was collected. While the recorded conversations made available through the Projet Phonologie du Français Contemporain provided approximately 30 $\sim$ 45 minutes of spontaneous speech per participant, the comparable amount of read speech provided was closer to just 2 minutes per participant. In the national assembly corpus, only a very few number of speakers spoke for more than a handful of turns. It could be valuable to obtain or construct a corpus that provides equal and adequate amounts of reading and spontaneous speech per participant in order to perform forced alignment.

While the variables that were selected for inclusion in the analyses were all expected to have some explanatory value in predicting the variable pronunciations of word-final consonant clusters and were easy to control for in these data, many more candidate predictors have been proposed in the literature on both schwa insertion and simplification following a word-final consonant cluster. Lexical effects, word frequency effects, word length and syntactic or prosodic effects have all been previously identified as have possible explanatory value along with the predictors examined here. The uncontrolled nature of the natural language speech captured in this corpus made controlling for these additional factors impossible. Since even the few variables selected for this study demonstrated the advantage gained by combining multiple predictors when exploring the variable pronunciations of word-final consonant clusters, it seems likely that the inclusion of further variables of interest can only further advance the understanding of these effects.

All of these limitations arise from a single source: The corpus, while larger than some and large enough for the purposes of the current study, could have been larger still. The national assemblies of both Québec and France (and other nationalities as well) provide publically available digital archives of all of the official proceedings, both debates and committees. These archives have been digitized going back many years and they are accompanied by transcriptions as well as in some cases, video. This allows for the possibility of obtaining a larger volume of data, as well as a larger volume of per participant data, with which to verify and expand upon the findings presented in this paper. Through the use of an accurate system of forced alignment, like the one created for and described in this paper, the list of research questions and hypotheses to be explored is very long indeed.

Appendix A

	p	b	t	d	k	g	f	v	s	z	ʃ	ʒ	ʧ	ʤ	m	n	ŋ	ɲ	l	ʁ	j	w	ɥ	i	y	e	ø	ɛ	ɛ̃	œ	œ̃	ə	a	ɑ	ɑ̃	u	o	ɔ	ɔ̃
Consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
Sonorant	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Syllabic	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Nasal	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	−	−	−	−	−	−	−	−	−	−	+	−	+	−	−	−	+	−	−	−	+
Continuant	−	−	−	−	−	−	+	+	+	+	+	+	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Lateral	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DelayedRelease	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	−	−	−	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Voice	−	+	−	+	−	+	−	+	−	+	−	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CG	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SG	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Labial	+	+	0	0	0	0	+	+	0	0	0	0	0	0	+	0	0	0	0	0	0	+	+	0	+	0	+	0	0	+	+	0	0	0	0	+	+	+	+
Round	−	−	0	0	0	0	−	−	0	0	0	0	0	0	−	0	0	0	0	0	0	+	+	0	+	0	+	0	0	+	+	0	0	0	0	+	+	+	+
Coronal	0	0	+	+	0	0	0	0	+	+	+	+	+	+	0	+	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Anterior	0	0	+	+	0	0	0	0	+	+	−	−	−	−	0	+	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Strident	0	0	−	−	0	0	0	0	+	+	+	+	+	+	0	−	0	0	−	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dorsal	0	0	0	0	+	+	0	0	0	0	0	0	0	0	0	−	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
High	0	0	0	0	+	+	0	0	0	0	0	0	0	0	0	−	+	+	0	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−
Low	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	+	+	+	−	−	−	−
Back	0	0	0	0	+	+	0	0	0	0	0	0	0	0	0	−	+	−	+	+	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+		
Tense	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	+	+	+	+	+	−	+	+	+	+	+	−	−
Reduced	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−	−	−	−	−	−	−	−	+	−	−	−	−	−	−	−

Table A.1: Phone by feature matrix

References

Adda-Decker, M. (2007). Problèmes posés par le schwa en reconnaissance et en alignement automatiques de la parole (problems caused by schwa in automatic speech recognition and alignment). In *Proceedings of Journées d'Études Linguistiques-2007*, pages 211–216, Nantes, France.

Adda-Decker, M., de Mareüil, P. B., Adda, G., and Lamel, L. (2002). Investigating syllabic structure and its variation in speech from french radio inteviews. In *Proceedings of ISCA Tutorial and Research Workshop on Pronunciation Modeling and Lexicon Adaptation*, Aspen, CO.

Adda-Decker, M., de Mareüil, P. B., and Lamel, L. (1999). Pronunciation variants in french: schwa and liaison. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, pages 2239–2242, San Francisco.

Adda-Decker, M., Gendrot, C., and Nguyen, N. (2008). Contributions du traitement automatique de la parole à l'étude des voyelles orales du français (contributions of automatic speech processing to the study of french oral vowels). *Traitement Automatique Langues*, 49:13–46.

Adda-Decker, M. and Lamel, L. (1999). Pronunciation variants across system configuration, language and speaking style. *Speech Communication*, 29(2-4):83–98.

Adda-Decker, M. and Snoeren, N. D. (2011). Quantifying temporal speech reduction in french using forced speech alignment. *Journal of Phonetics*, 39:261–270.

ANF (2011). Assemblée nationale de france. `http://www.assemblee-nationale.fr/index.asp`. Website of Assemblée nationale de France.

ANQ (2011). Assemblée nationale du québec. `http://www.assnat.qc.ca/fr/index.html`. Website of Assemblée nationale du Québec.

Armstrong, N. (1998). The sociolinguistic gender pattern in french: a comparison of two linguistic levels. *Journal of French Language Studies*, 8:139–58.

Armstrong, N. (2001). *Social and stylistic variation in spoken French: a comparative approach.* John Benjamins, Amsterdam.

Ayres-Bennett, W. and Carruthers, J. (2001). Schwa: a maverick vowel. In *Problems and perspectives: studies in the modern French language*, chapter 3. Harlow: Pearson Ed.

Barnes, J. and Kavitskaya, D. (2002). Phonetic analogy and schwa deletion in french. *Berkeley Linguistic Society*, 28:39–50.

Boughton, Z. (2008). Variable deletion of /l/ and /r/ in word-final clusters: The interplay of social and linguistic factors. In *AFLS 2008: Les voix du français: Usages et représentations.*

Bürki, A., Alario, F. X., and Frauenfelder, U. H. (2011a). Lexical representation of phonological variants: Evidence from pseudohomophone effects in different regiolects. *Journal of Memory and Language*, 64(4):424–442.

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., and Frauenfelder, U. H. (2011b). What affects the presence versus absence of schwa and its duration: A corpus analysis of french connected speech. *J. Acoust. Soc. Am.*, 130(6).

Bürki, A., Fougeron, C., Gendrot, C., and Frauenfelder, U. H. (2011c). Phonetic reduction versus phonological deletion of french schwa: Some methodological issues. *Journal of Phonetics*, 39:279–288.

Bürki, A., Gendrot, C., Gravier, G., Linares, G., and Fougeron, C. (2008). Alignement automatique et analyse phonétique: Comparaison de différents systèmes pour l'analyse du schwa (automatic speech alignment and phonetic analysis: A comparison of different systems for the analysis of schwa). *Trait. Automatique Langues*, 49:165–197.

Campbell, W. N. (1988). Extracting speech-rate values from a real-speech database. In *In Proc. ICASSP-88*.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

Côté, M.-H. (2000). *Consonant cluster phonotactics: a perceptual approach*. PhD thesis, MIT. Ch 2.

Côté, M.-H. (2004a). Consonant cluster simplification in québec french. *PROBUS*, 16:151–201.

Côté, M.-H. (2004b). Syntagmatic distinctness in consonant deletion. *Phonology*, 21:1–41.

Côté, M.-H. (2007). Rhythmic constraints on the distribution of schwa in french. In Camacho, J., FloresFerrn, N., Snchez, L., Dprez, V., and Cabrera, M., editors, *Romance Linguistics 2006*. John Benjamins, Amsterdam.

Côté, M. H. (2008). Empty elements in schwa, liaison and h-aspiré: The french holy trinity revisited. In Hartmann, J., Hegedüs, V., and van Riemsdijk, H., editors, *Sounds of Silence: Empty Elements in Syntax and Phonology*, pages 61–103. Elsevier, Amsterdam.

Côté, M.-H. (2009). Contraintes segmentales et variation dans la perte et la stabilisation du schwa en syllabe initial (segmental constraints and variation in the loss and stabilization of schwa in initial syllables). In Baronian, L. and Martineau, F., editors, *Le Français dun Continent à lAutre*, pages 93–121. Presses de lUniversite Laval, Québec, Canada.

Côté, M.-H. (2012). Laurentian french (québec): Extra vowels, missing schwas and surprising liaison consonants. In Gess, R., Lyche, C., and Meisenburg, T., editors, *Phonological variation in French: illustrations from three continents*, pages 235–274. John Benjamins, Amsterdam.

Coveney, A. (2001). *The Sounds of contemporary French*. Elm Bank Publications, Exeter.

Cucchiarini, C. and Binnenpoorte, D. (2002). Validation and improvement of automatic phonetic transcriptions. In *Proceedings of ISCLP*, pages 313–316, Denver, USA.

Dalton, W. (2011). An ultrasound imaging study of the tense-lax distinction in canadian french vowels. *Canadian Acoustics*, 39(3).

de Cornulier, B. (1978). Syllabe et suite de phonèmes en phonologie du français. In de Cornulier, B. and cois Dell, F., editors, *Études de phonologie française*, pages 31–70. CNRS, Paris.

de Mareüil, P. B. (2007). Traitement du schwa: De la synthèse à l'alignement automatique (the processing of schwa: From text-to-speech synthesis to automatic alignment). In *Proceedings of Journées d'Études Linguistiques-2007*, pages 181–189, Nantes, France.

de Mareüil, P. B., Habert, B., Bénard, F., Adda-Decker, M., Barras, C., Adda, G., and Paroubek, P. (2005). A quantitative study of disfluencies in french broadcast interviews. In *Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop*, pages 27–32.

Dell, F. (1985). *Les règles et les sons. Introduction á la phonologie générative*. Hermann, Paris.

Dell, F. (1995). Consonant clusters and phonological syllables in french. *Lingua*, 95:5–26.

DiCanio, C., Nam, H., Whalen, D. H., Bunnell, H. T., Amith, J. D., and García, R. C. (2013). Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *Journal of the Acoustical Society of America*, 134(3):2235–2246.

Dumas, D. (1986). Le statut des "deux a" en français québécois. *Revue québécoise de linguistique*, 15(2):167–197.

Durand, J. and Eychenne, J. (2004). Le schwa en français: pourquoi des corpus? *Corpus 3*, pages 311–356.

Durand, J., Laks, B., and Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. In Pusch, C. and Raible, W., editors, *Romanistische Korpuslingquistik - Korpora und gesprochene Sprache/Romance Corpus Linguistics Corpora and Spoken Language*, pages 93–106. Gunter Narr, Tubingen.

Durand, J., Laks, B., and Lyche, C. (2005). Un corpus numérisé pour la phonologie du français. In Williams, G., editor, *La linguistique de corpus*, pages 205–217. Presses Universitaires de Rennes, Rennes.

Durand, J., Laks, B., and Lyche, C. (2009). Le projet pfc: une source de données primaires structurées. In Durand, J., Laks, B., and Lyche, C., editors, *Phonologie, variation et accents du français*, pages 19–61. Hermès, Paris.

Elman, J. and McClelland, J. M. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, 27:143–165.

Evanini, K. (2009). *The permeability of dialect boundaries: A case study of the region surrounding Erie, Pennsylvania*. PhD thesis, University of Pennsylvania.

Eychenne, J. (2003). La phonologie du français au québec et le statut du e-caduc. In Durand, J., Laks, B., and Lyche, C., editors, *La prononciation du français*, volume 33, pages 72–77. Tribune internationale des langues vivantes.

Eychenne, J. (2005). The emergence of prosodic schwa in french: a typological approach. In *Bilbao-Deusto Student Conference in Linguistics*, Bilbao.

Eychenne, J. (2006). *Aspects de la phonologie du schwa dans le français contemporain. Optimalité, visibilité prosodique, gradience (Some aspects of French schwa phonology in contemporary French: Optimality, prosodic visibility, and gradience)*. PhD thesis, Université de Toulouse-Le Mirail, France.

Fouché, P. (1958). *Phonétique historique du français. vol II: Les voyelles.* C. Klincksieck, Paris.

Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.-F., and Gravier, G. (2005). Ester phase ii evaluation campaign for the rich transcription of french broadcast news. In *Proceedings of Interspeech 2005*, pages 1149–1152, Lisboa, Portugal.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6:110–125.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (1990). *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM.* National Institute of Standards and Technology. NTIS Order No. PB91-505065.

Gauvain, J.-L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., and Lamel, L. (2005). Where are we in transcribing french broadcast news? In *Proceedings of Interspeech*, Lisbon.

Gauvain, J.-L., Lamel, L. F., Adda, G., and Adda-Decker, M. (1994). Speaker-independent continuous speech dictation. *Speech Communication*, 15:21–37.

Ginsberg, R. (1968). La détente consonantique en français canadien et en français standard. In Léon, P. R., editor, *Recherches sur la structure phonique du français canadien*, pages 131–144. Didier, Montréal.

Goad, H. (2002). Markedness in rightedge syllabification: parallels across populations. *Canadian Journal of Linguistics*, 47:151–186.

Godfrey, J., Holliman, E., and McDaniel, J. (1992). Switchboard: telephone speech corpus for research and development. In *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, page 517520, San Francisco, CA.

Goldman, J.-P. (2011). Easyalign: an automatic phonetic alignment tool under praat. In *Interspeech*, Firenza, Italy.

Grammont, M. (1914). *Traité pratique de prononciation française (Practical treatise on French pronunciation)*. Delagrave, Paris.

Grammont, M. (1961). *La pronunciation française: traité pratique*. Delagrave, Paris.

Greenberg, S. (1998). Speaking in shorthand a syllable centric perspective for understanding pronunciation variation. In *Proceedings of the ESCA Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition*, Kekrade, Netherlands.

Grevisse, M. and Goosse, A. (2011). *Le Bon usage*. De Boeck Duculot, Louvain-la-Neuve.

Hansen, A. B. (1994). étude du e caduc - stabilisation en cours et variations lexicales (the study of caduc-e, stabilization in progress and lexical variations). *Journal of French Language Studies*, 4:25–54.

Hosom, J.-P. (2009). Speaker-independent phoneme alignment using transition-dependent states. *Speech Communication*, 51(4):352–368.

Johnson, K. (2004). Massive reduction in conversational american english. In Yoneyama, K. and Maekawa, K., editors, *Proceedings of the 10th International Symposium: Spontaneous Speech: Data and Analysis*, pages 29–54, Tokyo, Japan.

Kemp, W., Pupier, P., and Yaeger, M. (1980). A linguistic and social description of final consonant cluster simplification in montréal french. In Shuy, R. and Shnukal, A., editors, *Languages and the uses of language*, pages 12–40. Georgetown University Press, Washington.

Kipp, A., Wesenick, M., and Schiel, F. (1996). Automatic detection and segmentation of pronunciation variants in german speech corpora. In *Proceedings of ICSLP*, pages 106–109.

Kipp, A., Wesenick, M., and Schiel, F. (1997). Pronunciation modeling applied to automatic segmentation of spontaneous speech. In *Proceedings of eurospeech*, pages 1023–1026.

Kisseberth, C. (1970). On the functional unity of phonological rules. *Linguistic Inquiry*, pages 291–306.

Kuipers, C. and van Donselaar, W. (1997). The influence of rhythmic context on schwa epenthesis and schwa deletion in dutch. *Language and Speech*, 41:87–108.

Labov, W., Rosenfelder, I., and Fruehwald, J. (2013). One hundred years of sound change in philadelphia: Linear incrementation, reversal, and reanalysis. *Language*, 89(1):30–65.

Laks, B. (1977). Contribution empirique a l'analyse socio-différentielle de la chute de /r/ dans les groupes consonantiques finals. *Langue française*, 34:109–125.

Lamel, L., Gauvain, J.-L., and Eskénazi, M. (1991). Bref, a large vocabulary spoken corpus for french. In *Proceedings of the European Conference on SpeechTechnology, EuroSpeech*, pages 505–508, Genoa.

Lamel, L., Rosset, S., Bennacef, S., Bonneau-Maynard, H., Devillers, L., and Gauvain, J.-L. (1995). Development of spoken language corpora for travel information. In *EUROSPEECH*, Madrid.

Lanchantin, P., Morris, A., and Rodet, X.and Veaux, C. (2008). Automatic phoneme segmentation with relaxed textual constraints. In *Proceedings of the International Conference on Language Resources and Evaluation*, pages 2403–2407, Marrakech, Morocco.

Landis, J. and Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Lennes, M. (2011). Spect - the speech corpus toolkit for praat. `www.helsinki.fi/~lennes/praat-scripts/`.

Lyche, C. and Durand, J. (1996). Testing government phonology ou pourquoi le choix du schwa. In Durand, J. and Laks, B., editors, *Current Trends in Phonology. Models and Methods*, pages 443–471. European Studies research Institute.

Malécot, A. (1976). The effect of linguistic and paralinguistic variables on the elision of the french mute-e. *Phonetica*, 33:93–112.

Martin, P. (1998). à québec, a-t-on l'schwa? In Duhoux, Y., editor, *Langue et langues: hommage à Albert Maniet*, pages 163–180. Peeters, Leuven.

Milne, P. (2011a). The effects of syllable position on allophonic variation in québec french /ʀ/: A corpus analysis using a modified version of the penn phonetics lab forced aligner. In *NWAV40*, Washington, DC. Georgetown University.

Milne, P. (2011b). Finding schwa: Comparing the results of an automatic aligner with human judgments when identifying schwa in a corpus of spoken french. *Canadian Acoustics*, 39(3):190–192.

Milne, P. and Côté, M.-H. (2009). Les relations france-québec: insertion de schwa vs réduction des groupes consonantiques. In *Réunion 2009*, Paris, France. Phonologie Français Contemporain.

Morin, Y.-C. (1974). Règles phonologiques à domaine indéterminé : chute du cheva en français. *Cahier de linguistique*, 4:69–88.

Nadasdy, T. (2005). Le français en ontario. In Valdman, A., Auger, J., and Piston-Hatlen, D., editors, *Le français en Amérique du Nord: état présent*, pages 99–116. Presses de lUniversité Laval.

Nemoto, R., Adda-Decker, M., and Durand, J. (2010). Word boundaries in french: Evidence from large speech corpora. In *LREC*.

New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (2004). Lexique 2 : A new french lexical database. *Behavior Research Methods, Instruments, and Computers*, 36(3):515–524.

New, B., Pallier, C., Ferrand, L., and Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique*, 101:447–462. http://www.lexique.org.

Noske, R. (1993). *A theory of syllabification and segmental alternation. With studies on the phonology of French, German, Tonkawa and Yawelmani*. PhD thesis, Tilburg University. Ch 6.

Picard, M. (1991). La loi des trois consonnes et la chute du cheva en québecois. *Revue québecoise de linguistique*, 20(2):35–49.

Pooley, T. (1996). *Chtimi: the urban vernacular of northern France*. Multilingual Matters, Clevedon.

Pulgram, E. (1961). [untitled]. *Language*, 37(2):282–283.

Pustka, E. (2011). Le conditionnement lexical de l'élision des liquides en contexte post-consonantique final. *Langue française*, 169(1):19–38.

Racine, I. and Grosjean, F. (2002). La production du e caduc facultatif est-elle prévisible? un début de réponse (is the production of the optional e predictable? the beginnings of a reply). *Journal of French Language Studies*, 12:307–326.

Samuel, A. G. and Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language*, 48:416–434.

Sankoff, G. and Blondeau, H. (2007). Language change across the lifespan: /r/ in montreal french. *Language*, 83(3):560–588.

Saraclar, M. and Khudanpur, S. (2004). Pronunciation change in conversational speech and its implications for automatic speech recognition. *Computer Speech and Language*, 18:375–395.

Schuppler, B., Ernestus, M., Scharenborg, O., and Boves, L. (2011). Acoustic reduction in conversational dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39:96–109.

Schuppler, B., van Dommelen, W. A., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, 40:595–607.

Séguin, M.-C. (2010). *Catégorisation acoustique des voyelles moyennes antérieures arrondies en français laurentien*. PhD thesis, University of Ottawa.

Siegler, M. and Stern, R. (1995). On the effects of speech rate in large vocabulary speech recognition systems. In *In IEEE ICASSP-95*.

Tousignant, C. (1987). Relations sociales entre le /r/ vélaire et le /r/ rétroflexe anglais de montréal. *Revue québécoise de linguistique théorique et appliquée*, 6(2):39–55.

Tranel, B. (1987). *The sounds of French: An introduction*. Cambridge University Press, Cambridge.

Weintraub, M., Taussig, K., Hunicke-smith, K., and Snodgrass, A. (1996). Effect of speaking style on lvcsr performance. In *In ICSLP*, pages 16–19.

Yuan, J. and Liberman, M. (2008). Speaker identification on the scotus corpus. In *Proceedings of Acoustics '08*.

Zwicky, A. (1972). Note one a phonological hierarchy in english. In Stockwell, R. P. and Macauley, R. K. S., editors, *Linguistic change and generative theory*, pages 275–301. Indiana University Press, Bloomington.