

# Speech Reduction in Spontaneous French and Spanish

© 2011, Francisco Torreira

ISBN/EAN: 978-90-76203-42-3

Cover design: Harald Pieper

Printed and bound by Ipskamp Drukkers, Nijmegen

# Speech Reduction in Spontaneous French and Spanish

een wetenschappelijke proeve op het gebied van de Letteren

## Proefschrift

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,  
volgens besluit van het college van decanen  
in het openbaar te verdedigen op maandag 14 november 2011  
om 13.30 uur precies

door

Francisco José Torreira Martínez

geboren op 29 maart 1978  
te Cádiz, Spanje

Promotor: Prof. dr. L. Boves  
Copromotor: Dr. M. Ernestus

Manuscriptcommissie:

Prof. dr. C. Gussenhoven (Voorzitter)  
Dr. C. Fougeron (Université Sorbonne Nouvelle - Paris 3)  
Prof. dr. J. M. Pardo (Universidad Politécnica de Madrid)

# Contents

<b>Chapter 1: Introduction</b>	<b>1</b>
1.1 Speech reduction . . . . .	1
1.2 The nature of speech reduction phenomena . . . . .	3
1.3 Probabilistic and morphological effects on speech reduction . . . . .	5
1.4 Language-specific reduction . . . . .	6
1.5 Outline of this dissertation . . . . .	6
<b>Chapter 2: The Nijmegen Corpus of Casual French</b>	<b>9</b>
2.1 Introduction . . . . .	10
2.2 Corpus creation . . . . .	11
2.2.1 Participants . . . . .	11
2.2.2 Recording set-up . . . . .	12
2.2.3 Recording procedure . . . . .	12
2.2.4 Orthographic transcription . . . . .	16
2.3 Corpus contents . . . . .	20
2.4 Assessing casualness . . . . .	21
2.4.1 Disfluency words . . . . .	21
2.4.2 Word repetitions . . . . .	22
2.4.3 Lexical items . . . . .	23
2.4.4 Double negation . . . . .	25
2.4.5 Homogeneity across parts . . . . .	26
2.4.6 Homogeneity across speakers . . . . .	27
2.5 Discussion . . . . .	31
<b>Chapter 3: Probabilistic effects on French [t] duration</b>	<b>33</b>
3.1 Introduction . . . . .	34
3.2 Study 1 . . . . .	35
3.2.1 Method . . . . .	35
3.2.2 Results . . . . .	37
3.2.3 Discussion . . . . .	38
3.3 Study 2 . . . . .	39

3.3.1	Method . . . . .	39
3.3.2	Results . . . . .	40
3.4	General discussion and conclusions . . . . .	41
<b>Chapter 4: Vowel elision in casual French: the case of vowel /e/ in the word <i>c'était</i></b>		<b>43</b>
4.1	Introduction . . . . .	44
4.2	Realizations of <i>c'était</i> in connected speech . . . . .	46
4.3	Quantitative study . . . . .	47
4.3.1	Method . . . . .	49
4.3.2	Results: Distribution of acoustic characteristics . . . . .	53
4.3.3	Results: Conditioning factors . . . . .	57
4.4	Discussion . . . . .	60
<b>Chapter 5: Phrase-medial vowel devoicing in spontaneous French</b>		<b>65</b>
5.1	Introduction . . . . .	66
5.1.1	Vowel devoicing . . . . .	66
5.1.2	Vowel devoicing in French . . . . .	66
5.1.3	The present study . . . . .	67
5.2	Method . . . . .	67
5.2.1	Materials . . . . .	67
5.2.2	Measurements and annotation . . . . .	68
5.2.3	Results . . . . .	69
5.3	Discussion . . . . .	74
<b>Chapter 6: Realization of voiceless stops and vowels in conversational French and Spanish</b>		<b>77</b>
6.1	Introduction . . . . .	78
6.1.1	Phonetic reduction in conversational speech and purpose of this research . . . . .	78
6.1.2	Reduction of voiceless stops . . . . .	79
6.1.3	Vowel reduction . . . . .	80
6.2	Method . . . . .	81
6.2.1	Materials . . . . .	81
6.2.2	Measurements . . . . .	84
6.3	Analysis and results . . . . .	87
6.3.1	Analysis . . . . .	87
6.3.2	Voiceless stops . . . . .	88
6.3.3	Vowels . . . . .	93
6.4	Discussion and conclusion . . . . .	97

<b>Chapter 7: Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish</b>	<b>101</b>
7.1 Introduction . . . . .	102
7.1.1 Speech reduction and goal of the present study . . . . .	102
7.1.2 Realizations of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish . . . . .	103
7.1.3 Possible articulatory mechanisms behind /s/-weakening . . . . .	107
7.1.4 Morphological effects on speech reduction . . . . .	108
7.1.5 Probabilistic and lexical effects on speech reduction . . . . .	109
7.1.6 The Nijmegen Corpus of Casual Spanish . . . . .	110
7.1.7 Outline of this article . . . . .	111
7.2 Study 1: Exploratory overview of /s/-weakening . . . . .	111
7.2.1 Method . . . . .	111
7.2.2 Results . . . . .	116
7.3 Study 2: Morphological factors . . . . .	120
7.4 Study 3: Probabilistic and lexical factors . . . . .	122
7.5 General discussion . . . . .	125
<b>Chapter 8: Conclusion</b>	<b>129</b>
8.1 The Nijmegen Corpus of Casual French and the Nijmegen Corpus of Casual Spanish . . . . .	129
8.2 The nature of reduction phenomena . . . . .	130
8.3 Probabilistic, lexical, and morphological factors in speech reduction . . . . .	131
8.4 Language differences in speech reduction . . . . .	133
8.5 Concluding remarks . . . . .	133
<b>Summary</b>	<b>135</b>
<b>Samenvatting</b>	<b>141</b>
<b>Acknowledgments</b>	<b>147</b>
<b>Bibliography</b>	<b>148</b>
<b>Curriculum vitae</b>	<b>163</b>
A Curriculum vitae . . . . .	163
B List of publications . . . . .	163





## 1.1 Speech reduction

In everyday spontaneous conversations, speech sounds, syllables and words are often pronounced less clearly than in more formal speaking styles. In Spanish, for instance, the sounds /p/, /t/ and /k/ are often pronounced as voiced approximant sounds ([β], [ð] and [ɣ]) when produced between two vowels. This is illustrated in Figure 1.1, which shows a waveform and a spectrogram of the Spanish utterance *pequeña capital* /pekeɲa kapital/ ‘small capital’ in an example extracted from the Nijmegen Corpus of Casual Spanish. In this utterance, only the first /p/ and /t/ sounds were produced with a voiceless stop closure, as evidenced by the stretches of silence in the waveform (flat signal) and the lack of acoustic energy in the spectrogram (white space) in the time regions marked with black arrows. All the other stop consonants, marked with grey arrows, exhibit periodic energy extending well into the medium frequency range (1-5 KHz), indicating that they were produced with glottal vibrations and without full oral closures. This utterance was therefore not produced in close correspondence to the canonical forms of its words ([pekeɲa kapital]), but rather with three reduced consonants ([peyeɲa ɣaβital]).

Studies of conversational speech show that reduction phenomena such as the one presented in Figure 1 are much more common than is generally thought (Ernestus, 2000; Kohler, 2000; Bell et al., 2003). For instance, for a large database of American English, Johnson (2004) reports reductions affecting no less than 60% of all word tokens, and absences of sounds in at least 20% of the word tokens. In another large-scale corpus study, Schuppler et al. (2011) document numerous instances of speech reduction in spontaneous Dutch affecting around 40% of the analyzed word tokens; they conclude that reduction phenomena in this language are more pervasive than previously documented. Despite the ubiquity of reduction phenomena in everyday speech, however, little is known about how reduced sounds, syllables and words come to be produced by speakers.

During the last century, formal linguists have tended to disregard reduced pronunciations as mere by-products of ‘linguistic performance’ irrelevant to the enterprise

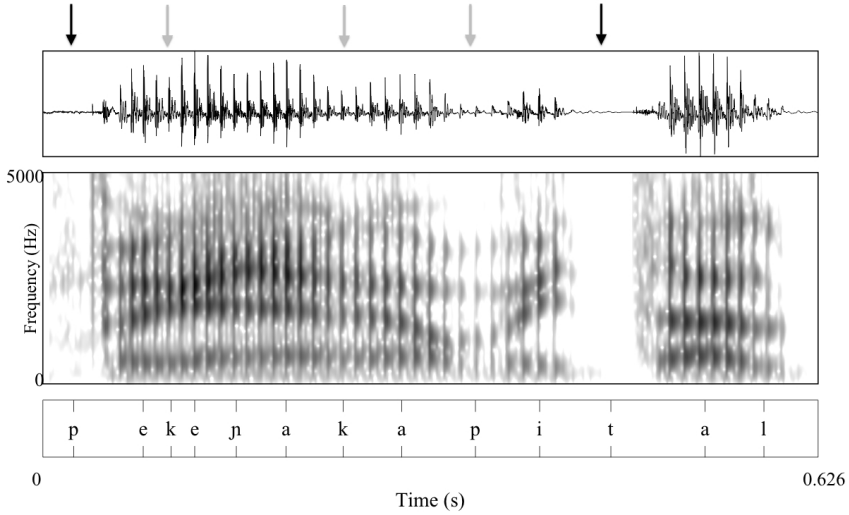


Figure 1.1: Spanish utterance containing reduced (grey arrows) and unreduced (black arrows) stop consonants.

of describing the structural properties of language (Jakobson and Halle, 1968). In order to develop workable theories of linguistic structure, many linguists have been inclined to disregard the roles of factors claimed to be external to language *per se* such as speech rate, memory restrictions or inattention (Chomsky and Halle, 1968). Disciplines devoted to the study of ‘linguistic performance’ such as phonetics and psycholinguistics have also paid little attention to reduction and connected-speech phenomena, leaving an important gap in our understanding of verbal communication. In the case of phonetics, one possible reason for this is that research has focused mainly on discovering and studying invariant units of speech production and perception (e.g. segments, syllables, tones, gestures, etc.), and that the variability present in speech, including that induced by reduction phenomena, has often been regarded as something to avoid rather than study in its own sake.

The methodology used in most experimental studies in speech science may also explain why reduction phenomena have received so little attention. The vast majority of studies in phonetics and psycholinguistics are based on read utterances elicited in the laboratory. Most of the speech materials collected using this methodology contain few and mild cases of reduction if at all, and this may have led researchers to neglect reduction phenomena given their apparent marginality. In the last decade, this situation has led some to question the ecological validity of most phonetic research as it is carried out nowadays, since it tells us little about how everyday verbal

communication takes place (Kohler, 2000; Johnson, 2004).

Another reason why speech reduction has received little attention is that it appears to be largely unnoticeable for listeners during speech comprehension. This is evidenced by experimental results indicating that listeners restore reduced words to their full forms at some stage of the speech comprehension process (Kemps et al., 2004). Although speech scientists should not be treated as naive language users, their linguistic experience may have led them to view reduction processes as more marginal than they really are, and this may have had an influence on their research agendas.

In summary, several factors have led speech reduction to be neglected by researchers in spite of its pervasiveness in everyday verbal communication. As a consequence, much remains to be explored regarding its nature and its conditioning factors. The aim of this dissertation is to contribute to filling this gap by investigating several reduction phenomena in French and Spanish, languages that have attracted little research on speech reduction in comparison with Germanic languages such as English, Dutch or German. It is part of a larger project entitled *Acoustic Reduction in European Languages*, funded by the European Science Foundation in the form of a European Young Investigator Award to Mirjam Ernestus.

As explained above, only spontaneous speech data collected under naturalistic conditions can reveal the pervasiveness of speech reduction phenomena in everyday language. Given the absence of accessible databases of spontaneous French and Spanish, I have created two new corpora, the Nijmegen Corpus of Casual French and the Nijmegen Corpus of Casual Spanish. Chapter 2 presents a detailed description of the collection and contents of the Nijmegen of Corpus of Casual French. Since the Nijmegen Corpus of Casual Spanish was collected following the same procedure as the Nijmegen Corpus of Casual French, information about the Spanish corpus is provided more briefly in Chapter 6 as part of a study on intervocalic /s/-weakening in this language.

## 1.2 The nature of speech reduction phenomena

Reduction processes have traditionally been described in terms of categorical segmental processes such as deletion and assimilation. For example, in American English, the word *support*, which in citation form is expected to be pronounced [səpɔrt], is often pronounced without an audible schwa vowel ([spɔrt], Dalby, 1986). In this example, the vowel [ə] is said to be ‘deleted’. Assimilation occurs when one segment becomes more similar to a neighboring segment. For instance, the sound [s] in the utterance *miss you* is often palatalized under the influence of upcoming [j], resulting in a pronunciation that can be transcribed as [mɪfju].

Segmental transcriptions of reduction phenomena such as the two presented above present a fundamental problem in that they often lead to considerable data reduction. For instance, when the word *support* is pronounced with a ‘deleted’ schwa, and transcribed as [sɔrt], it is still different from the word *sport* in terms of temporal and spectral details. Since this difference cannot be accurately captured, and therefore studied, through the use of segmental letter-like symbols, one is often led to believe either that this difference does not exist, or that it does not matter to such a point that detailed articulatory or acoustic descriptions should be provided by the researcher. However, differences such as this one do matter, because listeners treat them as part of the linguistic message, and also because they reveal systematic patterns of articulation important from the perspective of speech production (Manuel, 1992; Zsiga, 1995; Davidson, 2006).

A way to overcome this problem is to use both gradient and categorical descriptive and analytical tools, and to depart from traditional models of speech by referring to proper articulation and well-defined acoustic events rather than to phoneme-sized units. An example of such an approach is provided by Browman and Goldstein (1990, 1992), who propose that two distinct articulatory processes operating on invariant lexical entries are responsible for the reductive phenomena observed in casual speech. First, there may be a reduction in the magnitude of speech gestures. For example, a stop closure can be reduced to an oral configuration more typical of an approximant ( $[g] > [\gamma]$ ), or even to such an extent that it leaves little or no acoustic traces at all ( $[g] > \emptyset$ ). The second process underlying phonetic reduction is the increase in overlap between neighboring articulatory gestures. In the utterance *miss you* mentioned above, the tongue blade is required to perform different actions by the adjacent sounds [s] and [j]. In such a situation, gradient gestural blending may occur resulting in an articulation difficult to capture with letter-like symbols as discussed above. According to Browman and Goldstein, therefore, phonetic reduction results from changes in the size and timing of articulatory gestures during the production of connected speech. Depending on their magnitude, these gradient processes may result in acoustic patterns traditionally described as coarticulation, assimilation, or even the complete ‘deletion’ of segments and syllables.

Despite the convincing evidence that Browman and Goldstein offer in their reinterpretation of common segmental deletions and assimilations in English, there are reasons to think that categorical processes might also be at work in specific cases. For instance, several assimilation and deletion processes have been found to be categorical rather than gradient upon experimental inspection (e.g. schwa elision in French, Bürki et al. (2007); voicing assimilation in French, Hallé and Adda-Decker (2007); nasal place assimilation in Spanish, Honorof (1999); place assimilation in Korean, Kochetov and Pouplier (2008), among others). Although Browman and Goldstein

may be right in that most connected-speech reduction phenomena traditionally considered as categorical are of a gradient nature, there is nevertheless evidence that categorical reductions also occur.

The issue of whether a specific reduction phenomenon is categorical or continuous is relevant for modeling speech production. This issue is explicitly addressed in Chapters 4 and 5 of this dissertation. On the basis of acoustic measures of spontaneous speech data, Chapter 4 investigates if the elision of vowel /e/ in the French word *c'était* is the product of articulatory reduction, or if vowel /e/ is absent from speakers' articulatory plans in a considerable number of *c'était* pronunciations. In Chapter 5, I investigate the devoicing of high vowels in phrase-medial position in French.

### 1.3 Probabilistic and morphological effects on speech reduction

The reduction mechanisms described above can affect all words in a given language. Interestingly, studies indicate that reduction does not affect all words equally. A number of recent corpus studies have shown that probabilistic factors such as the frequency of occurrence and contextual predictability of specific linguistic units (e.g. words, syllables, segments) may affect articulation (Bell et al., 2003, 2009; Pluymaekers et al., 2005; Aylett and Turk, 2004, among others). Bell et al. (2003) found that the ten most frequent English function words (e.g. *I, of, that*) were more likely to be reduced in contexts of high probability given the preceding and/or the following word. Pluymaekers et al. (2005) found a shortening effect of the frequency of a word on the duration of several affixes and affix-internal segments drawn from a corpus of spontaneous Dutch. Similarly, Bell et al. (2009) report that English content and function words tend to be shorter the higher their lexical frequency and conditional probability (i.e. probability of a word given its preceding or following words). These studies suggest that more frequent and predictable linguistic materials tend to receive shorter and weaker pronunciations, a claim also known as the Probabilistic Reduction Hypothesis (Bell et al., 2003).

The question arises if such probabilistic effects can be also found for Romance languages like French and Spanish, which notably differ from the previously studied Germanic languages in terms of their prosodic and rhythmic properties. Since many of the studies cited above have used global estimates of reduction (i.e. word and syllable duration measured across large corpora), the extent to which probabilistic factors condition specific reduction phenomena also deserves to be studied in more detail. These questions are addressed several times throughout this dissertation. Chapter 2 explicitly investigates if the duration of French [t] consonants is

affected by probabilistic factors. In Chapter 5, which describes phrase-medial vowel devoicing in French, I examine if devoicing tends to occur more often in function words than in content words. Finally, in Chapter 7, which documents the weakening of intervocalic /s/ consonants in Spanish, I investigate if /s/ consonants are affected by grammatical class, lexical frequency and contextual predictability.

Several studies have shown that reduction phenomena can be conditioned by morphological factors. For instance, several researchers have pointed out that the deletion of word-final /t/ and /d/ in English, a common reduction, is more likely to apply when the coronal consonant is part of a monomorphemic word (as in *mist*) than when it is a verbal suffix (as in *missed*) (Guy, 1994; Bybee, 2002; Labov, 2004). A functional explanation of this phenomenon is that segments carrying grammatical information (such as /t/ and /d/ suffixes) are not likely to be hypoarticulated or deleted by speakers. Chapter 7 tests this idea by studying the weakening of intervocalic /s/ in Spanish, a segment which can occur both as part of a stem and as plural marker or verbal suffix. In this chapter I also investigate if /s/ suffixes that are redundant based on their morphosyntactic context are more reduced than non-redundant suffixes.

## 1.4 Language-specific reduction

Although it is reasonable to think that speech reduction results largely from universal production and perceptual constraints, there are also good reasons to think that it must be language-specific to a certain extent. For instance, numerous phonetic studies have found that the strength of many coarticulatory processes differs among languages (e.g. Öhman, 1966; Boyce, 1990; Manuel, 1990; Solé, 1995; Manuel, 1999; Beddor et al., 2002), and it is well known that the outcome of diachronic weakening processes is never completely predictable. These facts lead us to hypothesize that speech reduction is language-specific and learned to a certain degree. More particularly, I hypothesize that similar speech sounds may undergo different degrees of reduction in different languages. Chapter 8 addresses this issue by comparing the realization of vowels and voiceless stops in French and Spanish spontaneous speech.

## 1.5 Outline of this dissertation

This dissertation investigates reduction phenomena in seven corpus studies on French and Spanish spontaneous speech. Chapter 2 presents the Nijmegen Corpus of Casual French. On the basis of this corpus and of a corpus of journalistic speech, Chapter 3 investigates the conditioning of French [t] duration by probabilistic factors. Chapter 4 studies the elision of vowel /e/ in the French word *c'était*, while Chapter 5 investigates the devoicing of phrase-medial high vowels in spontaneous French. Chap-

ter 6 compares the reduction of voiceless stops and vowels in spontaneous French and Spanish. Finally, the weakening of intervocalic /s/ consonants in spontaneous Spanish is investigated in Chapter 7. Chapter 8 provides a general discussion of the findings presented in previous chapters and a general conclusion.





# The Nijmegen Corpus of Casual French

---

## Chapter 2

This chapter is a reformatted version of:  
Francisco Torreira, Martine Adda-Decker and Mirjam Ernestus (2010).  
The Nijmegen Corpus of Casual French. *Speech Communication*, 52(3):201–22.

This article describes the preparation, recording and orthographic transcription of a new speech corpus, the Nijmegen Corpus of Casual French (NCCFr). The corpus contains a total of over 36 hours of recordings of 46 French speakers engaged in conversations with friends. Casual speech was elicited during three different parts, which together provided around ninety minutes of speech from every pair of speakers. While Parts 1 and 2 did not require participants to perform any specific task, in Part 3 participants negotiated a common answer to general questions about society. Comparisons with the ESTER corpus of journalistic speech show that the two corpora contain speech of considerably different registers. A number of indicators of casualness, including swear words, casual words, *verlan*, disfluencies and word repetitions, are more frequent in the NCCFr than in the ESTER corpus, while the use of double negation, an indicator of formal speech, is less frequent. In general, these estimates of casualness are constant through the three parts of the recording sessions and across speakers. Based on these facts, we conclude that our corpus is a rich resource of highly casual speech, and that it can be effectively exploited by researchers in language science and technology.

## 2.1 Introduction

French is one of the best documented languages in the world. Accordingly, researchers interested in spoken French have a choice among several speech corpora for their studies (e.g. ESTER (Galliano et al., 2005), PFC (Durand et al., 2005), see <http://catalog.elra.info/> for more). However, no existing corpus of French contains the large amounts of casual speech necessary for detailed research on the characteristics of this register, including inter- and intra-speaker variability. This article describes a new corpus of European French that fills this gap.

The specific characteristics of a given corpus present advantages and disadvantages depending on the researcher's goals. For instance, the ESTER corpus, with around 90 hours of journalistic recordings, mainly comprises prepared speech, either planned or read, from several European and North-African French-speaking radio stations. It is a valuable source for researchers interested in journalistic speech covering a broad range of topics with a huge lexical variety, produced by a large population of professional and occasionally intervening speakers in various audio conditions.

The PFC (Phonologie du Français Contemporain) corpus, which is still under development, contains recordings of speakers from diverse geographic and social backgrounds in the French-speaking world. In the future, it will consist of a total of several hundred hours of speech of different registers, thus providing a reference corpus for linguists who are interested in social and regional language variation. Every informant in the PFC corpus contributes an average of 20 minutes of speech, including both read speech and conversations with the interviewer. The conversations differ in their degree of casualness, as the interviewers and informants are not all on familiar terms. So far, only a relatively small percentage of these data has been orthographically transcribed. The recordings, which are field data collections, are of variable acoustic quality and hence not always appropriate for detailed acoustic analysis.

The motivation behind the creation of the Nijmegen Corpus of Casual French (NCCFr from now on) was to provide large amounts of high-quality recordings of casual speech suitable for phonetic analysis. The uniqueness of our corpus can be characterized as follows:

- It contains large amounts of casual speech. All of the recorded speech has been orthographically transcribed.
- It contains high-quality recordings captured with head-mounted microphones in a sound-attenuated room.

- It contains speech from 46 speakers sharing the same geographic and educational background. This allows researchers to study inter-speaker variation in a corpus controlled in terms of regional and social variation.
- It contains large amounts of data for every speaker (around 90 minutes of recorded conversation for every pair of speakers). This allows researchers to study within-speaker variability.
- It contains audio as well as video data, which can be used to study facial and body gestures during verbal communication.

Information about how to obtain a copy of the corpus can be found online at <http://mirjamernestus.ruhosting.nl/Ernestus/NCCFr>.

The present article provides a detailed description of the creation and characteristics of our corpus (Sections 2.2 and 2.3). In Section 2.4 we provide evidence of the casual register of the speech contained in the corpus by comparing the NCCFr and the ESTER corpus in terms of several uncontroversial indicators of casualness (e.g. lexical items sensitive to register, double negation, disfluencies). We also assess the variability of these indicators throughout the different parts of the recordings (see below for details) and across speakers.

## 2.2 Corpus creation

### 2.2.1 Participants

The corpus creation was initiated in November 2007. Twenty-three confederates were recruited at the University of Paris 3 Sorbonne Nouvelle, either by e-mail or personally. These confederates were briefly interviewed and asked to find two friends willing to participate in recordings of natural conversations. These friends will be referred to as *speakers* from now on. Every recording consists of a conversation among these three participants: a confederate and two speakers. All participants complied with the following conditions:

- They knew the two other participants in the recording well.
- They were of the same sex as the two other participants in the recording.
- They had completed the secondary education cycle in France.
- They had been raised in Central/Northern France.
- They reported not suffering from any pathology related to speech or hearing.

In total there were 46 speakers (24 female and 22 male). Thirty-four speakers came from the Paris region. The remaining 12 came from other regions in Central and Northern France. Except for two female speakers in their fifties, all speakers were university students aged between 18 and 27. The gender, age and regional background of each speaker is provided in Table 2.1.

### **2.2.2 Recording set-up**

The recording room was sound-attenuated and had an approximate size of 4 x 3 m. The participants sat on chairs around a table placed in the middle of the room. The confederate always sat on the south side of the table, while the speakers occupied the chairs on the north and west sides. Figure 2.1 shows the layout of the recording room.

For technical reasons, only two participants could be recorded. Given this limitation, we decided to dispense with the speech of the confederates, who, contrary to the speakers, were not naive about our goals and procedures. However, the confederates also wore a microphone in order to reinforce the speakers' impression that they did not have a special role in the conversation. Speakers were recorded in separate audio channels. The recording equipment consisted of an Edirol R-09 solid-state stereo recorder, Samson QV head-mounted unidirectional microphones and a stereo microphone preamplifier. Microphones were placed at an average distance of 5 cm from the left corner of the speakers' lips. The sampling rate used was 48 KHz, while quantization was set to 32 bits.

The conversations were filmed using a Canon XM2 Mini-DV video camera. The camera was placed on a tripod in a corner of the recording room as shown in Figure 2.1. The positioning of the camera allowed us to film the two speakers, but not the confederate. This is illustrated in Figure 2.2. Since awareness of being filmed could compromise the casualness of the conversations, we tried to make the speakers believe that the camera was turned off during the recordings. As a first step, a small piece of duck tape was placed on each of its lights. Additionally, an unplugged cable was left hanging from the camera in order to reinforce the impression that it was turned off. It should also be noted that there were several other objects in the room, including a desktop computer, several loudspeakers and other audio equipment. Our camera may have then appeared to be merely another piece of unused electronic equipment in the recording room.

### **2.2.3 Recording procedure**

The recording procedure was established after a series of pilot recordings was run. During these pilot recordings, we noticed that it was difficult to obtain casual speech

Speaker	Gender	Age	Region	Speech	Speaker	Gender	Age	Region	Speech
M01L	M	23	Île-de-France	45:27	F12R	F	21	Poitou-Charentes	55:45
M01R	M	25	Île-de-France	43:23	F13L	F	19	Île-de-France	39:38
M02L	M	24	Île-de-France	55:15	F13R	F	18	Nord-Pas-de-Calais	31:49
M02R	M	24	Île-de-France	49:24	F14L	F	20	Île-de-France	28:46
F03L	F	50	Île-de-France	45:09	F14R	F	23	Île-de-France	55:33
F03R	F	40	Île-de-France	40:28	M15L	M	27	Limousin	28:24
F04L	F	25	Île-de-France	19:49	M15R	M	23	Centre	52:53
F04R	F	25	Île-de-France	40:50	F16L	F	19	Île-de-France	44:08
F05L	F	19	Île-de-France	26:39	F16R	F	21	Île-de-France	58:17
F05R	F	18	Île-de-France	28:38	M17L	M	20	Île-de-France	33:52
F06L	F	21	Île-de-France	39:48	M17R	M	20	Île-de-France	33:02
F06R	F	21	Île-de-France	49:38	M18L	M	20	Île-de-France	40:18
F07L	F	20	Île-de-France	47:25	M18R	M	20	Île-de-France	27:46
F07R	F	20	Île-de-France	40:14	M19L	M	19	Haute-Normandie	29:23
F08L	F	21	Île-de-France	40:49	M19R	M	26	Île-de-France	27:40
F08R	F	20	Île-de-France	21:29	M20L	M	22	Bourgogne	34:53
M09L	M	24	Île-de-France	29:22	M20R	M	22	Bretagne	30:23
M09R	M	24	Île-de-France	35:33	M21L	M	21	Île-de-France	37:54
F10L	F	19	Île-de-France	43:36	M21R	M	19	Bretagne	24:04
F10R	F	20	Île-de-France	41:51	M22L	M	19	Basse-Normandie	27:20
M11L	M	18	Île-de-France	46:38	M22R	M	19	Haute-Normandie	27:39
M11R	M	18	Île-de-France	46:18	M23L	M	20	Île-de-France	34:17
F12L	F	22	Île-de-France	53:56	M23R	M	23	Île-de-France	33:41

Table 2.1: Gender, age, regional background (region of longest residence) and total amount of recorded speech (minutes:seconds) for every speaker in the corpus. Speaker code: M = male, F = female; L = left channel, R = right channel. The number in the speaker code indicates the recording number.

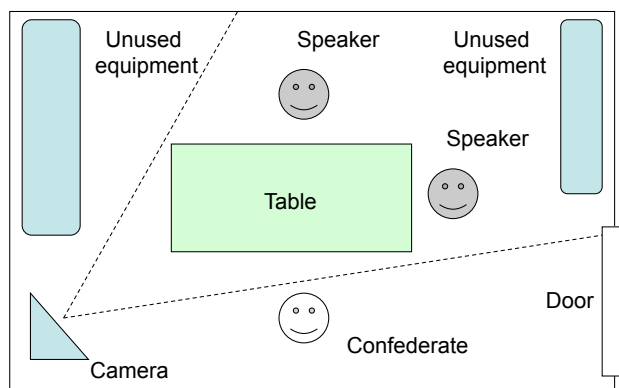


Figure 2.1: Layout of the recording room



Figure 2.2: Snapshot extracted from one of the films in the corpus.

for long periods in the absence of any explicit task or changes in the recording setting. We also noticed that the speech recorded during the initial moments of the session was often far from casual. In order to obtain lively casual speech from our speakers for 90 minutes, we divided the session into three different parts. In the first part, in order to elicit highly casual speech right from the beginning of the recording session, the two speakers were unexpectedly left alone on the false grounds that the confederate's microphone was defective. In the second part, the confederate returned to the booth and all participants engaged in free conversation. In the third part, participants were explicitly asked to perform a communicative activity in which they had to express and negotiate their views on real issues.

We now describe the recording preparations and each of these three parts in more detail. The recordings were conducted by the first author (FT from now on). FT is not a native speaker of French, but is highly proficient in this language.

*Preparations:* Confederates arrived at the Institut de Linguistique et Phonétique Générales et Appliquées (ILPGA) thirty minutes earlier than their friends. At this time, FT informed the confederates that it was their task to elicit natural speech from their friends, by raising familiar topics whenever the conversation seemed to approach a dead end. In order to maximize the amount of recorded speech from the speakers, confederates were instructed to try not to monopolize the conversations. The confederates were also informed that the conversations would be filmed, and where to sit so that only the other participants would appear in the film. They were asked not to unveil any of these details to their friends until the end of the recording. Finally, they were briefly instructed about the activity planned for the third part of the recording (see below for details). Moreover, they were asked to inform their friends that the instructions for this activity were the only reason for coming to the ILPGA earlier than them. At the end of the instruction section, confederates were asked to wait for the other participants in the entrance hall of the ILPGA.

At the time of the appointment, FT met the three participants at the entrance hall and asked them to wait while he got the keys of the recording room. He then returned to the recording room, started the video recording, turned off the lights and locked the door. Back at the entrance hall, he invited the participants to follow him to the recording room, making sure that the confederate would be the first person to enter in order to prevent the other participants from taking the confederate's seat. Once in the room, participants were asked to stay seated and not to touch their microphones or play with any other object (e.g. keys, watch) during the recording.

*Part 1:* After adjusting the recording volume during the first two minutes of the conversation, FT entered the recording room and informed the participants that the confederate's microphone was not working properly. He then asked the confederate to come out of the room in order to test a new microphone. At this moment,

the speakers left in the room did not know with certainty whether they were being recorded. It was precisely then that the recording was started. In our opinion, this situation elicited very natural speech from the beginning of the recording (see Section 2.4 below).

*Part 2:* After a period of ten to thirty minutes depending on the liveliness of the conversation, confederates were asked to go back into the room and join their friends. The conversation then held by the three friends constituted the second part of the recordings. The topics addressed during this part were usually a follow-up to those addressed in the first part, but with the novelty of a new participant. Among the conversation topics addressed by the speakers during this part were their college exams, the ongoing strike, parties, and travel plans. No instructions were provided about the topics to be discussed during this part of the conversation.

*Part 3:* After a period of thirty to forty minutes, FT entered the room and provided the participants with a sheet of paper describing the activity for the remaining part of the recording session. The participants were asked to choose at least five questions about political and social issues from a list (see Appendix), and then negotiate a unique answer for every question. In order to encourage them to negotiate rather than just discuss their views, we informed that they would have to write down their answers at the end of the recording session. The average duration of this part was around forty minutes.

At the end of the recording, we revealed our procedures to the participants and paid 30 euros to each of the speakers and 45 euros to the confederate as a compensation for their time. We then handed them a consent form and explained to them that they should only sign it if they fully agreed with its content, and that refusing to sign it would not cause them any trouble. Furthermore, we offered participants the opportunity to add restrictions to the distribution of the corpus. All of the participants signed the consent form. Two participants required that their recordings be not distributed online.

## **2.2.4 Orthographic transcription**

### **Transcription protocol**

The corpus was orthographically transcribed by two professional transcribers using the TRANSCRIBER software (Barras et al., 2001) following transcription guidelines developed at LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur) in line with the French GARS conventions (Blanche-Benveniste, 1990). The transcribers were recruited on the basis of earlier transcription experience in several French and European projects with interactive spontaneous speech (e.g. the ESLO/ESLO2 project (Serpellet et al., 2007), the SNCF Recital project



Event type	Symbol	Example
Mispronunciation	* prefix	* <i>légaliser</i> for [legazile]
Proper name	^^ prefix	^^ <i>Joffroy</i>
Verlan	^^ prefix	^^ <i>chelou</i> (for <i>louche</i> )
Standardized abbreviation	\$ suffix	<i>fac</i> \$ (for <i>faculté</i> )
Truncated words	()	<i>re(gardes)</i> (for <i>regardes</i> )
Interjection	& prefix	& <i>ben</i> , & <i>pf</i> , & <i>eu</i>

Table 2.2: Transcription symbols.

(see [http://recherche.sncf.com/uk/projets/uk\\_recital.html](http://recherche.sncf.com/uk/projets/uk_recital.html)), the ARISE project (Lamel et al., 2000)).

The speech of each speaker was orthographically transcribed in a separate annotation file using mono-channel audio streams. The audio stream was manually segmented, separately for each speaker, into small chunks of a few seconds. Most chunks contain either speech or long silent pauses, but some consist entirely of speaker noises (e.g. laughter). The transcription guidelines stated that speech in a chunk should have a clear degree of syntactic and semantic coherence and contain no long stretches of silence. In total, over 83,000 chunks were marked, with an average duration of 3.12 seconds.

Transcribers were asked to provide standard orthographic transcriptions following the French *Robert* dictionary (Le Petit Robert, 2007) wherever possible. Hence all pronunciation variants of the same word (e.g. those resulting from the addition of final schwas (Fagyal, 1998)) were annotated with the same orthographic form. However, not all speech events can be handled by a normative written language dictionary. Transcription problems arise for mispronounced words, words with an uncertain spelling (proper names, neologisms, onomatopoeia, slang...) and for unintelligible parts. The guidelines provided a series of special symbol affixes to annotate these speech events. Table 2.2 lists the most important ones. Notice that the same affix was used for proper names with an uncertain spelling and *verlan* words (for an explanation of the term *verlan* see Section 2.4.3 below). Since proper names always start with a capital letter, this convention does not lead to confusion between these two types of words.

Although the TRANSCRIBER tool proposes a list of specific noise labels, our transcribers were encouraged to fall back to a generic noise label [*b*] for all noises except for frequent and easily identifiable noises such as respiration and laughter. The label [*r*] was used for respiration noises and the label [*rire*] ('laughter') for laughter. These three labels (i.e. [*rire*], [*r*] and [*b*]) account for 87% of all noise labels in the NCCFr corpus.

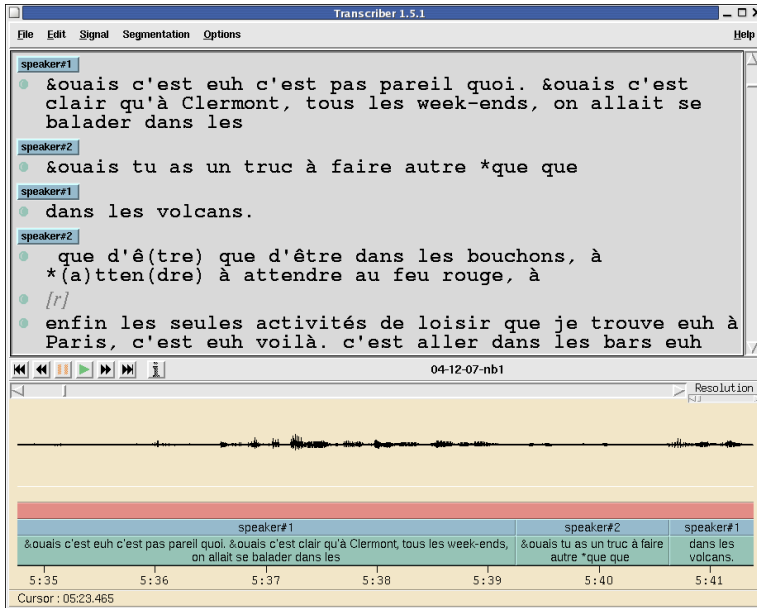


Figure 2.3: Manual transcription sample illustrating the segmentation and transcription steps. The two speaker streams are merged to restore the conversation structure.

Transcribers were asked to restore common elisions and contractions to their full orthographic forms. For instance, the guidelines specified that expressions characteristic of casual speech such as *y a* ‘there is’ or *j’sais pas* ‘I don’t know’ should be transcribed in their full forms as *il y a* and *je sais pas*. The reason for requiring standard full forms is that providing detailed transcriptions is very time consuming and error prone (Ernestus, 2000). Moreover, non-standard transcriptions make searching for particular lexical items difficult. However, one exception was made to this rule: the guidelines recommended that cases of obvious *ne* deletion (in the French double negations, such as *ne . . . pas* and *ne . . . plus*, see Section 2.4.4) should not be restored in the transcription. In case of doubt the *ne* particle should be transcribed.

Figure 2.3 shows an excerpt of a conversation with transcribed chunks of various relatively small lengths. To restore the conversation structure, the individual speaker transcription streams have been merged in the figure. The produced transcription files are in a machine readable XML mark-up language format, and are also available in Praat TextGrid format (Boersma and Weenink, 2009).

## Transcription quality check

The quality and consistency of manual transcripts can be assessed via automatic alignment with the acoustic signal: successfully aligned parts of the corpus guarantee a good fit between the manual transcripts and the recorded speech signal. We followed this approach in order to check the quality of our transcriptions. We first created a pronunciation dictionary containing all words in the corpus and their canonical pronunciations, which were then used in an automatic alignment with the speech signal.

The corpus contains 15,919 distinct word types, including over 10% of word fragment types (truncated pronunciations). About 14,000 entries already existed in the LIMSI transcription system vocabulary (which comprises around 200,000 entries). The additional 2,000 items were checked and added to the system vocabulary with appropriate pronunciations. Among these new entries, around 1,000 correspond to word fragment types. The other entries include *verlan* words, interjections, onomatopoeia and apocopes.

We then segmented the audio stream into words given the orthographic transcription and the pronunciation dictionary using the LIMSI recognition system (Gauvain et al., 2005). If a given transcription does not fit with the corresponding audio chunk, the alignment system will tend to reject the chunk without producing the alignment. The quality of the transcriptions can therefore be measured via chunk rejection rates. Only 41 speech chunks out of around 83,000 were rejected, corresponding to less than three minutes of speech. This strongly suggests that the orthographic transcription is of a high quality.

Transcribers often feel uncomfortable with the transcription rule prescribing the restoration of omitted and contracted words. For this reason, we also checked the extent to which the transcribers followed this rule by manually examining the two most common types of restoration: (1) *il y a* ‘there is’ and morphologically-related word sequences (e.g. *il y avait* ‘there was’), and (2) the pronoun *tu* ‘you’ followed by a verb starting with an /a/ vowel (e.g. *tu as* ‘you have’), in which the *tu* subject pronoun tends to be pronounced as [t], becoming a homophone with the object pronoun *t’*. Whereas for *il y a* both transcribers observed the rule in almost 100% of the cases, instances of *tu* followed by an /a/-initial verb were only restored in less than a third of the occurrences (842 out of 2,316). A possible explanation for this tendency might be that *t’* exists as an orthographically correct form for the object pronoun: *il t’a vu* (‘he has seen you’) is perfectly correct, but *t’as vu le film* (‘you have watched the film’) is not correct in written French. The manual transcripts have been updated to restore these elisions and contractions.

	Total	Average	sd	min	max
<i>Effective speech</i>	26h 07' 04"	1h 08' 08"	12' 25"	50' 13"	1h 30' 20"
<i>Overlapping speech</i>	3h 21' 50"	8' 46"	4' 44"	3' 30"	23' 46"
<i>Non-effective-speech</i>	10h 05' 25"	26' 19"	9' 47"	9' 54"	46' 19"
Total	36h 12' 29"	1h 34' 27"	7' 45"	1h 22' 50"	1h 52' 25"

Table 2.3: Amounts of *effective speech*, *overlapping speech* and *non-effective-speech* in the corpus, along with averages per recording, plus standard deviations and ranges. *Non-effective-speech* includes speech from confederates.

## 2.3 Corpus contents

The NCCFr consists of 23 recordings involving a total of 69 participants (23 confederates and 46 speakers). As explained in Section 2.2.2, only the speech of the two speakers was recorded. In most cases, however, the speech of the confederate was captured by the speakers' microphones and can be well interpreted from the speakers' recordings.

Table 2.3 shows the amount of speech contained in the corpus, both in total and averaged by recording. *Effective speech* includes all stretches of the recording containing speech by one of the two speakers, or by both at the same time (*overlapping speech*). This was calculated by adding the durations of all chunks in the transcriptions of each speaker containing at least one lexical item and subtracting the duration of *overlapping speech*. *Overlapping speech* was calculated by summing up the durations of stretches of conversation where both speakers spoke simultaneously. *Overlapping speech* involving a confederate and a speaker could not be estimated, since the speech of the former was not transcribed. *Non-effective-speech* includes stretches of conversation not containing *effective speech* by any of the two speakers. It does not only include silence, laughter and other speaker noises, but also non-overlapping speech from confederates. Overall, the corpus contains over 36 hours of recorded conversations, with over 26 hours of *effective speech*, over 3 hours of *overlapping speech* and around ten hours of *non-effective-speech* including silence and speech from confederates. The considerable amount of *overlapping speech* indicates that the corpus contains highly interactive speech (Schegloff, 2000).

Table 2.4 shows the total and average durations of each of the three parts of the recorded conversations, along with their average amounts of *effective speech* and *non-effective-speech*. Notice that *non-effective speech* in Part 1 refers to stretches of the conversations containing silence, laughter or other speaker noises, while *non-effective-speech* in Parts 2 and 3 also contains speech turns from confederates (remember that confederates were not recorded and only participated in Parts 2 and 3). It can be seen from this table that Parts 2 and 3 appear very similar in the

	Total	Average	<i>Effective Speech</i>	<i>Non-effective-speech</i>
Part 1	7h 32' 07"	19' 39"	16' 40" (85%)	2' 58" (15%)
Part 2	12h 52' 11"	33' 34"	23' 22" (70%)	10' 13" (30%)
Part 3	15h 48' 09"	41' 13"	28' 05" (69%)	13' 08" (31%)

Table 2.4: Duration of each recording part in the corpus and on average per recording, along with average amounts of *effective speech* and *non-effective-speech* for each part. *Non-effective-speech* includes speech from confederates.

percentages of effective speech that they contain, and that the inclusion of a confederate in these two parts leads to a similar decrease in *effective speech* with respect to Part 1. This was confirmed by a series of two-tailed t-tests showing that the percentage of *effective speech* in Part 1 differed significantly from those of Parts 2 ( $t = 3.06, p < .005$ ) and 3 ( $t = 3.6, p < .001$ ), but it did not differ between the latter two ( $t = 0.33, p = .74$ ).

The right-most column in Table 2.1 shows the total amount of recorded speech for every speaker. These amounts ranged from roughly twenty to sixty minutes, with an average of 38 minutes and 27 seconds and a standard deviation of 10 minutes.

## 2.4 Assessing casualness

In spite of our efforts to create an informal atmosphere during the recording sessions (for instance by inviting groups of friends), it is possible that speakers felt intimidated or inhibited by the awareness of being recorded. Therefore, in the absence of any proof to the contrary, the casualness of the speech contained in our corpus may be legitimately questioned. In this section we examine several indicators of spontaneity and casualness that can be extracted automatically from an orthographic transcription, and compare their values in our corpus and in the ESTER corpus of journalistic speech.

### 2.4.1 Disfluency words

We believe that genuine casual speech should be, among other things, unprepared and spontaneous. For this reason, we first quantify the incidence of disfluencies by identifying transcribed filled pauses (e.g. *eah, hum, ben*), which are known to be more frequent in spontaneous speech than in more careful and formal styles (e.g. Clark and Wasow, 1998; Shriberg, 2001; Fox Tree, 1995; Clark, 1996). Following Jousse et al. (2008), we measured the frequencies of the most common word types used by transcribers to annotate filled pauses and hesitations, that is *ben, euh* and

	NCCFr		ESTER	
<i>ben</i>	2.77	(1,292)	.26	(176)
<i>ehh</i>	24.77	(11,546)	8.14	(5,452)
<i>hum</i>	9.42	(4,391)	.11	(73)
Rep. bigrams	12.94	(6,034)	3.76	(2,522)
Rep. trigrams	1.74	(815)	.49	(333)

Table 2.5: Frequencies of *ehh*, *hum*, *ben* and repetition bigrams and trigrams in the two corpora per thousand words. Numbers within brackets indicate absolute numbers of occurrences.

*hum*. These words will be referred to as *disfluency words* from now on. The transcription guidelines of the NCCFr and the ESTER corpus do not differ in how they specify the annotation of filled pauses, and we have not noticed salient differences in the transcription of filled pauses between the two corpora. Therefore, we assumed that the transcribers of both corpora annotated filled pauses with the same accuracy and following similar principles. The first three lines in Table 2.5 suggest that the NCCFr corpus contains considerably more filled pauses and hesitations than the ESTER corpus, and thus confirm our expectations.

#### 2.4.2 Word repetitions

We also counted the number of word bigrams and trigrams consisting of identical words occurring in each corpus. Following Jousse et al. (2008), we assume that word repetitions mostly result from breakdowns during online speech planning and are therefore characteristic of spontaneous speech. The word bigrams *vous vous* and *nous nous*, which form grammatical sequences (e.g. *vous vous voyez* ‘you see yourself’, *nous nous connaissons* ‘we know each other’), were excluded from the repetition bigram count. It should be also noticed that repetitions can be used as a stylistic device to intensify the meaning of a word (e.g. *trop, trop* ‘very, too much’, *partout, partout, partout* ‘everywhere’), or as a backchannel utterance (e.g. *oui, oui...* ‘yes’). Importantly, however, the latter two types of word repetitions are also characteristic of casual speech. We therefore did not exclude these sequences from our counts. The last two rows in Table 2.5 show that sequences of repeated words are more frequent in our corpus than in the ESTER corpus. Again, these numbers suggest that the NCCFr contains more spontaneous speech than the ESTER corpus.

	NCCFr		ESTER	
<i>chier</i>	.23	(110)	.00	(0)
<i>con</i>	.21	(102)	.00	(2)
<i>cul</i>	.06	(31)	.00	(0)
<i>merde</i>	.32	(152)	.00	(1)
<i>putain</i>	.79	(370)	.00	(0)

Table 2.6: Frequencies of occurrence per thousand words for five swear words in the NCCFr and the ESTER corpus. Numbers within brackets indicate absolute numbers of occurrences.

### 2.4.3 Lexical items

An obvious way of assessing the casualness of a corpus is to check the extent to which it contains lexical items typical of casual speech. In order to do this, we examined the frequency of occurrence of swear words and *verlan* (see below for details), and also compared the use of informal and formal words with a similar meaning. We determined which casual words and swear words would be considered for analysis by asking four native speakers of French to provide two lists. The first list should contain a subjective choice of the ten most common French swear words, while the second should consist of formal and informal content words having similar meanings (e.g. *chose / truc* ‘thing’). The lists of swear words were very similar, as six terms were present in all of them. On the other hand, only two pairs of formal and informal words were present in all of the second lists.

From the first lists, we selected for analysis those swear words that occurred at least ten times in either the NCCFr or the ESTER corpus. The threshold was set at ten so that interpretable comparisons could be made (comparisons of very low frequencies, say three and one, would have been hard to interpret). This threshold also allowed for a reasonable number of comparisons between the corpora. Table 2.6 shows the frequency of occurrence for each of these swear words in both corpora. We were surprised to find out that swear words were highly frequent in our corpus (e.g. on average, *putain* occurs roughly once every six minutes of conversation). In our view, such a frequent usage of swear words constitutes strong evidence of the casual speech register of our recordings (Egins and Slade, 1997).

From the second lists, we only retained those pairs of which each member appeared at least ten times in one of the two corpora. When two pairs shared the same formal word, they were reorganized into a triplet (e.g. formal: *garçon*; informal: *mec, gars* ‘lad’). We added two pairs of function words (i.e. *cela / ça* and *oui / ouais*) which in our opinion are very good indicators of register as well. Notice that our subjects had been asked to provide pairs of content words, and had therefore

Word	NCCFr		ESTER		Gloss
<i>ami(s)</i>	.13	(65)	.14	(94)	friend(s)
<b><i>pote(s)</i></b>	.16	(87)	.00	(1)	
<i>argent</i>	.21	(98)	.11	(74)	money
<b><i>tune(s)</i></b>	.05	(29)	.00	(0)	
<i>cela</i>	.01	(6)	.55	(369)	that (Pron.)
<b><i>ça</i></b>	17.75	(8,276)	1.72	(1,152)	
<i>chose(s)</i>	1.25	(587)	.53	(358)	thing(s)
<b><i>truc(s)</i></b>	3.01	(1,400)	.00	(1)	
<i>fille</i>	1.13	(531)	.07	(51)	girl(s)
<b><i>nana(s)</i></b>	.11	(52)	.00	(0)	
<i>fou</i>	.15	(71)	.02	(18)	crazy
<b><i>dingue</i></b>	.08	(39)	.00	(2)	
<i>garçon(s)</i>	.58	(271)	.02	(20)	lad(s)
<b><i>gars</i></b>	.48	(226)	.00	(2)	
<b><i>mec(s)</i></b>	.67	(315)	.00	(2)	
<i>livre(s)</i>	.05	(27)	.14	(97)	book(s) (N)
<b><i>bouquin(s)</i></b>	.01	(48)	.00	(2)	
<i>mange(r)</i>	.17	(81)	.03	(24)	eat(s) (V)
<b><i>bouffe(r)</i></b>	.08	(41)	.00	(3)	
<i>oui</i>	6.32	(2,949)	.83	(558)	yes
<b><i>ouais</i></b>	17.89	(8,343)	.02	(15)	
<i>travail</i>	.19	(90)	.35	(235)	work (N)
<b><i>boulot</i></b>	.08	(38)	.00	(4)	
<i>très</i>	1.33	(622)	1.64	(1,099)	very
<b><i>vachement</i></b>	.30	(141)	.00	(0)	

Table 2.7: Frequencies per thousand words for casual (in bold) words and their standard variants in the two corpora. Numbers in brackets indicate absolute numbers of occurrences. (V = verb; N = noun; Pron. = pronoun).



<i>Verlan</i>	Frequency	Standard form
<i>ouf(s)</i>	80	<i>fou</i> ‘crazy’
<i>meuf(s)</i>	63	<i>femme(s)</i> ‘woman/women’
<i>relou</i>	32	<i>lourd</i> ‘heavy-going’
<i>chelou</i>	14	<i>louche</i> ‘dodgy’
<i>vénière(s)</i>	11	<i>énervé(s)</i> ‘angry’
<i>rebeu</i>	10	<i>arabe</i> ‘arab’

Table 2.8: Frequent *verlan* words and their numbers of occurrences in the NCCFr.

not mentioned any of these two function words. Table 2.7 shows the frequencies of occurrence of these words in both corpora. This table shows that all casual words are more frequent in the NCCFr than in the ESTER corpus. Moreover, some casual words in our corpus are more frequent than their more formal synonyms (e.g. *ça*, *ouais*, *truc* and *mec* occur more often than *cela*, *oui*, *chose* and *garçon*). So far, these facts lead us to conclude that our speakers did not generally aim at a formal register of speech during the conversations.

We also checked the usage of *verlan* in our corpus. *Verlan* is a language game typically consisting in the inversion of segments and syllables in a word, often accompanied by other changes, affecting for instance the quality of vowels. The name *verlan* /vɛʁlɑ̃/ itself is an example of such inversion, as it comes from *l’envers* /lɑ̃vɛʁ/ ‘the inverse’. Importantly for our purposes, the use of *verlan* can be used as an indicator of casualness, as it is common in slang and youth language (Valdman, 2000). *Verlan* word types used in the NCCFr were identified on the basis of the prefix *~* in the orthographic transcriptions (see Section 2.2.4). There was a total of 14 word types and 232 tokens of *verlan* words. The most frequent ones ( $n > 10$ ) are listed in Table 2.8 with their number of occurrences. It should be noticed that none of these words appeared in the ESTER corpus. The occurrence of *verlan* in the NCCFr corpus constitutes further evidence that it contains highly casual speech.

#### 2.4.4 Double negation

Negation in French requires the use of two grammatical particles, the first of which must be *ne* (or its contracted form *n’* before a vowel). For instance, in the utterance *Je ne veux pas dormir* ‘I don’t want to sleep’, the negation particle *pas* appears after the verb *veux*, while the negative particle *ne* precedes it. In the same way, the word *ne* occurs along other negative particles such as *rien* ‘nothing’, *jamais* ‘never’ or *aucun* ‘any’. Importantly for our purposes, casual French is characterized by the frequent elision of the particle *ne* (Coveney, 1996; Armstrong and Smith, 2002). For instance,

	NCCFr		ESTER	
<i>ne/n'</i>	1.32	(676)	3.10	(3,664)
<i>pas</i>	19.31	(9,001)	4.55	(3,503)
<i>rien</i>	1.19	(557)	.28	(193)
<i>jamais</i>	.68	(321)	.26	(179)
<i>aucun(e)</i>	.21	(97)	.35	(239)
Double Neg. %	6.7		89.0	

Table 2.9: Frequencies of negation particles in the two corpora per thousand words, and estimated percentage of double negation (Double Neg. %). Numbers within brackets indicate absolute numbers of occurrences.

in informal settings *Je veux pas* 'I don't want' is often heard instead of *Je ne veux pas*.

We investigated how often negation occurred in both corpora without the first element *ne*. Our goal was to identify rough differences in the use of double negation between the two corpora, rather than make our estimates of double negation as accurate as possible. Therefore, instead of checking every instance of double negation manually, we automatically extracted the frequency of *ne* in each corpus and compared these with the automatically extracted frequencies of negation particles *pas* 'not', *rien* 'nothing', *jamais* 'never' and *aucun(e)* 'any'. Negation particles whose orthographical form may also occur with other meanings (e.g. *personne* 'person' and 'nobody'; *que* 'that' and 'only') were not examined. An exception was made for *pas*, since it occurs far more often as a negative particle ('not') than as a noun ('step').

Table 2.9 shows the frequencies of occurrence of *ne* and of other negation particles in both the NCCFr and the ESTER corpus. An estimate of the percentage of double negation usage was computed by dividing the number of *ne* occurrences by the total number of occurrences of the other negation particles. As expected, double negation turned out to be very infrequent in our corpus (6.7%), suggesting that the register of the recorded conversations was highly casual and informal. On the other hand, the ESTER corpus exhibits a high rate of double negation (89%), confirming that a more formal register was used in the journalistic speech materials.

#### 2.4.5 Homogeneity across parts

Tables 2.4-2.8 show that the NCCFr contains highly casual speech in spite of the fact that speakers were conscious of being recorded. Since the recordings consisted of three different parts, we investigated whether these parts differed in their degree of casualness. It might be expected, for instance, that Part 3 contained less casual

speech, since it involved discussing a number of prescribed topics. We therefore examined the distribution of lexical and disfluency indicators across the different parts.

In previous subsections in which we compared the NCCFr and the ESTER corpus, we examined indicators that occurred at least ten times in one of the two corpora. The same restriction cannot be applied to a comparison of the three parts in our recordings, since frequencies slightly above ten are too low to obtain interpretable differences in this case (e.g. an indicator with four occurrences in Part 1, two occurrences in Part 2 and six occurrences in Part 3 does not provide information about whether the three parts are different). We therefore decided to investigate only those indicators that appeared in our corpus at least 100 times (*chier, con, merde, putain, cela / ça, garçon(s) / gars / mec(s), oui / ouais, très / vachement, ben, euh, hum*, word repetitions). The percentage of double negation, which could be reasonably well estimated for each part, was also included in this comparison.

Table 2.10 shows our findings. The usage of swear words and casual words does not exhibit significant differences across parts. The only exception perhaps is the word *garçon*, which was over twenty times more frequent in Part 3 than in Parts 1 and 2 combined. This increase may be explained by the fact that one of the questions included in the activity performed during Part 3 explicitly mentioned the word *garçon* (*Pourquoi les garçons et les filles ne sont-ils pas éduqués de la même manière?* ‘Why aren’t boys and girls raised in the same way?’). As in the case of swear words and casual words, word repetitions and double negation appear to be equally distributed across the three parts.

We tested for systematic differences in the frequency of casualness indicators across the three parts by fitting a mixed-effects linear model with log normalized frequency as the predicted variable, recording part as predictor and casualness indicator (e.g. *ça, truc(s), chier, euh*) as random factor. Since double negation was estimated as a percentage, it was not included in the analysis. From the pairs and triplets of formal and informal words, only informal words were retained for analysis. No statistical effect of recording part on log normalized frequency was identified ( $F(2, 36) = 0.79, p > .1$ ), suggesting that parts did not differ systematically in their degree of casualness.

#### **2.4.6 Homogeneity across speakers**

We finally assessed the distribution of indicators of casual speech across speakers. Our goal was to check if the casual characteristics of the NCCFr revealed by our pre-

	Part 1		Part 2		Part 3	
<i>chier</i>	.10	(13)	.14	(24)	.35	(73)
<i>con</i>	.26	(32)	.17	(30)	.19	(40)
<i>merde</i>	.38	(47)	.26	(45)	.28	(60)
<i>putain</i>	1.21	(151)	.72	(124)	.45	(95)
<i>cela</i>	.00	(1)	.00	(2)	.00	(3)
<b>ça</b>	15.87	(1,984)	15.19	(2,633)	17.36	(3,655)
<i>chose(s)</i>	.99	(124)	.91	(157)	1.45	(306)
<b>truc(s)</b>	3.58	(448)	2.76	(479)	2.25	(473)
<i>garçon(s)</i>	.10	(12)	.00	(2)	1.22	(257)
<b>gars</b>	.30	(37)	.64	(111)	.37	(78)
<b>mec(s)</b>	.46	(57)	.70	(121)	.65	(137)
<i>oui</i>	6.44	(805)	4.91	(851)	6.14	(1,293)
<b>ouais</b>	18.32	(2,290)	15.66	(2,715)	15.85	(3,338)
<i>très</i>	1.15	(144)	1.49	(259)	1.04	(219)
<b>vachement</b>	.46	(58)	.20	(35)	.23	(48)
<i>ben</i>	2.07	(259)	2.47	(428)	2.87	(605)
<i>euh</i>	24.11	(3,014)	22.02	(3,817)	22.38	(4,713)
<i>hum</i>	1.05	(131)	.76	(132)	.65	(136)
Rep. bigrams	11.29	(1,412)	11.99	(2,080)	12.07	(2,542)
Rep. trigrams	1.43	(179)	1.58	(275)	1.71	(361)
Double negation	4.9%		6.5%		7.9%	

Table 2.10: Frequencies of casualness indicators per thousand words within each recording part (lexical items, disfluency words, word repetitions) and percentage of double negation. Numbers within brackets indicate absolute numbers of occurrences. Double negation was estimated as explained in Section 2.4.2. Only lexical items that appeared at least 100 times in the NCCFr were included in this table. Rep. stands for repetition.

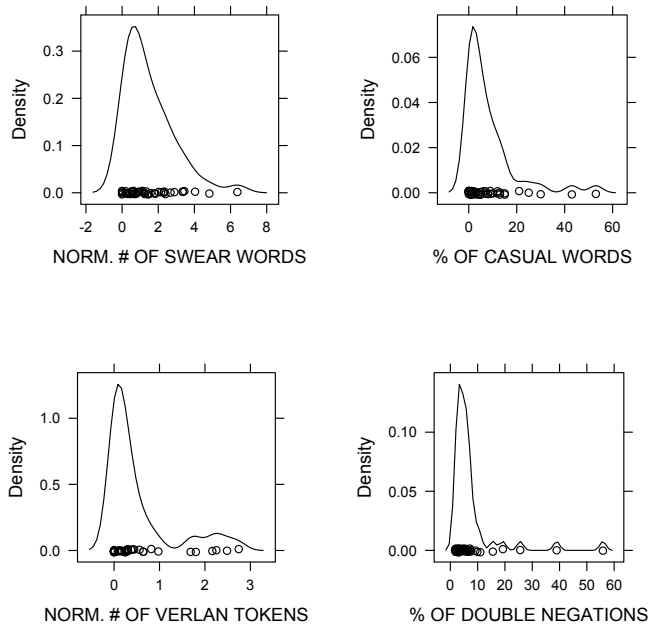


Figure 2.4: Kernel density plots of within-speaker frequencies of swear words (*putain*, *merde*, *chier*, *con* and *cul*) and *verlan* (normalized per thousand words), and percentages of casual content words and of double negation. Circles represent individual speakers. Except for double negation, higher values indicate higher degrees of casualness. NORM. # stands for normalized number.

vious analyses were due only to a small group of speakers. Figures 2.4 and 2.5 show kernel density plots<sup>1</sup> of within-speaker estimates for every indicator of casualness.

The top left panel of Figure 2.4 shows that a few speakers used more swear words than the rest, but overall the distribution of swear word frequencies is skewed only very slightly. Importantly, only five out of the 46 speakers did not pronounce any of the five swear words we selected for analysis.

Casual word use (use of the casual members of casual/non-casual word pairs) was estimated following a procedure different from the one used in the previous section. This time we added up the total number of tokens of casual and formal

<sup>1</sup>Kernel density plots display the estimated probability density function (y-axis) of a continuous random variable (x-axis), and have a purpose similar to that of histograms. The kernel density plots shown here were computed using the *density* function in the statistical software R with default parameters. For further details, see the R manual (R Development Core Team, 2008) and Sarkar (2008).

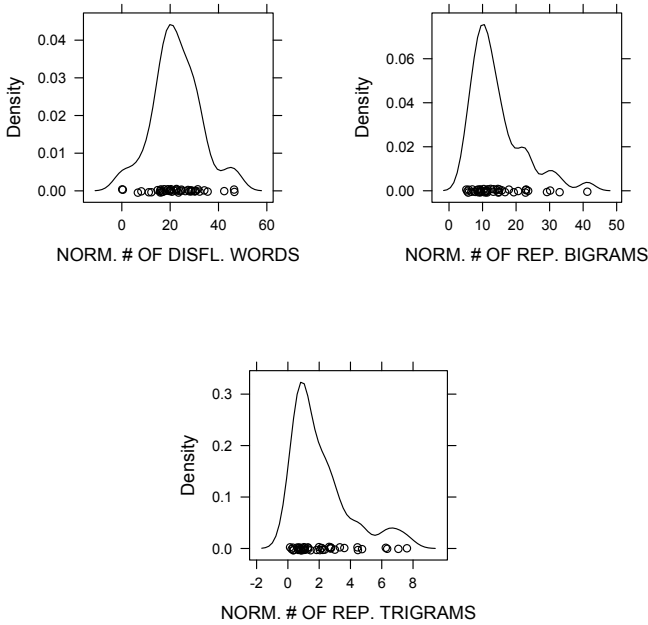


Figure 2.5: Kernel density plots of within-speaker frequencies of disfluency words (*euh*, *hum*, *ben*) and repetition bigrams and trigrams (normalized per thousand words). Circles represent individual speakers. NORM. # stands for normalized number.

content words listed in Table 2.7 pronounced by each speaker, and then calculated the percentage of casual words over this total. The reason for doing this was that, in order to plot the data, we needed a single score for each speaker, rather than multiple scores corresponding to different formal/informal pairs and triplets. Function words (i.e. *oui* / *ouais* and *cela* / *ça*) were considered separately from other casual words and are not included in the figure, since they were overwhelmingly more frequent than content words and would have had too big an impact on the measure. Casual word use ranged from 0% to 92%. Only 8 speakers did not pronounce any of the casual words considered for analysis and therefore scored very low on this indicator of casualness. With respect to the function words, the word *ça* was used by all speakers, while the few occurrences ( $n = 6$ ) of the more formal variant *cela* were shared by three speakers. The word *ouais* showed more variability, with 32 speakers showing a use between 30% to 95%, with a mean of 69.8%, and 14 speakers not using *ouais* at all. Interestingly, these 14 speakers used *oui* as often as the other

participants. *Verlan* was used by 60% of the speakers.

The other indicators also showed that most speakers used casual speech. Double negation use was generally low across speakers, as expected from our previous analyses, with only a small number of significantly deviant speakers: three speakers displayed double negation rates between 15% and 30%, and two showed surprisingly high rates (38.9% and 55.8%). Furthermore, all speakers exhibited at least five repetition bigrams per thousand words, and disfluency words were used by all except two speakers.

We finally checked whether speakers with low scores for a specific indicator of casualness also exhibited low scores for other indicators. This appeared to be the case only for one speaker who pronounced zero casual words, two swear words and displayed a high percentage of double negations (55.88%). Notice, however, that this speaker's double negation percentage was still much lower than that found in the ESTER corpus (89%). All other speakers who happened to display a low score for one indicator did not have particularly low scores for the other indicators. We therefore conclude that, in spite of individual differences in terms of specific indicators, the vast majority of speakers produced highly casual speech.

## 2.5 Discussion

In the previous sections we have described a new speech corpus, the Nijmegen Corpus of Casual French. The corpus contains a total of over 36 hours of orthographically-transcribed recordings involving 23 pairs of speakers with similar social and geographical backgrounds. These numbers should make it possible to model and study in detail the characteristics of spontaneous speech and inter- and intra-speaker variation. Our corpus can also be used to study gender differences, since gender was explicitly controlled for in our selection of speakers.

Our comparison of the NCCFr and the ESTER corpus of journalistic speech in terms of several indicators of casualness shows that our new corpus contains speech of a more casual nature. The high frequencies of swear words, casual words, *verlan*, disfluency words and word repetitions along with the low usage of double negation suggests that speakers generally aimed at a casual speech register in spite of the awareness of being recorded. The analyses in Sections 2.4.5 and 2.4.6 further suggest that this casual register was present throughout the different parts of each recording and in all speakers (excepting perhaps one male speaker). The NCCFr can therefore be used as a resource to investigate all sorts of linguistic phenomena related to casual speech, such as speech reduction (Ernestus, 2000; Johnson, 2004), disfluencies (Clark and Wasow, 1998), or the prosodic and syntactic characteristics of unprepared speech, among many other possible topics.

Every recording session was divided into three parts so that natural speech could be for long periods of time. A welcome consequence of this division is that specific parts of the corpus can be used to study specific phenomena. For instance, Part 1, in which speakers were left alone unaware of being recorded, is a good resource for researchers interested in talker interaction, turn-taking and conversation analysis in general (e.g. Local, 2003, 2007; Plug, 2005). Parts 2 and 3 can also be used for the same purposes, but the presence of a confederate, whose speech was not directly recorded, may complicate the study of these subjects. Part 3, in which participants were asked to choose and discuss specific topics, can be used to study argumentation and strategies used by speakers to convince their interlocutors. It can also be used to study the phonetics of specific content words, since many groups of speakers produced the same content words while discussing the same questions during Part 3. The description of the corpus provided in this article should allow researchers to judge which part best suits their purposes.

Finally, we hope that the corpus will be of use for researchers in different fields of speech technology. For instance, given the challenge that spontaneous speech presents to ASR systems (Moore, 2003, 2005), annotated resources such as the NCCFr) may help to improve current technology.

In conclusion, the Nijmegen Corpus of Casual French is a rich source of high-quality speech data that will help researchers to study spontaneous speech from many perspectives.



## Probabilistic effects on French [t] duration

---

### Chapter 3

This chapter is a reformatted version of:  
Francisco Torreira and Mirjam Ernestus (2009). Probabilistic effects on French [t] duration.  
In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, pages 448–451. Causal Productions Pty Ltd.

The present study shows that [t] consonants are affected by probabilistic factors in a syllable-timed language as French, and in spontaneous as well as in journalistic speech. Study 1 showed a word bigram frequency effect in spontaneous French, but its exact nature depended on the corpus on which the probabilistic measures were based. Study 2 investigated journalistic speech and showed an effect of the joint frequency of the test word and its following word. We discuss the possibility that these probabilistic effects are due to the speaker's planning of upcoming words, and to the speaker's adaptation to the listener's needs.

### 3.1 Introduction

A number of recent corpus studies have shown that probabilistic factors such as the frequency of occurrence and contextual predictability of specific linguistic units (e.g. words, syllables, segments) may affect articulation (Pluymaekers et al., 2005; Bell et al., 2003, 2009; Aylett and Turk, 2004). For instance, Pluymaekers et al. (2005) found a shortening effect of the frequency of a word on the duration of several affixes and affix-internal segments drawn from a corpus of spontaneous Dutch. Similarly, Bell et al. (2009) reports that English content and function words tend to be shorter the higher their lexical frequency and conditional probability (i.e. probability of a word given its preceding or following words). These studies suggest that more frequent and predictable linguistic materials tend to receive shorter and weaker pronunciations, a claim also known as the Probabilistic Reduction Hypothesis (Bell et al., 2003). The present work extends this line of research by investigating if probabilistic effects can be found in French on the duration of intervocalic [t] consonants in two different speech styles.

Predictability effects on articulation have been claimed to arise during the planning process in speech production (Bell et al., 2009). When speech planning is more difficult (e.g. because the words being planned are less probable), speakers may locally slow down their rate of articulation in order to preserve an undisrupted flow of speech. This hypothesis is supported by findings that predictability effects on content words are caused by the right, not the left, context (i.e. the following rather than the preceding word) (Pluymaekers et al., 2005; Bell et al., 2009).

If word frequency and predictability effects arise during speech planning, it is expected that probabilistic effects will not be restricted to languages for which they have been established so far. Furthermore, it is expected that probabilistic effects will be greater in unprepared spontaneous speech than in read and prepared speech. In order to investigate these issues, we examined a new language and analyzed materials extracted from corpora of casual and journalistic speech.

The reason for investigating French is that the studies cited above, along with nearly all others, have exclusively focused on Germanic languages. Importantly, Germanic and Romance languages differ in terms of their temporal structure. For instance, French and Spanish, which are said to be syllable-timed, exhibit less variability in syllabic durations within an utterance than Dutch, English and German (Grabe and Low, 2002). As a consequence, probabilistic effects on the duration of syllables and segments might be smaller in French than in Germanic languages.

In Study 1 we extracted data from the Nijmegen Corpus of Casual French (NCCFr from now on), which contains highly casual conversations among Parisian university students (Torreira et al., 2010, see Chapter 2). In Study 2, we used speech materi-

als extracted from the ESTER corpus of journalistic speech (Galliano et al., 2005), which mostly contains read news broadcasts from several French-speaking radio stations.

We selected [t] as our object of study, since it can be found in numerous different French words representing a wide range of frequencies. Moreover, [t] closure, the main acoustic correlate of the duration of French [t], can be accurately measured from a minimally clean acoustic signal. In the light of the findings mentioned above, [t] closures can be expected to display shorter durations when occurring in more frequent or predictable words.

## 3.2 Study 1

In Study 1, we investigate whether probabilistic effects on articulation can be found in casual French.

### 3.2.1 Method

Excerpts containing intervocalic [t] consonants were extracted from the NCCFr Torreira et al. (2010), which includes 35 hours of conversation among friends. All [t] consonants occurred within words of minimally two syllables, and were preceded and followed by one of the vowels /a/, /e/, /ɛ/, /o/, /ɔ/. Word-medial positions were favored over word-initial and word-final positions in order to avoid word-edge effects. Words with [t] surrounded by high vowels were excluded in order to avoid cases of palatalization, which might affect the realization of closures. Nasal vowels were disconsidered in order to avoid cases of nasality spreading into [t] closures, a situation that would make our measurements of oral closures less accurate. Only word types with at least 5 tokens were considered for analysis. For word types that complied with all these conditions, we randomly extracted a maximum of 15 tokens. The average number of tokens per word type was 8.24, with a standard deviation of 4.03.

All word types were content words. We decided to discard two semantically weak words (*était*, and *totalelement*) on the grounds that they might behave differently from the vast majority of word types Bell et al. (2009). Therefore, our materials exclusively consist of a homogeneous group of semantically strong content words. Examples of used word types include *arrêté* /aʁɛtɛ/ ‘stopped’, *gâteaux* /gato/ ‘cake’, *hôtel* /otɛl/ ‘hotel’ and *photo* /foto/ ‘photograph’. The few tokens that displayed a partial closure were discarded, and only [t] consonants displaying full closures were retained for analysis. In total, our dataset contained 29 distinct word types and 239 tokens spoken by 44 different speakers.

The duration of [t] closures was measured manually from speech waveforms. We provided a prosodic annotation to each of the extracted tokens including pitch accents, intonational boundaries and the number of syllables of the word containing [t]. Pitch accents were annotated whenever they occurred on the vowels preceding and following [t]. No distinctions were made between different sorts of pitch accents. However, it may be noted that all accents on vowels preceding [t] were initial accents found at the beginning of accentual groups (also called *accents d'intensité* in the French intonational literature) (Jun and Fougeron, 2002), while accents on vowels following [t] could be either group initial or group final. This distribution follows from the fact that all of our [t] consonants were word-medial, and that final accents must always be word-final. Intonational boundaries were marked whenever a boundary tone occurred between the test word and preceding or following words. No distinctions were made between different levels of prosodic phrasing. The vast majority of test words were not preceded by an intonational boundary (97%), but most of them were followed by one (74%). The number of syllables in the test word was calculated automatically from canonical pronunciations. Finally, speech rate was computed as the number of syllables per second within a stretch of speech containing the test word free of pauses or disfluencies.

Our probabilistic measures consisted of word and word bigram frequencies, both normalized per million words (for word  $i$ : frequencies of  $w_i$ ,  $C(w_{i-1}w_i)$  and  $C(w_iw_{i+1})$ ). Two sets of frequency measures were computed. The first set was computed using the transcriptions of the NCCFr corpus. Since this corpus may be judged rather small for the purpose of estimating word and word bigram frequencies (it only has around 550,000 word tokens), we also used probabilistic estimates computed jointly from the ESTER corpus and the PAROLE French Corpus, which both consist mostly of journalistic texts. These two journalistic corpora together contain over 15 million words. The first set of probabilistic measures is more appropriate for modeling casual speech (as it is based on casual speech data), while the second set may be more reliable given the bigger size of the corpus used for its estimation.

For the computation of word bigram frequencies, the different types of punctuation marks used in the transcriptions of the corpora were considered word types of their own. Word bigrams can therefore be composed of a test word and a punctuation mark. Whenever a test word was preceded or followed by a pause not marked by a punctuation mark, the corresponding word bigram frequency was left blank in our data matrix.

Finally, a check was performed in order to assess the reliability of our manual measurements and annotations. Eighty-nine tokens (15% of the dataset) were randomly selected, and independently analyzed by a trained transcriber unaware of the purposes of our study. No major disagreement was found between our measurements

and those of the assistant (97.2% of [t] closure duration measurements differed by 10 ms or less between the two annotators. Cohen's  $\kappa$  values of .94 and .67 were respectively obtained for the annotation of a preceding and a following intonational boundary).

### 3.2.2 Results

Mixed-effects linear regression with speaker and word type as random factors was used in order to estimate the effects of the different predictors on [t] closure duration. All probabilistic variables were logarithmically transformed, since their distributions had long right tails spreading over several orders of magnitude, and since language users are known to be sensitive to logarithmic values rather than raw frequencies. For ease of exposition, we do not continuously refer to these variables as log-transformed variables and rather use their original names (e.g. word frequency instead of log word frequency). We first fitted a control model with speech rate and the prosodic factors present in our annotation, and then added separately each of the two sets of probabilistic estimates. Whenever two predictors, say A and B, were significantly correlated, we orthogonalized them by replacing variable A in the regression model with the residuals of a linear model in which A was predicted by B. We only report the results of models which have been trimmed of outliers (i.e. of data points with values beyond two residual standard errors of the values predicted by the model).

#### Control model

Unsurprisingly, speech rate was a statistically significant predictor of [t] closure duration ( $\beta = -4.25, t = -5.92, p < .0001$ ). From the set of prosodic factors considered in the annotation, only the presence of an accent on the vowel preceding [t] and the number of syllables in the word yielded statistically significant effects (preceding accent:  $\beta = 8.63, t = 2.56, p < .05$ ; number of syllables.:  $\beta = -5.3, t = -3.04, p < .005$ ). All other factors yielded regression coefficients in the expected direction, but none of them approached statistical significance. Only speech rate, the presence of a preceding accent and the number of syllables in the test word were retained in the model, to which we then added the probabilistic predictors.

#### Probabilistic factors

From the set of probabilistic factors estimated from the NCCFr, only word bigram frequency  $C(w_i w_{i+1})$ , that is the joint frequency of the test word and its following word, approached statistical significance ( $\beta = -2.06, t = -1.81, p = .07$ ): there was a tendency for [t] closures to be shorter when occurring in more frequent word

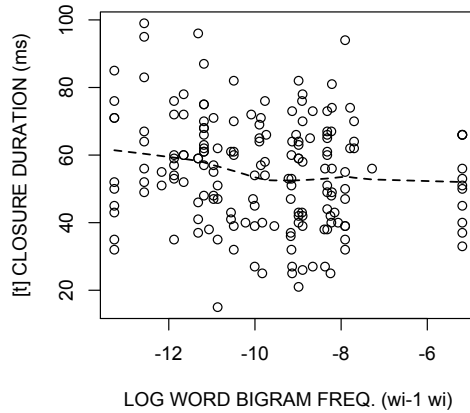


Figure 3.1: *Study 1: duration of [t] closure as a function of log word bigram frequency  $C(w_{i-1}w_i)$  as estimated from journalistic sources. The dashed line represents a lowess regression curve.*

bigrams. Word frequency and word bigram frequency  $C(w_{i-1}w_i)$  did not appear to affect [t] closure duration (word frequency:  $\beta = -0.96, t = -0.6, p = .54$ ; word bigram frequency  $C(w_iw_{i+1})$ :  $\beta = -0.5, t = -0.54, p = .58$ ).

From the set of probabilistic factors estimated from the journalistic corpora, only word bigram frequency  $C(w_{i-1}w_i)$ , that is the joint frequency of the test word and its *preceding* word, reached statistical significance ( $\beta = -1.19, t = -2.13, p < .05$ ): the duration of [t] closures tended to be shorter the more frequent the word bigram. This tendency is illustrated in Figure 3.1. Word frequency and word bigram frequency  $C(w_iw_{i+1})$  did not yield statistically significant effects on [t] closure duration (word frequency:  $\beta = -1.07, t = -1.63, p = .1$ ; word bigram frequency  $C(w_iw_{i+1})$ :  $\beta = -0.32, t = -0.4, p = .68$ ).

### 3.2.3 Discussion

In Study 1, effects of word bigram frequency on the duration of [t] closures were identified. These results show that a syllable-timed language like French may also exhibit probabilistic effects on articulation. Interestingly, different results were obtained depending on which probabilistic estimates were used. An effect of word bigram frequency  $C(w_{i-1}w_i)$  was identified when the probabilistic estimates were computed from a large corpus of journalistic texts. On the other hand, when we used

probabilistic estimates computed from the NCCFr, a marginal effect of word bigram frequency  $C(w_i w_{i+1})$  was found. It may be argued that the journalistic speech estimates used in Study 1 are not appropriate for modeling casual pronunciations and that the effect of word bigram frequency  $C(w_{i-1} w_i)$  is not genuine, but rather arises from an underlying correlation between word bigram frequency  $C(w_{i-1} w_i)$  and word bigram frequency  $C(w_i w_{i+1})$ . The reason why the latter did not yield a statistical effect might be attributed to the fact that it was poorly estimated. Another possibility is that both word bigram effects are genuine.

### 3.3 Study 2

In Study 2, we investigated whether probabilistic effects could also be found in journalistic speech, where speech planning may be less difficult and affect articulation less.

#### 3.3.1 Method

The methodology used in Study 2 was similar to that used in Study 1, except for that the materials were extracted from the ESTER corpus of journalistic speech. Possible word types complied with the criteria exposed in Section 3.2.1. Consequently, the dataset consisted of semantically strong content words with [t] consonants in word-medial position surrounded by non-high oral unrounded vowels. As in Study 1, only word types with at least 5 tokens were considered for analysis, and a maximum of 15 tokens was used for each word type. Examples of used word types include *atomique* /atɔ̃mik/ ‘atomic’, *attaque* /atak/ ‘attack’, *hôtel* /otɛl/ ‘hotel’ and *bateau* /bato/ ‘ship’. The resulting dataset consisted of 63 word types and 592 tokens spoken by 376 speakers.

The annotation procedure was identical to that used in Study 1. The probabilistic measures were estimated only from the journalistic PAROLE and ESTER corpora, which were deemed to be appropriate in terms of register and size. We decided not to use frequency measures computed from the NCCFr, since this corpus was relatively small and inappropriate in terms of register. As in Study 1, the phrasal position of the test words was not balanced. Most of the word tokens (72%) were followed by an intonational boundary, while only a few of them were preceded by one (6%). Again, we performed a reliability check on 15% of the tokens with the help of a second annotator. Inter-annotator agreement was satisfactory (95.4% of [t] closure duration measurements differed by 10 ms or less between the two annotators. Cohen’s  $\kappa$  values of .88 and .57 were respectively obtained for the annotation of a preceding and a following intonational boundary).

Table 3.1: *Control model in Study 2: Regression coefficients along with  $t$  and  $p$  values.*

Factor	$\hat{\beta}$	$t$	$p$
Speech rate	-4.58	-7.77	< .0001
Fol. accent	12.67	6.57	< .0001
Fol. inton. bound.	9.31	5.32	< .0001
Prec. accent	4.89	2.23	< .05
Number of syl.	-4.54	-3.25	< .005

### 3.3.2 Results

As in Study 1, mixed-effects linear regression with speaker and word type as random factors was employed to identify statistically significant predictors of [t] closure duration and estimate their effects. Again, we first fitted a control model with speech rate and prosodic factors, and then added the probabilistic measures to this model.

#### Control model

Table 3.1 summarizes the regression coefficients,  $t$  values and  $p$  values of the statistically significant predictors. All control factors, except the presence of a preceding intonational boundary, yielded statistically significant effects in the expected direction. Speech rate and the number of syllables in the word were negatively correlated with closure duration. On the other hand, the presence of neighboring accents and of a following intonational boundary tended to increase closure duration.

#### Probabilistic factors

Word bigram frequency  $C(w_i w_{i+1})$  yielded a statistically significant effect ( $\beta = -1.12$ ,  $t = -2.57$ ,  $p < .05$ ). Along the lines of the Probabilistic Reduction Hypothesis, [t] closures tended to be shorter the more frequent the word bigram (see Figure 3.2). On the other hand, word frequency and word bigram frequency  $C(w_{i-1} w_i)$  did not have a statistically significant effect on [t] closure duration (word frequency:  $\beta = -0.84$ ,  $t = -0.81$ ,  $p = .41$ ; word bigram frequency  $C(w_{i-1} w_i)$ :  $\beta = -0.27$ ,  $t = -0.63$ ,  $p = .52$ ).

Since word bigram frequency can be expected to be lower across stronger prosodic boundaries, there is a possibility that the effect of word bigram frequency  $C(w_i w_{i+1})$  is in fact a prosodic effect, since in our annotation we only coded one level of intonational phrasing. In order to address this issue, we tested the effect of word bigram frequency  $C(w_i w_{i+1})$  on the subset of test words not followed by an intonational boundary. The effect approached statistical significance ( $\beta = -1.04$ ,  $t = -1.81$ ,  $p = .07$ ). The fact that the regression coefficient obtained for word bigram frequency  $C(w_i w_{i+1})$



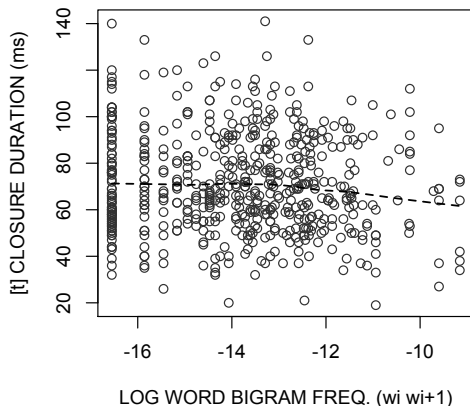


Figure 3.2: *Study 2: duration of [t] closure as a function of log word bigram frequency  $C(w_i w_{i+1})$ . The dashed line represents a lowess regression curve.*

in this model remained similar to that obtained using the full dataset suggests that the loss of statistical significance may be due to the reduced sample size ( $n = 140$ ; 23% of the full dataset).

### 3.4 General discussion and conclusions

This study investigated whether a word's predictability affects the duration of its [t] consonants in connected French. In Study 1, which used casual speech materials, an effect of word bigram frequency  $C(w_{i-1} w_i)$  was identified when the probabilistic estimates were computed from a large corpus of journalistic texts. When we used probabilistic estimates computed from casual speech materials, a marginal effect of word bigram frequency  $C(w_i w_{i+1})$  was found. In Study 2, which used journalistic speech materials and probabilistic estimates, we found an effect of word bigram frequency  $C(w_i w_{i+1})$ . In all cases, the duration of [t] closures tended to be shorter the higher the word bigram frequency. We conclude that French, a syllable-timed language, is also subject to the sort of probabilistic effects on articulation which have been identified for Germanic languages (Pluymaekers et al., 2005; Bell et al., 2003, 2009), which are said to be stress-timed.

Study 1 shows the importance of using appropriate corpora to estimate the probabilities of linguistic materials. This poses a serious challenge to researchers investi-

gating casual speech, since for most languages no sufficiently large corpora of casual speech are available from which robust frequency estimates can be computed. Using inappropriate probabilistic measures may lead to erroneous or conflicting results.

Even though in Study 1 different results were obtained depending on which probabilistic measures were used (i.e. based on casual or journalistic speech), a consistent picture emerges if we only consider the results obtained with frequency estimates computed from corpora matched to the data in terms of speech style. In this case, we find that the duration of [t] closures in Studies 1 and 2 is sensitive to word bigram frequency  $C(w_i w_{i+1})$ , replicating previous findings that predictability effects for content words are essentially caused by the upcoming context (Pluymaekers et al., 2005; Bell et al., 2009).

This interpretation supports the view that probabilistic effects are mainly due to speech planning constraints. To this respect, the results of Study 2 need some comment. Since our journalistic speech data mostly consisted of read texts, one may wonder to what extent the word bigram frequency effect found in Study 2 may be due to planning constraints. A possible explanation is that, even if planning constraints play a lesser role in read than in casual speech, the effect is strong enough to manifest itself in the speech of radio news readers.

Another possibility is that the probabilistic effect in Study 2 is not solely due to planning constraints, but also to adaptation on the part of news readers to their listeners' needs. Less frequent word sequences tend to be more informative from the point of view of the listener, and may therefore receive longer articulations. Although on-line adjustments of this sort are considered to be computationally complex and to take place slowly and imperfectly in connected speech (Bard et al., 2000), it is not inconceivable that professional news readers are able to successfully perform such adjustments. More research is needed to identify the sources of probabilistic effects on articulation in different speech registers.

## Vowel elision in casual French: the case of vowel /e/ in the word *c'était*

---

### Chapter 4

This chapter is a reformatted version of:  
Francisco Torreira and Mirjam Ernestus (2011). Vowel elision in casual French: the case of vowel /e/ in the word *c'était*. *Journal of Phonetics*, 39(1):50–58.

This study investigates the reduction of vowel /e/ in the French word *c'était* /setɛ/ 'it was'. This reduction phenomenon appeared to be highly frequent, as more than half of the occurrences of this word in a corpus of casual French contained few or no acoustic traces of a vowel between [s] and [t]. All our durational analyses clearly supported a categorical absence of vowel /e/ in a subset of *c'était* tokens. This interpretation was also supported by our finding that the occurrence of complete elision and [e] duration in non-elision tokens were conditioned by different factors. However, spectral measures were consistent with the possibility that a highly reduced /e/ vowel is still present in elision tokens in spite of the durational evidence for categorical elision. We discuss how these findings can be reconciled, and conclude that acoustic analysis of uncontrolled materials can provide valuable information about the mechanisms underlying reduction phenomena in casual speech.

## 4.1 Introduction

One of the main characteristics of spontaneous connected speech is the ubiquitous presence of reduced pronunciations. For a corpus of spontaneous American English, Johnson (2004) reports absences of segments in at least 20% of the word tokens, and deviations from canonical citation forms involving some sort of reduction for 60% of all word tokens (e.g. [p<sup>h</sup>ɛɹ̩] for *apparently*; [k<sup>h</sup>z] for *because*, [p<sup>h</sup>t<sup>h</sup>k<sup>h</sup>əʋ] for *particular*). Although the pervasiveness of such reductions in everyday speech is now widely acknowledged (e.g. Ernestus, 2000; Kohler, 2000; Shockey, 2003; Johnson, 2004), little is known about how they are produced and about their exact phonetic characteristics. In fact, detailed descriptions of lenition phenomena in conversational speech are scarce, especially for non-Germanic languages such as French (excepting perhaps well-studied phenomena as schwa deletion (Bürki et al., 2007) and voicing assimilation in consonant clusters (Duez, 1995; Hallé and Adda-Decker, 2007), which for many researchers might not fall under the hood of casual speech processes).

This article contributes to filling this gap. We investigate the phonetic characteristics and conditioning factors of a reduction phenomenon affecting vowel /e/ in the word *c'était* /setɛ/ 'it was' in casual French. This reduction process, despite its high incidence in conversational French, has not received any attention in the phonetic and phonological literatures, and therefore offers a fresh ground for investigation.

Of particular interest to us is whether the elision of vowel /e/ results from the categorical absence of this segment in a reduced phonological form of the word *c'était* /stɛ/, or if it can be explained as a case of extreme gradient reduction. Browman and Goldstein (1990) made the hypothesis that casual speech reduction phenomena are mainly due to two gradient processes: reduction in gestural magnitude and increase in gestural overlap. Reduction in gestural magnitude occurs for instance when a consonantal constriction is not fully achieved, as when the word *sugar* is produced with a velar approximant rather than with a stop. This mechanism can be thought to underlie most undershoot phenomena. Gestural overlap involves the timing between articulatory movements, and may lead to different kinds of reduction including apparent segmental assimilations and deletions. For instance, the articulatory gestures in consonant clusters consisting of coronal and bilabial consonants (e.g. /tm/ in the utterance *perfect memory* or /dp/ in *hundred pounds*) are sometimes overlapped in such a way that the coronal constriction is acoustically masked by the bilabial closure, even if it is effectively achieved by the speaker.

The view that reduction phenomena in casual speech are mainly gradient has been supported by numerous studies. Davidson (2006) showed that pretonic schwa deletion in English (as in the words *potato* or *support*), which had been traditionally

analyzed as the output of a phonological deletion rule (e.g. Zwicky, 1972), is the endpoint of a phonetic reduction continuum. Davidson's conclusions were based on the acoustic detail of segmental sequences with and without acoustically realized schwas, and of underlying consonant clusters lacking schwas (e.g. as in the word *sport*). This study showed, among other things, that voiceless stops such as /p/ in the word *support* /səpɔrt/ maintain their characteristic aspiration when following [s] in cases of schwa elision, in contrast with underlying /sp/ clusters, which notably lack aspiration after the [p] release. Other connected speech phenomena now claimed to be gradient rather than categorical include nasal place assimilation in English (Nolan (1992); Byrd (1996), but see Ellis and Hardcastle (2002) for evidence of categorical assimilation in some speakers), palatalization of /s/ before /j/ in English (Zsiga, 1995), and progressive voice assimilation in German (Kuzla et al., 2007).

Although many reduction phenomena are now classified as gradient, it is uncontroversial that some reduced forms typical of casual connected speech such as English *gonna* (standard form: *going to*) and *wanna* (standard form: *want to*) do not result from online articulatory reduction. These cases can be viewed as fossilized versions of extreme gradient reductions in frequent words. Numerous lexicalized reduced forms can be found in other languages as well (e.g. *tuurlijk* for *natuurlijk* 'of course' in Dutch, *pa* for *para* 'for' in Spanish, *cê* for *você* 'you' in Brazilian Portuguese). Moreover, some sandhi and elision processes have been claimed to be categorical rather than gradient (e.g. schwa elision in French, Bürki et al. (2007, 2010); voicing assimilation in French, Hallé and Adda-Decker (2007); nasal place assimilation in Spanish and Italian, Honorof (1999); Farnetani and Busà (1994a,b); /s/ to /ʃ/ accommodation in English, Nolan et al. (1996); place assimilation in Korean, Kochetov and Pouplier (2008), among others). There is therefore evidence that connected speech reduction phenomena may be both gradient and categorical.

According to traditional descriptions of French, in this language only schwa vowels can be elided categorically (Tranel, 1987; Coveney, 2001; Fagyal et al., 2006). Therefore, complete elision of [e] in the word *c'était* can be expected to occupy the endpoint of a gradient reduction continuum. However, *c'était* is a very recurrent expression in everyday speech, similar in this respect to English expressions with phonologically reduced variants such as *going to* and *want to*. For this reason, the possibility that the word *c'était* has a phonologically reduced variant without vowel /e/ should not be discarded a priori, even if the vowel subject to elision is not a schwa.

We investigate the nature of vowel elision in the word *c'était* using acoustic data from a corpus of casual speech. From a methodological perspective, we are interested in probing to what extent this sort of question, mostly addressed using laboratory speech materials and articulatory techniques (e.g. Davidson (2006), Manuel (1992), Fougeron and Steriade (1997), Barnes and Kavitskaya (2002), Côté and Mor-

rierson (2007)), can be investigated on the basis of more naturalistic data. Although carefully controlled materials and articulatory instrumentation offer obvious advantages for the investigation of finely grained phonetic phenomena, we suspect that many reduction phenomena characteristic of casual speech, such as the elision of vowel [e] in the word *c'était*, cannot be properly studied in a highly controlled laboratory setting. Let us assume for a moment that vowel [e] in the French word *c'était* can be elided categorically *only in extremely relaxed and casual speech*. Under this hypothesis, which seems quite plausible to us, a controlled experiment using invasive instrumentation would probably inhibit speakers from adopting the speech register in which categorical elision is allowed to occur, and would lead to the wrong conclusion that [e] elision is gradient.

## 4.2 Realizations of *c'était* in connected speech

In this section we present several pronunciations of the word *c'était* illustrating the reduction phenomenon studied in this article. The examples in Figures 4.1-4.3 were extracted from the Nijmegen Corpus of Casual French (Torreira et al., 2010, see Chapter 2), and depict occurrences of the word *c'était* in non-prepausal position. The vocalic portion observable at the beginning of each example belongs to a word preceding *c'était*. This vocalic portion was included in the examples so that the realization of the word *c'était* can be fully appreciated from its beginning. Note also that all waveforms and spectrograms are drawn to scale in the time dimension.

Figure 4.1 shows two cases in which the word *c'était* received unreduced realizations containing clear [e] vocalic features between [s] and [t] (i.e. voicing, formant structure). Figure 4.2 shows realizations more reduced than those in Figure 4.1. Example 4.2a shows a pronunciation of *c'était* in which a few low-amplitude voicing periods attributable to /e/ can be observed only at the beginning of the stop closure. Resonances in the F2 and F3 regions appear to be reinforced before the closure and the onset of voicing (see arrows in the figure), suggesting that an opening of the vocal tract occurred between [s] and [t]. Example 4.2b does not display any periodicity attributable to vowel /e/, but it does exhibit an increase in energy in the F2 region before the oral closure (see arrow in the figure), indicating that, as in Example 4.2a, the speaker may have produced a vocal opening gesture between [s] and [t].

Figure 4.3 illustrates cases in which the word *c'était* is further reduced. Example 4.3a does not display any trace of periodicity or formant structure attributable to /e/. Moreover, the duration of the voiceless frication portion is considerably shorter than that of Examples 4.2a and 4.2b above, making the presence of a reduced devoiced vowel between [s] and [t] rather unlikely. Example 4.3b exhibits additional reduction in the form of an incomplete [t] closure. In cases such as these, in which the main

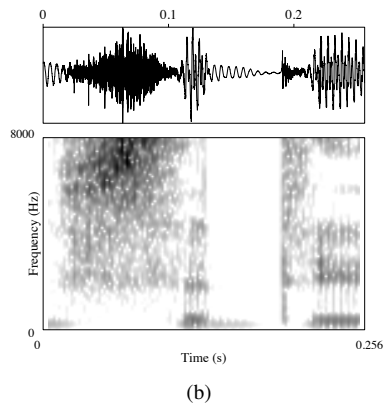
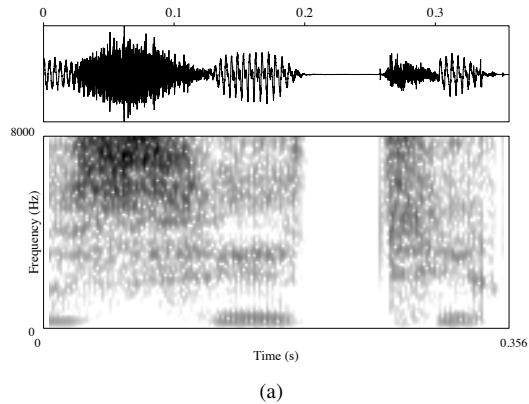


Figure 4.1: Waveforms and spectrograms of unreduced pronunciations of the word *c'était* /setɛ/.

spectral and durational correlates of /e/ are absent from the signal, it is legitimate to wonder if the speaker attempted to pronounce a vowel at all. On the other hand, the fact that cases of intermediate reduction such as those in Figure 4.2 also occur in casual French suggests that those in Figure 4.3 might simply occupy the endpoint of an articulatory reduction continuum.

### 4.3 Quantitative study

We now investigate the reduction of vowel /e/ in the word *c'était* /setɛ/ using quantitative methods. The study consists of two parts. First, we examine the distributions

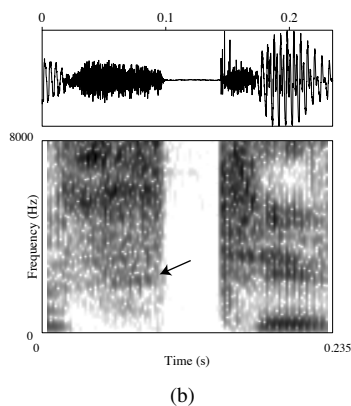
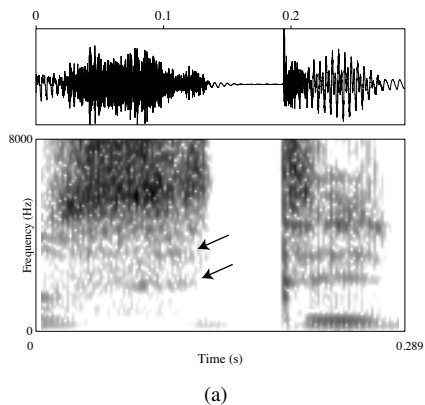


Figure 4.2: Examples of reduced realizations of the word *c'était* /setɛ/. Arrows indicate traces of vowel /e/ between [s] and [t] in the form of F2 and F3 resonances.

of several acoustic measures taken in the region of the word *c'était* where /e/ is expected to manifest itself. We also compare some of these measures between tokens of reduced *c'était* and words beginning with underlying /st/ clusters. In the second part of the study, we investigate by means of multiple regression which among a series of prosodic, speech rate and segmental factors condition the occurrence of elision and the duration of [e] in pronunciations of the word *c'était*. Our focus is on data and analyses allowing us to determine whether /e/ elision in the word *c'était* is a gradient or a categorical phenomenon.



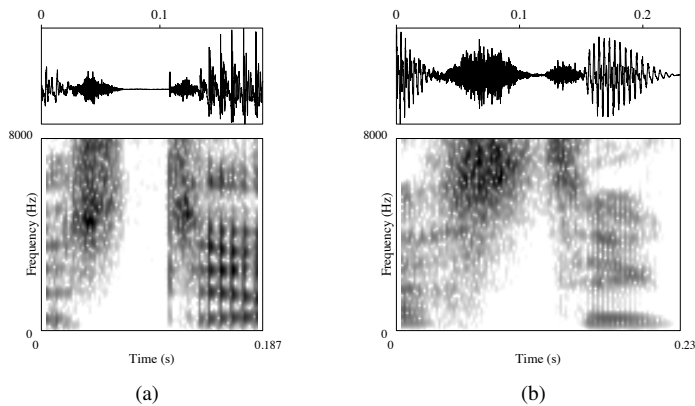


Figure 4.3: Examples of extremely reduced pronunciations of the word *c'était* /setε/.

### 4.3.1 Method

#### The Nijmegen Corpus of Casual French

All materials were extracted from the Nijmegen Corpus of Casual French (NCCFr). This corpus contains 35 hours of conversations featuring 46 French speakers (24 female, 22 male). At the time of the recording, all speakers were living in the Paris region. While most of them had been born in Île-de-France, some came from other regions in Central and Northern France. Except for two speakers, all had ages between 18 and 28. The recording equipment consisted of an Edirol R-09 solid-state stereo recorder, Samson QV head-mounted unidirectional microphones and a stereo microphone preamplifier. Microphones were placed at an approximate distance of 5 cm from the left corner of the speakers' lips. The sampling rate used was 48 KHz, while quantization was set to 32 bits. A detailed description of the preparation and recording of the NCCFr corpus can be found in (Torreira et al., 2010, see Chapter 2).

#### Selection of materials

Excerpts containing the word *c'était* were randomly extracted from the NCCFr. We considered materials from all speakers excepting the two who were considerably older than the rest (i.e. 52 and 55 years old) on the grounds that they might behave differently. Excerpts in which *c'était* carried a pitch accent were disconsidered, since their low number did not allow for statistical modeling. For the same reason, we excluded tokens in which intervocalic /s/ was voiced, or in which speakers failed to achieve a full [t] closure after clearly pronouncing [e]. Excerpts in which *c'était*

was immediately preceded by a voiceless obstruent were also discarded, since the beginning of [s], an important measure in our analyses, could not be determined reliably in these cases. Finally, excerpts in which *c'était* was part of an accentual phrase containing disfluencies, laughter or intrusive overlapping speech were also excluded. The resulting dataset contained 450 tokens from 44 different speakers.

In order to compare [st] clusters in reduced *c'était* pronunciations (as in Examples 4.2b and 4.3a) with underlying /st/ clusters, we extracted all tokens of words beginning with underlying /st/ clusters followed by a non-high vowel. Word types with high vowels after /st/ were discarded because /t/ was often severely palatalized in this context. Examples of investigated word types include *stage* 'internship', *station* 'station', *stérile* 'sterile' and *stéréotype* 'stereotype'. After discarding tokens occurring in sentences with disfluencies and laughter, those preceded by a voiceless obstruent, and tokens spoken by the two oldest speakers, a total of 104 underlying /st/ clusters from 15 word types remained for analysis.

## Measurements

The acoustic measurements taken include [e] and [s] durations, the duration of the [t] closure, the duration of voicing during the [t] closure and the spectral center of gravity in the second half of the interval marked as [s]. All measurements were done manually unless indicated otherwise.

It should be noted that the durational intervals defined below, particularly [e] and [s], do not correspond strictly to traditional segmental intervals, but rather to intervals delimited by specific acoustic landmarks attributable to the articulation of specific phonological segments. These landmarks were chosen on the basis of a clear association with the articulation of /s/, /e/ and /t/ in uncontrolled casual speech (i.e. onset of high-frequency noise as a landmark of /s/, voicing onset as a landmark of vowel /e/, start of stop closure as a landmark of /t/). For instance, in the case of [s] to [e] transitions, sudden and instantaneous increases in formant amplitudes indicative of a vocal tract opening, which can provide a reasonably consistent landmark in laboratory speech recordings, could not be used reliably in our materials. Sudden and simultaneous increases in formant amplitude as in the [se] transition of Figure 4.1b were rare in our dataset.

*Duration of [s]:* The beginning of [s] was marked at the onset of aperiodic energy in the high frequencies (> 4 KHz), while its end was defined as the onset of periodicity attributable to vowel /e/. In tokens lacking any periodicity between [s] and [t], the end of [s] was marked at the beginning of the [t] closure. Tokens lacking both a portion of [e] periodicity and a complete [t] closure did not receive an [s] duration value. According to this definition, the interval marked as [s] may contain a voiceless vocalic part.

*Duration of [e]:* The duration of [e] was defined as the interval extending from the onset of periodicity after the aperiodic portion of [s] up to the start of the [t] closure. We refer to this measurement as [e] duration rather than [e] periodicity only for the sake of simplicity. Due to this definition, the initial part of this interval may still contain frication attributable to /s/. Also, cases with an [e] duration of 0 ms may actually contain a completely devoiced vowel, which in our measurement scheme is included in the [s] interval.

*Duration of [s(e)]:* The duration of [s(e)] includes the interval extending from the onset of [s] frication to the start of the [t] closure, regardless of whether periodicity is present within this interval. Tokens without a full [t] closure were not assigned any [s(e)] duration value, since the presence of a full [t] closure was needed to determine the end of the [s(e)] interval.

*Duration of [t] closure:* The start of the [t] closure was marked at the offset of energy in the F2-F3 region, while the end was marked at the beginning of [t] release. Tokens without a clear [t] closure (e.g. Figure 4.3b) did not receive a [t] closure duration value.

*Intrusive voicing:* A considerable number of [t] closures had a portion of periodicity attributable to the preceding /e/ segment. We refer to this portion of periodicity as intrusive voicing. Since the end of periodicity during the closure was often difficult to determine manually due to a very gradual decay, its duration was measured automatically by using the auto-correlation method for pitch detection available in Praat, with default settings except for time step (5 ms), silence threshold (0.03) and voicing threshold (0.04). The details of this algorithm can be found in Boersma (1993).

*Spectral balance in the second half of [s]:* For underlying /st/ clusters and tokens of *c'était* without voicing between [s] and [t], we calculated the spectral center of gravity in the last part of the [s] interval. Speech signals were low-pass filtered (10 KHz), and then an FFT spectrum was computed from a Hamming window placed over the second half of the [s] interval (between the center of [s] and the beginning of the [t] closure). This measure should identify gross spectral differences between the underlying and elision [st] clusters in the region where acoustic traces of a reduced/devoiced vowel might be present. If elision clusters are produced with a relaxation or opening of the [s] constriction before the closing gesture of [t], or a more posterior articulation of [s] (resulting from coarticulation with vowel /e/), a downward shift in spectral balance should be observed.

### **Contextual annotation**

Each token was provided with an annotation of several prosodic characteristics and of the preceding segmental context. The main use of this annotation was to provide a

Phrase-medial	
... <i>parce que c'était un aquarium.</i>	'because it was an aquarium'
... <i>et c'était assez.</i>	'and it was enough'
... <i>mais c'était pas vrai.</i>	'but it was not true'
AP-initial	
<i>moi aussi, c'était comme ça.</i>	'me too it was like that'
<i>plus exactement, c'était comment?</i>	'more exactly, how was it?'
<i>en plus, c'était la première fois.</i>	'on top of that, it was the first time'
IP-initial	
<i>Si. C'était marrant.</i>	'Yes. It was funny.'
<i>Non. C'était pendant le PACS.</i>	'No. It was during the PACS.'
[pause] <i>C'était foutu pour eux.</i>	[pause] 'They were screwed.'

Table 4.1: Examples of *c'était* in each annotated phrasal position.

pool of potential predictors of [e] reduction that could be used in the regression part of this study (Section 4.3.3).

*Prosodic annotation:* Phrasal prosody is known to affect articulation and therefore the occurrence of reduction phenomena. Most importantly, phrase-initial segments tend to be articulated with increased articulatory strength (e.g. Fougeron, 2001; Keating et al., 2003; Cho and McQueen, 2005). For this reason, we marked whether *c'était* was preceded by a major prosodic boundary or if it was in phrase-medial position (note that *c'était* tokens in phrase-final position were not included in the dataset). Two different prosodic boundaries were distinguished: Accentual Phrase (AP) boundaries were marked after continuation rises, while Intonational Phrase (IP) boundaries were placed after final falls (typically in sentences of declarative modality, but also in some questions) or final rises in certain questions (typically yes-no questions). Table 4.1 illustrates utterances representative of each phrasal position type. The dataset contained a total of 71 tokens in phrase-medial position, 161 tokens in AP-initial position and 218 tokens in IP-initial position.

The beginning and end of the AP containing *c'était* were marked following the same criteria. The number of canonical syllables (referred to simply as syllables from now on) in each AP was also annotated, as well as the number of syllables separating the first syllable of *c'était* from the beginning and end of its AP. AP length ranged from 3 to 11 syllables, with an average of 4.97 syllables. Long APs occurred when speakers did not place intonational boundaries in the usual locations (i.e. at major syntactic boundaries), a phenomenon that typically occurs at fast speech rates (Fougeron and Jun, 1998).

*Segmental context:* We annotated whether the word *c'était* was immediately preceded by a consonant, a vowel or a silent pause.

### Reliability check

The manual measurements and annotations mentioned above were all made by the first author. A check was performed in order to assess the reliability of these measurements and annotations. One hundred *c'était* tokens (21.9% of the dataset) were randomly selected, and independently analyzed by an assistant unaware of the purposes of our study. We computed the mean differences and correlations between the continuous measurements mentioned above ([s], [e] and [t] closure durations and the number of syllables in the AP). For the other annotations (the location of AP edges, the presence and type of major prosodic breaks before *c'était*, the presence of a complete [t] closure in *c'était*), we computed the percentages of agreement between the two annotators.

No major discrepancies were found between the two annotators. Duration measurements were generally highly correlated ( $r = .96$  for [e] duration,  $r = .92$  for [s] duration,  $r = .88$  for closure duration,  $r = .78$  for the number of syllables in the AP) and their mean absolute differences were always below 15% of the mean of the variable being checked (3.6 ms for [e] duration, 6.1 ms for [s] duration, 4.4 ms for closure duration, 0.67 for the number of syllables in the AP). The annotation of phrasal position was highly consistent between the two annotators (93%), as was the annotation of complete [t] closures (99%). The annotation of the beginning and end of the AP showed an agreement of 82%.

### 4.3.2 Results: Distribution of acoustic characteristics

#### Duration of [e]

We first examined the distribution of [e] duration in the dataset. As explained above, we defined this duration as the interval extending from the onset of voicing after the voiceless [s] segment to the beginning of the [t] closure. It should be noted that all cases without a complete [t] closure ( $n = 95$ , 21.1% of the total dataset) received an [e] duration value of 0 ms, since none of them exhibited any voicing between the beginning of [s] and the onset of voicing in the vowel following the reduced [t].

Figure 4.4 shows a kernel density plot of [e] durations. Interestingly, 62% of the data points ( $n = 279$ ) did not have any voicing between [s] and [t]. The rest of the data points ( $n = 171$ ) exhibit positive values around a mean of 37 ms. Inspection of tokens with low [e] duration values ( $< 10$  ms) revealed that many of them had a considerable stretch of voicing extending into the [t] closure (defined and measured as intrusive voicing; see Section 4.3.1). Figure 4.2a illustrates this phenomenon. For

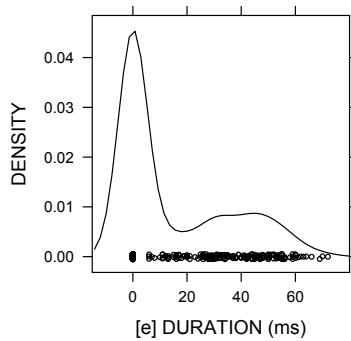


Figure 4.4: Kernel density plot of [e] duration. Circles represent individual data points.

the sake of simplicity, we henceforth refer to tokens with 0 ms of periodicity between [s] and [t] as *elision* cases, and to tokens with positive periodicity in this interval as *non-elision* cases.

### Duration of [s(e)]

We then examined the distribution of [s(e)] duration, which we defined as the interval extending from the onset of [s] frication to the start of the [t] closure, regardless of whether it included any periodicity. If elision results from the absence of a complete segmental slot corresponding to vowel /e/, a bimodal distribution of [s(e)] duration should be found, and elision tokens should have shorter durations than non-elision tokens. On the other hand, if tokens with [e] duration of 0 ms actually contain a devoiced [e] vowel resulting from gradient reduction, the distribution of [s(e)] durations is expected to be unimodal.

The distribution of [s(e)] duration (solid line in Figure 4.5) exhibits an asymmetrical pattern, with a clear mode between 100 and 150 ms, and a salient bump between 50 and 100 ms. Interestingly, the main mode and the bump correspond closely to the modes of the distributions of the non-elision and elision groups (see dotted and dashed lines in Figure 4.5), suggesting that the asymmetrical pattern of [s(e)] duration is due to the pooling of two different durational populations (elision and non-elision groups). In our view, this distribution is consistent with the categorical absence of a temporal slot corresponding to /e/.

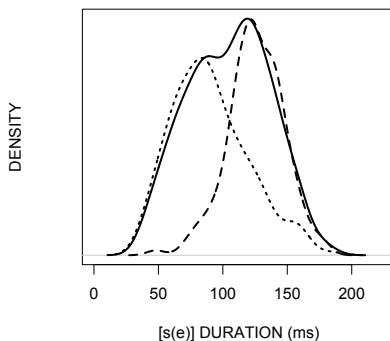


Figure 4.5: Kernel density plots of [s(e)] duration for pooled data (solid line), elision cases (dotted line) and non-elision cases (dashed line).

### Duration of [s] and [t]

Previous work on segmental durations in English and Dutch has shown that consonants tend to be shorter when part of a consonant cluster (Klatt, 1974; Crystal and House, 1988; Waals, 1999). The same durational pattern may also hold for French. We therefore wondered if [s] and [t] closures in elision tokens were shorter than in non-elision tokens. If so, the difference might be taken as evidence that elision *c'était* tokens contain genuine /st/ clusters.

Mixed-effects linear models (Bates and Sarkar, 2006) were fitted with speaker as a random factor, elision as the predictor and [s] and [t] closure durations (in ms) as dependent variables. Tokens without a full [t] closure ( $n = 95$ ), all of them elision cases, were excluded from this comparison, since they did not offer measurable [s] and [t] closure intervals. The resulting dataset contained 184 elision and 171 non-elision tokens. Results indicated that both [s] and [t] closure durations were slightly shorter in elision than in non-elision cases ([s]:  $\beta = -4.97, t = -2.16, p < .05$ ; [t] closure:  $\beta = -4.69, t = -3.16, p < .005$ ), supporting the possibility that in elision cases they are genuine /st/ clusters, rather than apparent /st/ clusters containing an extremely reduced /e/ vowel. Importantly, similar results were obtained when several covariates including speech rate (see Section 4.3.3 below for details) were included in the regression models.

### Comparison of elision and underlying /st/ clusters

In order to further test the hypothesis that vowel /e/ was categorically absent in a subset of pronunciations of the word *c'était*, we compared [st] clusters resulting from [e] elision in *c'était* with underlying /st/ clusters in word-initial position. Three acoustic parameters were examined: [s] duration (in ms), [t] closure duration (in ms) and the spectral center of gravity (in Hz) computed in a Hamming window placed over the second half of [s] (where traces of [e] in *c'était* might be present). Mixed-effects regression with speaker as random factor was used in order to test for statistical differences between the two groups. In the case of spectral center of gravity, the gender of the speaker was used as a covariate.

Only tokens with a complete [t] closure were considered for analysis, since [s] and [t] closure durations were not available for tokens with an incomplete closure. The resulting dataset contained 184 *c'était* elision tokens and 73 tokens of words containing underlying /st/ clusters. Since underlying /st/ clusters were never preceded by a major prosodic break (because they always occurred in content words preceded by articles or pronouns), we controlled for the preceding prosodic context in the following way. We first checked if the preceding prosodic context had a statistical effect on each of our dependent variables in the subset of *c'était* tokens. Only [s] duration was affected by the preceding prosodic context ( $F(2, 181) = 5.8, p < .005$ ). For the analysis of [s] duration, therefore, only *c'était* tokens not preceded by a prosodic boundary were retained for comparison with underlying /st/ clusters. For [t] closure duration and the spectral center of gravity, we concluded that the preceding prosodic context was not a confounding factor and proceeded to an analysis of all available data.

Neither [s] nor closure duration differed between underlying clusters and clusters resulting from [e] elision ([s] duration:  $\beta = -3.07, t = -0.69, p > .1$ ; closure duration:  $\beta = -0.001, t = -0.001, p > .1$ ). On the other hand, we found a marginal effect of cluster type on the spectral center of gravity of the second half of [s] ( $\beta = 292.5, t = 2.01, p < .05$ ). This measure tended to be slightly higher in underlying /st/ clusters than in reduced *c'était* pronunciations. These spectral differences were confirmed by observation of the dynamics of energy contours in low-pass filtered signals (3.25 KHz) from the beginning of [s] up to the [t] release. The details of this analysis, which involves Functional Data Analysis (Ramsay and Silverman, 1997), can be found in Gubian et al. (2009).

We then investigated whether the identified spectral differences were caused by the strengthening of low formants (F1-F3) during the last part of [s] in elision *c'était* tokens. Automatic extraction of formant frequencies and bandwidths in this region was attempted through several standard procedures, but turned out to be erratic and did not provide any usable output. We then visualized spectrograms of tokens with



low spectral center of gravity values. Only in 12 cases did we observe consistent energy patterns that might be attributable to F2 or F3. Importantly, in the majority of these cases formant-like resonances did not appear during the last part of [s], but were present throughout the whole interval marked as [s]. This pattern was found also in some underlying /st/ clusters, suggesting that it may not always be caused by the articulation of a vocalic gesture between [s] and [t].

We conclude from the comparison of underlying and elision clusters that these two groups do not differ significantly in duration. On the other hand, we observed slight but statistically significant spectral differences consistent with the possibility that elision clusters contain a reduced /e/ vowel.

### **Intrusive voicing**

In most tokens with [e] voicing (94.3%), the initial part of the [t] closure contained some voicing periods attributable to the preceding vowel. We checked if shorter [e] durations were accompanied by longer intrusive voicing durations. If so, it might be argued that reduction in [e] duration could result from increased articulatory overlap between the voicing gesture of [e] and the upcoming stop closure. Our data suggest that this is not the case, since the durations of these two intervals did not appear to be correlated ( $r = -0.09, p = .22$ ).

### **4.3.3 Results: Conditioning factors**

In the preceding subsections we examined the phonetic characteristics of the word *c'était*. In this section we investigate which factors condition the occurrence of tokens with elided /e/ vowels. In particular, we examined whether the occurrence of [e] elision on the one hand and [e] duration on the other are conditioned by the same factors. If elision is the result of extreme gradient reduction, the occurrence of elision and [e] duration are expected to share at least some conditioning factors.

We used as predictors the prosodic and segmental variables presented in Section 4.3.2 (i.e. phrasal position, number of syllables in the AP, distance in number of syllables from *c'était* to the start and end of its AP, and the preceding segmental context) as well as speech rate. Speech rate was calculated by dividing the duration in seconds of a given AP, excluding its last syllable, by its number of canonical syllables minus one. The last syllable of the AP was excluded because, due to final lengthening, the inclusion of this syllable would make the estimation of speech rate less precise in phrase-initial and phrase-medial position (where the analyzed *c'était* tokens were always found). Speech rate estimates calculated in this way may be correlated with [e] duration and the occurrence of elision, since [e] duration and elision (defined as [e] duration of 0 ms), were used for the computation of speech

rate. For this reason, we also subtracted the duration of [e] from the overall AP duration.

Separate regression models were fitted for the two dependent variables: the occurrence of [e] elision (defined as the absence of voicing between [s] and [t]) and the duration of [e] (defined as the duration of voicing between [s] and [t]) in tokens with positive values (non-elision tokens). Mixed-effects regression with speaker as a random effect was used in all the analyses reported below. For the analysis of the occurrence of [e] elision, a binary variable, we used mixed-effects logistic regression.

Whenever two potentially significant predictors, say A and B, were significantly correlated, we orthogonalized them by replacing variable A in the corresponding regression model with the residuals of a linear model in which A was predicted by B. These residuals capture the information in A that cannot be attributed to its correlation with B. Furthermore, two prosodic variables, the number of syllables in the AP and the number of syllables from *c'était* to the end of the AP, were highly skewed to the right. These variables were log-transformed so that their distributions better approached a normal shape.

### Elision of [e]

The dataset used in the analysis of [e] elision included 279 elision and 171 non-elision tokens. Only two predictors were statistically significant. First, elision was more likely the higher the speech rate ( $\beta = 0.28, z = 3.52, p < .0005$ ). This does not mean, however, that elision was rare at slow speech rates. Figure 4.6a shows boxplots of speech rate for the elision and non-elision groups. It can be seen from this figure that, although there was a general tendency for speech rate to be higher in the elision group, the majority of elision tokens occurred at relatively slow and normal speech rates ( $< 10$  syllables per second).

Second, the distance from *c'était* to the end of its AP had a positive effect on the probability of elision ( $\beta = 2.17, z = 5.13, p < .0001$ ): elision is more likely the farther away *c'était* is located from the end of the AP. Figure 4.6b illustrates the effect of this factor on the probability of [e] elision.

Elision was not affected by phrasal position (IP-initial vs. AP-initial vs. phrase-medial), by the presence of a silent pause before *c'était*, nor by the preceding segmental context (vowel vs. consonant) ( $p > .1$  in all cases). The number of syllables in the AP was not a significant predictor when an orthogonalized version of the number of syllables to the end of the AP, a factor with which it was highly correlated ( $r = .94$ ), was included in the model.

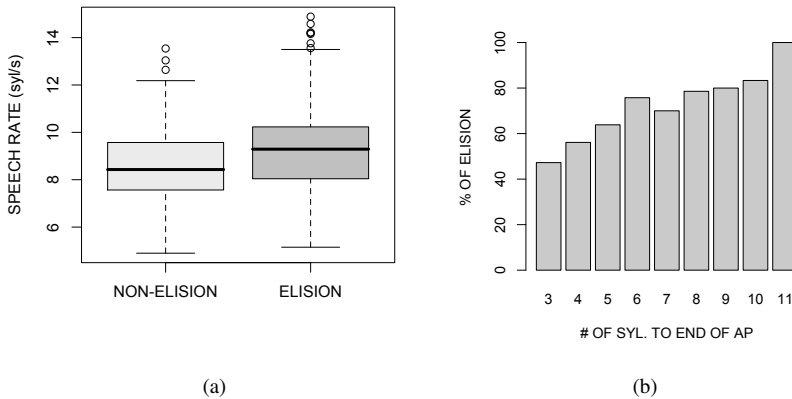


Figure 4.6: (a) Boxplot of speech rate values for elision and non-elision realizations of *c'était*. (b) Barplot of percentages of elision as a function of the number (#) of syllables (syl.) from *c'était* to the end of the AP.

### Duration of [e]

As explained in Section 4.3.1, the interval defined as [e] duration extended from the onset of voicing after the voiceless [s] segment to the beginning of the [t] closure. The label [e] duration was used for the sake of simplicity (rather than ‘periodicity duration between voiceless [s] and closure onset’), and should not be understood as vowel duration in a general sense. Only non-elision tokens ( $n = 171$ ) were used in the analysis of [e] duration. The durations approached a normal distribution and ranged from 6 to 72 ms. Contrary to [e] elision, [e] duration was not sensitive to speech rate ( $\beta = 0.04, t = 0.06, p = .95$ ) or to the distance from *c'était* to the end of the AP ( $\beta = -1.34, t = -0.33, p = .73$ ). Instead, it was shorter when *c'était* followed an IP boundary than when it was in phrase-medial position ( $\beta = -7.08, t = -2.1, p < .05$ ). This suggests that elision and the shortening of [e] duration are qualitatively different phenomena, since they appear to be conditioned by different factors.

Segments may be expected to be longer in IP-initial position than in phrase-medial position as a consequence of initial prosodic strengthening. For this reason, we wondered if the shortening effect of a preceding IP boundary on [e] also held for [s]. Contrary to [e] duration, [s] duration was found to be longer in IP-initial position than in phrase-medial position ( $\beta = 7.05, t = 2.19, p = .05$ ). Furthermore, we found that [e] duration was inversely correlated with the duration of [s] ( $r = -.28, t = -3.96, p < .0005$ ). This raised the question whether the effect of

phrasal position on [e] duration was not actually an effect of [s] duration. We fitted a final model predicting [e] duration with phrasal position and an orthogonalized version of [s] duration (the residuals of a linear model predicting [s] duration with phrasal position) as predictors. The results showed that only [s] duration was a significant predictor of [e] duration ( $\beta = -0.28, t = -4.03, p < .0001$ ). This suggests that the shortening of [e] was caused by the strengthening of [s], regardless of phrasal position, rather than by phrasal position per se, and that the shortening of [e] is not due to general reduction and hypoarticulation.

#### 4.4 Discussion

This study has shown that the word *c'était* /setɛ/ is frequently reduced in conversational French. More than half of the tokens of this word extracted from the Nijmegen Corpus of Casual French presented no apparent (or at most very weak) traces of vowel /e/ between [s] and [t]. Moreover, over a third of these tokens also lacked a complete [t] closure. These findings are in agreement with previous observations that reduction phenomena are more pervasive in everyday connected speech than is usually thought (Ernestus, 2000; Kohler, 2000; Shockey, 2003; Johnson, 2004).

In order to investigate whether vowel /e/ in cases of vowel elision was categorically absent or only gradually reduced, we measured several acoustic parameters in the region in which it is expected to manifest itself. All durational measurements support the hypothesis that elision involves the categorical absence of a segmental slot corresponding to vowel /e/ between /s/ and /t/. First, the distribution of [s(e)] durations was bimodal, with each of its modes roughly coinciding with the modes of the distributions of elision and non-elision tokens. Since gradient reduction can be expected to result in a unimodal distribution of durations, this bimodality can be seen as evidence of the categorical absence of /e/ in a subset of *c'était* pronunciations. Second, it was found that the duration of [s] and [t] closures were shorter in elision cases than in non-elision cases (also after covariates such as speech rate were controlled for). These shorter durations of [s] and [t] in elision cases may be attributed to the fact that [s] and [t] form a genuine consonant cluster, rather than an apparent /st/ cluster containing a reduced /e/ vowel. In languages like English and Dutch, for instance, consonants [s] and [t] have been found to be shorter when part of a complex onset than when part of a simple onset (Klatt, 1974; Crystal and House, 1988; Waals, 1999). This interpretation was further supported by the lack of durational differences between [st] clusters arising from elision and underlying /st/ clusters. Finally, the shortening of [s] in elision *c'était* tokens is at odds with an interpretation of elision as the result of gradient vowel reduction and devoicing. If the interval marked as [s] contained a vowel devoiced as a consequence of articu-

latory reduction, [s] in elision tokens should have been longer, not shorter, than in non-elision tokens.

Also in support of a categorical view of [e] elision, it was found that the probability of elision and [e] duration were conditioned by different factors. The occurrence of [e] elision was favored by fast speech rates and in utterances in which the word *c'était* occurred far from the end of its AP, suggesting that elision could be used by speakers as a strategy to reduce the overall duration of an ongoing utterance. On the other hand, the duration of [e] appeared to be inversely correlated with the duration of [s]. Importantly, in the case of non-elision tokens, [s] was not longer at faster speech rates and when far from the end of the AP, as would be expected if the occurrence of elision and the shortening of [e] were the same phenomenon. All of this suggests that the shortening of [e] durations is due to articulatory overlap between [e] and strengthened realizations of [s], rather than to general reduction and hypoarticulation.

The findings reviewed so far, all based on durational measures, clearly support a categorical account of the elision of /e/ in the word *c'était*. However, we also found that the spectral balance in the last part of the interval marked as [s] tended to be lower in reduced *c'était* pronunciations than in underlying clusters. This may be taken as a sign that a reduced or overlapped vocalic gesture is present between the gestures of /s/ and /t/ in reduced *c'était* pronunciations, since either a relaxation or a slight opening of the [s] constriction before the closing gesture of [t], or a retracted place of articulation of [s] as a consequence of coarticulation with vowel [e], can be held responsible for the observed downward shift in spectral balance.

However, it should be noted that the spectral differences between elision and underlying clusters may have been caused by uncontrolled factors, rather than by the reduced articulation of an underlying /e/ vowel in elision clusters. Underlying /st/ clusters, which in our dataset were always found at the beginning of content words, may tend to be pronounced with more articulatory effort than [st] clusters in semantically weak words like *c'était*. The spectral differences between elision and underlying clusters may be due to such a difference in articulatory effort (van Son and van Santen, 2005) rather than to the presence of an underlying /e/ vowel. Moreover, it should be noted that the identified spectral differences were relatively small (around 300 Hz according to our regression model), and that spectrographic analysis revealed traces of reduced and devoiced vowels such as those in Figure 4.2 only in a few cases. For these reasons, we believe that these differences in spectral balance between elision and underlying clusters should be interpreted with caution.

If these spectral differences are indeed caused by an underlying vowel, the question arises then how a simple mechanism of gradual articulatory reduction or overlap can account for the duration data discussed earlier above, which indicate that, at the

temporal level, elision and vowel shortening are qualitatively different phenomena. One possibility is through gestural reorganization. Speakers might sharply reduce the word *c'était* by changing the phase relationship between the gestures of /s/ and /t/ into one similar to that of an underlying /st/ cluster. Such change in intergestural coordination would result in a gestural sequence with the durational properties of underlying clusters, but with a nonetheless different phonological and articulatory specification containing a vowel /e/. Under this hypothesis, the gestures associated to /e/ would never be categorically deleted or absent, but only downsized as a consequence of extensive overlap with the consonantal gestures of /s/ and /t/, hence the slight spectral differences between elision and underlying clusters. We are not aware of any previous research showing that such discrete changes in intergestural coordination can be the cause of reduction phenomena in everyday speech. Further research is needed to investigate this possibility.

One goal of the present study was to test the usefulness of acoustic analysis for investigating the nature of reduction phenomena in casual spontaneous speech. An advantage of acoustic analysis is that it can be applied to recordings of spontaneous conversations, which exhibit natural speech behavior, and moreover, which document speech phenomena that cannot be easily elicited under laboratory conditions. Our findings show that valuable insights into casual speech reduction phenomena can indeed be gained by using quantitative methods, rather than by providing detailed descriptions of incidental observations. In particular, analyses of the frequency distributions and conditioning factors of relevant acoustic phonetic parameters proved to be informative tools.

On the other hand, we acknowledge that acoustic data are often difficult to interpret in articulatory terms, making conclusions about the production mechanisms behind reduction phenomena unwarranted. Another limitation of acoustic data drawn from corpora of spontaneous speech is that they are often unbalanced. For instance, only 73 underlying /st/ tokens were available for our comparison of underlying and elision [st] clusters. These limitations point to the necessity of developing new experimental paradigms allowing for the use of invasive instrumentation (e.g. electromagnetic articulography, ultrasound) which do not seriously inhibit the production of casual spontaneous speech.

To sum up, our results have shown that vowel /e/ is frequently reduced in casual renditions of the French word *c'était*. Phonetic and articulatory analyses have shown that many casual speech phenomena are gradient in nature. In our case, however, all analyses based on durational data clearly pointed towards a categorical absence of vowel /e/ in a subset of *c'était* tokens. Further research is needed to explain the slight spectral differences observed between elision and underlying /st/ clusters.

---

Overall, our findings have shown that acoustic analysis can provide valuable information about the mechanisms behind reduction phenomena in casual speech.





# Phrase-medial vowel devoicing in spontaneous French

---

Chapter 5

This chapter is reformatted version of:  
Francisco Torreira and Mirjam Ernestus. Phrase-medial vowel devoicing in spontaneous French. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, pages 2006-2009.

This study investigates phrase-medial vowel devoicing in European French (e.g. /ty pø/ [t̚ypø] ‘you can’). Our spontaneous speech data confirm that French phrase-medial devoicing is a frequent phenomenon affecting high vowels preceded by voiceless consonants. We also found that devoicing is more frequent in temporally reduced and coarticulated vowels. Complete and partial devoicing were conditioned by the same variables (speech rate, consonant type and distance from the end of the Accentual Phrase). Given these results, we propose that phrase-medial vowel devoicing in French arises mainly from the temporal compression of vocalic gestures and the aerodynamic conditions imposed by high vowels.

## 5.1 Introduction

### 5.1.1 *Vowel devoicing*

Vowel devoicing occurs in many languages across the world, such as Japanese (Hirose, 1971; Yoshioka, 1981; Beckman and Shoji, 1984), Korean (Sun-Ah Jun, 1998; Mo, 2007), Greek (Dauer, 1980), Turkish (Jannedy, 1995) and Québec French (Cedergren, 1985). In this study, we investigate the occurrence and characteristics of phrase-medial vowel devoicing in European French, a language for which only phrase-final devoicing has been thoroughly studied in the past (Fagyal and Moisset, 1999; Smith, 2003).

Vowel devoicing may originate from different mechanisms, but it is not always clear which mechanism is responsible for vowel devoicing in a particular language. First, it can be the result of a high-level phonological process or rule by which the voicing feature of a vowel is modified to match that of an adjacent voiceless consonant. In support of this mechanism, Hirose (1971) found that the vocal folds remain abducted during the production of Japanese devoiced vowels. Second, vowel devoicing can also be regarded as a lower-level phonetic effect. For instance, Jannedy (1995); Sun-Ah Jun (1998) have argued that vowel devoicing in Turkish and in Korean results from the overlap of the glottal gestures of vowels with those of their adjacent voiceless consonants.

Another phonetic explanation involves the aerodynamics of vowel production. If the transglottal pressure differential necessary for voicing is not achieved, for instance during the production of an extremely constricted and short high vowel, voicing will not be possible even if the vocal folds are properly adducted. The cross-linguistic observation that vowel devoicing typically affects unstressed high vowels lends support to this aerodynamic explanation in many cases.

### 5.1.2 *Vowel devoicing in French*

European French is well known for its use of vowel devoicing in phrase-final position as a prosodic and discourse marker (Fagyal and Moisset, 1999; Smith, 2003). This kind of devoicing affects high and mid vowels before a pause, and can be characterized as a long and sustained fricative noise preceded by a portion of a voiced vowel, rather than as a reduced or elided vowel.

On the other hand, phrase-medial devoicing has been only briefly reported for this language, although it appears to be recurrent in connected speech: in a study of short vowels (< 30 ms), automatically extracted from a corpus of conversational speech, Meunier et al. (2008) found that devoicing affected 20% of the high vowels examined, and in some rare cases, non-high vowels as well. In the vast majority of

cases, the devoiced vowels were preceded, but not necessarily followed, by voiceless consonants.

### 5.1.3 The present study

In the present study we further document and investigate the occurrence of phrase-medial vowel devoicing in European French. By phrase-medial devoicing we understand devoicing in non-utterance-final position, as in the utterance /ty pø/: [tʏpø̥] *tu peux* ‘you can’. We first examine the frequency of complete and partial devoicing in a corpus of spontaneous casual French. Second, we investigate what factors condition the likelihood of complete devoicing and the extent of partial devoicing. Then we look for relationships between devoicing and syllable, stop closure and vowel durations. Finally, we examine whether devoicing is accompanied by increased consonant-vowel (CV) coarticulation, as captured by spectral measures.

## 5.2 Method

### 5.2.1 Materials

Based on the observations made by Meunier et al. (2008), we decided to focus on vowels and on contexts for which phrase-medial devoicing is recurrent, that is on *unaccented high vowels preceded by a voiceless consonant* (/p/, /t/, /k/, /f/, /s/ or /ʃ/). In order to consistently measure the end of the target vowels, we decided to further restrict our study to vowels followed by a consonant involving a complete oral closure (i.e. a stop, either oral or nasal, and voiced or voiceless) in the same or following word. We considered both function and content words. On the basis of the restrictions above, only three function words were studied: *qui* /ki/, *si* /si/ and *tu* /ty/ ‘you’. Since we restricted ourselves to unaccented function words, in our data the word *qui* was always a relative pronoun, and the word *si* was always a conditional conjunction (‘if’) or an adverb (‘so’). As for content words, only words with target vowels in non-final position (where they cannot receive a phrasal accent) were considered. Tokens containing laughter, disfluencies or intrusive background noise were discarded. We also discarded tokens in which the postvocalic consonant did not exhibit a clear closure in the signal. Target vowels meeting the conditions above were randomly extracted from the Nijmegen Corpus of Casual French (Torreira et al., 2010, see Chapter 2) until 250 content words (151 word types) and 300 function words (100 words for each function word type) were available for analysis.

### 5.2.2 Measurements and annotation

The first author made the following measurements and annotations for every consonant-vowel-consonant sequence ( $C_1VC_2$ ) in the dataset:

- Closure duration: only for  $C_1$  stops ( $n = 192$ ).
- VOT: only for  $C_1$  stops, from stop release to onset of voicing in V, or in its absence, to the start of the upcoming  $C_2$  closure.
- Voicing duration: from the onset of voicing corresponding to the target vowel to the onset of the  $C_2$  closure.
- Vowel duration: only for tokens with a stop as  $C_1$ , the sum of VOT and voicing duration, corresponding to the stretch of the signal in which the vocal tract was unconstricted between  $C_1$  and  $C_2$ .
- Voiceless frication duration: only for  $C_1$  fricatives, from increase in acoustic energy above 4 KHz to onset of voicing in the vowel, or in its absence, to the start of the following stop closure.
- Syllable duration: the sum of closure/voiceless frication, VOT and voicing duration.
- Prosodic context: presence of an Accentual Phrase (AP) or Intonational Phrase (IP) boundary before/after the target word. The annotation of AP and IP boundaries was carried out following Fougeron and Jun (1998).
- Number of syllables in the AP containing the target word, and distance in number of syllables from the target syllable to the AP start and end.
- Speech rate: (number of syllables in AP - 1) / (Duration of AP excluding last syllable). We excluded the last syllable in the AP from the computation of our speech rate estimate to avoid variability due to final lengthening.

A check was performed in order to assess the reliability of the measurements and annotations. One hundred tokens from the dataset were randomly selected, and independently analyzed by a trained transcriber unaware of the purposes of our study. No major disagreement was found between our measurements and those of the independent transcriber (100% of closure duration, 91% of VOT and 96% of voicing duration measurements differed by 10 ms or less between the two annotators; the annotation of the location of preceding and following intonational boundaries coincided in 82% and 92% of the cases, and the annotation of the type of boundary in 92% and 77% of the cases).

### 5.2.3 Results

#### Frequency of complete devoicing and distribution of voicing duration

More than a third of the target vowels ( $n = 199$ , 36.3%) were completely devoiced (no voicing was found between the prevocalic and postvocalic consonants). The voicing durations of the remaining vowels formed a unimodal distribution with a mean of 30 ms and a standard deviation of 14 ms. A considerable number of tokens had voicing durations below 15 ms ( $n = 49$ ). In agreement with Meunier et al. (2008), these numbers indicate that vowel devoicing is a recurrent phenomenon in European French. Figures 5.1 and 5.2 illustrate cases of complete and partial devoicing in our data.

#### Conditioning variables

We investigated which of the variables that we had annotated conditioned the occurrence of complete devoicing and the duration of vocalic voicing. If complete and partial devoicing are different points in a phonetic continuum, they should be conditioned by the same factors. Based on previous research, we expect devoicing to occur more often at faster speech rates and in non-prominent prosodic locations (far from phrasal prosodic edges, word-medially; Sun-Ah Jun, 1998; Jannedy, 1995). If devoicing is seen as a reduction phenomenon, it can also be expected to occur more often in function than in content words. We first fitted a logistic regression model with complete devoicing as the response and the following predictor variables: speech rate, vowel type (/i/, /y/ or /u/),  $C_1$  manner,  $C_1$  place of articulation,  $C_2$  voicing, preceding and following prosodic context, AP length, distance to AP end, distance to AP start, position in the word (word-initial vs. word-medial syllable), word class (content vs. function word). Speaker was included in the model as a random factor.

Three variables were found to favor the probability of complete devoicing. First, complete devoicing was more probable the faster the speech rate ( $\beta = -0.16, z = -3.53, p < .0005$ ). Second, devoicing tended to occur more often after fricatives than after stops ( $\beta = -0.52, z = -2.65, p < .01$ ). Third, devoicing was more probable the further away the target vowel was from the right edge of its AP ( $\beta = -0.63, z = -4.15, p < .0001$ ). No other variable was a significant predictor in the model.

We then fitted a linear regression model using only tokens with positive voicing durations. This model had voicing duration as the response, and the same variables listed above as predictors. The obtained model was similar to the logistic regression model above, except for a significant interaction between  $C_1$  manner and word class. Voicing duration was shorter at faster speech rates ( $\beta = -1.07, t = -2.99, p < .005$ ) and the further away the target vowel was from the end of the AP ( $\beta = -4.15, t =$

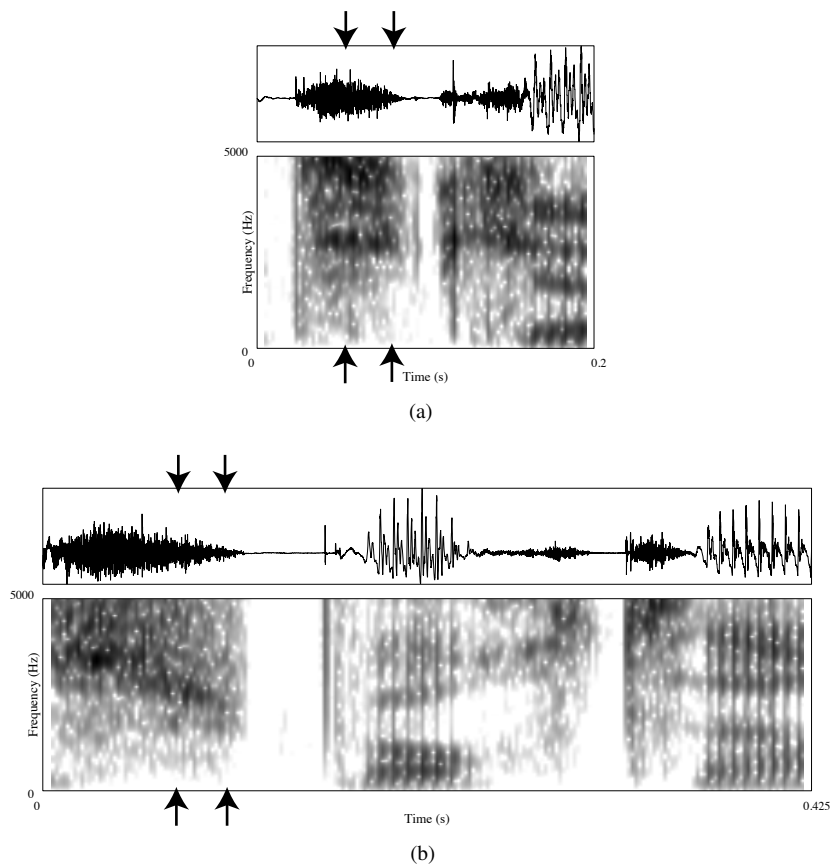


Figure 5.1: Examples of complete vowel devoicing in the words (a) *ticket* /tikɛ/ ‘ticket’ and (b) *supporter* /супрəтєв/ ‘sports fan’. Arrows indicate the approximate location of the devoiced vowel.

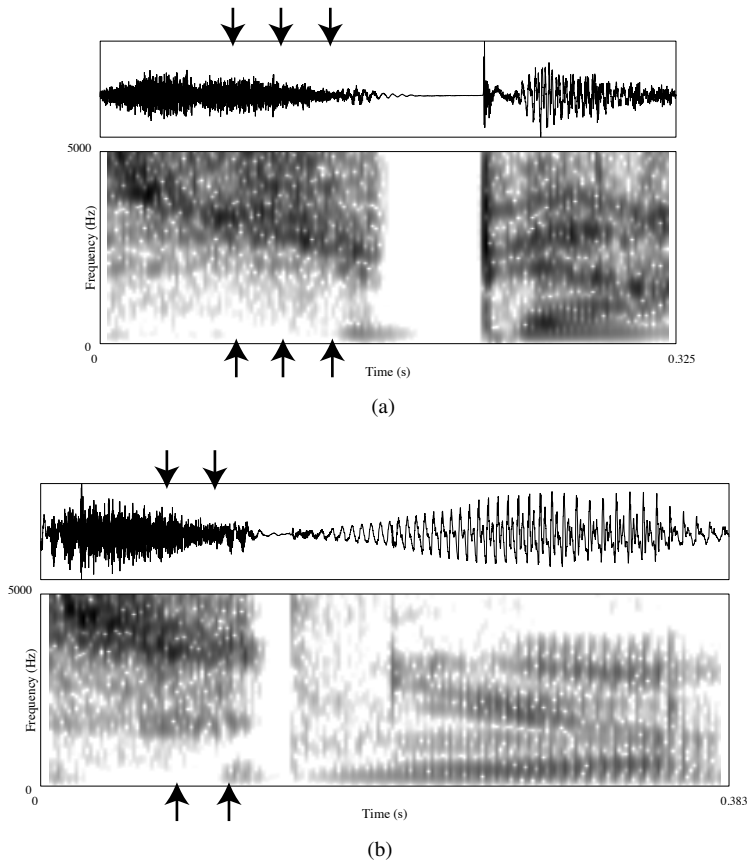


Figure 5.2: Examples of partial vowel devoicing in the words (a) *super* /sypɛs/ ‘super’ and (b) *soutenu* /sutny/ ‘formal’. Arrows indicate the approximate location of the devoiced vowel.

$-3.096, p < .005$ ). As for the interaction, voicing duration tended to be shorter after fricatives than after stops ( $\beta = -4.4, t = -2.07, p < .05$ ), but this effect did not appear to hold for function words. Since the only function word having a fricative was *si*, it cannot be concluded whether the interaction should be attributed to the word *si*, to consonant /s/ or to fricatives in function words.

We conclude from these analyses that complete and partial devoicing were generally sensitive to the same factors, namely speech rate,  $C_1$  manner and the distance from the target syllable to the end of the AP. This suggests that phrase-medial vowel devoicing in European French is a continuous process, with completely devoiced vowels at its endpoint.

### **Relationships between devoicing and syllable, closure and vowel durations**

Previous research on several languages has shown that vowel devoicing is usually accompanied by a temporal compression of the vowel and its surrounding consonants (Jannedy, 1995; Sun-Ah Jun, 1998). We checked whether this is also true for French by looking for statistical relationships between the occurrence of complete devoicing and syllable, closure and vowel durations.

A logistic regression model with complete devoicing as response and syllable duration as predictor indicated that complete devoicing was more likely the shorter the target syllable ( $\beta = -0.026, z = -7.72, p < .0001$ ). In order to control for potential confounds, we fitted a second model in which syllable duration had been normalized with respect to its segmental composition, speech rate and its distance to the end of the AP. This was done by using as predictor the residuals of a linear model predicting syllable duration with these variables. Again, we found that devoicing was more likely the shorter the syllable ( $\beta = -0.022, z = -6.31, p < .0001$ ). These results are illustrated in Figure 5.3, which shows syllable duration as a function of devoicing. This finding suggests that phrase-medial devoicing in French involves some sort of temporal compression at the supraglottal level.

We then investigated whether this temporal compression equally affected the consonantal and vocalic parts of the syllable. Because the consonantal and vocalic parts in a syllable with a fricative onset and a devoiced vowel cannot be measured reliably on the basis of a discrete acoustic landmark (e.g. a stop release), we only considered syllables with intervocalic stop onsets ( $n = 315$ , of which 99 had a completely devoiced vowel). We fitted regression models with closure and vowel duration as responses, the occurrence of complete devoicing as the main predictor and speech rate,  $C_1$  and V types, and distance from the target syllable to the end of the AP as covariates. The model predicting closure duration showed that this variable was not affected by devoicing ( $\beta = 3.04, t = 1.27, p > .2$ ). In other words, the duration of



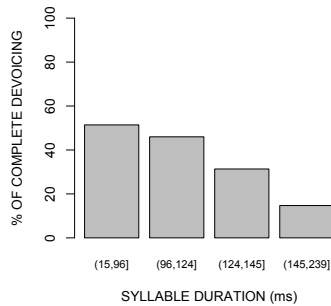


Figure 5.3: Percentage of complete devoicing as a function of syllable duration for each quartile.

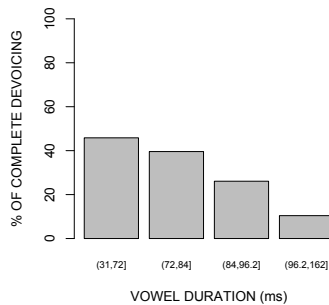


Figure 5.4: Percentage of complete devoicing as a function of vowel duration for each quartile.

stop closures did not appear to differ between syllables with a devoiced vowel and syllables with a voiced vowel. On the other hand, we found a significant effect of devoicing on vowel duration (defined here as the interval from the  $C_1$  stop release to the  $C_2$  closure onset, see above). Devoiced vowels tended to be significantly shorter than voiced vowels ( $\beta = -28.62, t = -28.37, p < .0001$ ). It appears therefore that devoicing is accompanied by the temporal reduction of only the unconstricted part of the syllable, rather than of the syllable as a whole. This finding is illustrated by Figure 5.4, which shows the percentage of complete devoicing as a function of vowel duration.

### Complete devoicing and the spectral characteristics of C<sub>1</sub>

It has been claimed for Japanese, Turkish and Korean that vowel devoicing occurs as the result of increased gestural overlap between vocalic and consonantal gestures (Beckman and Shoji, 1984; Jannedy, 1995; Sun-Ah Jun, 1998). We now investigate whether this is true for French. Under this hypothesis, it can be expected that devoiced vowels will exhibit signs of increased CV coarticulation. Since the acoustic effects of coarticulation depend greatly on the specific consonant and vowel sequence under examination, we focused only on sequences for which at least 100 tokens were available for analysis: /ki/, /ty/ and /si/. The vast majority of tokens were instances of the function words *qui*, *tu* and *si*. In the case of /ki/ and /ty/ we calculated the spectral center of gravity (CoG) and the spectral peak (SP) in a 20 ms Hamming window left-aligned with the stop release. In the case of /si/, we calculated the spectral center of gravity and the spectral peak 25% into the voiceless frication interval (measured from the left) in a 0.5-10 KHz band. In all three syllables, we examined whether devoicing had an effect on these spectral measures. In the case of /ki/, increased coarticulation between /k/ and /i/ should translate into increased /k/ fronting and higher values for the two spectral parameters considered. On the other hand, in the case of /ty/ and /si/, increased palatalization of /t/ and /s/ should result into a downward shift of these parameters. In all models, gender was included as a covariate, and speaker as a random factor.

For /ki/ and /ty/, we did not find any statistical spectral differences between tokens with devoiced and voiced vowels. However, in all four models the signs of the obtained regression coefficients were consistent with the hypothesis of increased coarticulation for syllables with devoiced vowels (/ki/  $\beta = 90.5$  for SP and  $\beta = 213.4$  for CoG; /ty/  $\beta = -152.6$  for SP and  $\beta = -180.4$  for CoG). For /si/, we found statistically significant differences in the direction predicted by the hypothesis (SP:  $\beta = -582.7$ ,  $t = -2.02$ ,  $p < .05$ ; CoG:  $\beta = -436.5$ ,  $t = -2.62$ ,  $p < .01$ ). It should be noted that the number of available tokens was considerably higher for /si/ than for /ki/ and /ty/ (/si/:  $n = 145$ , /ki/:  $n = 104$ , /ty/:  $n = 115$ ). For this reason, it can be speculated that statistically significant differences for /ki/ and /ty/ might have been observed if we had used bigger samples. In summary, we found some acoustic evidence of increased CV coarticulation in cases of complete devoicing.

### 5.3 Discussion

We have shown that phrase-medial devoicing occurs frequently in connected European French. In our data, over a third of the studied high vowels are completely devoiced, and an important number of tokens are partially devoiced in a considerable degree, both in function and content words. Devoicing occurs for syllables of

all durations, but it tends to be more frequent in syllables with temporally reduced vowels. Study of the spectral characteristics of  $C_1$  provided some evidence of increased coarticulation between  $C_1$  and V in cases of complete devoicing, especially for syllables with fricative onsets. Finally, we found that the occurrence of complete and partial devoicing are conditioned by similar variables (speech rate, manner of articulation and distance to the upcoming AP boundary). Surprisingly, we did not find differences in the likelihood and extent of devoicing between function and content words.

In view of these results, we propose that phrase-medial vowel devoicing in European French is a phonetic phenomenon arising mainly from temporal compression and affecting a class of vowels which, by their constricted nature, are unfavorable to the production of voicing for aerodynamic reasons. Although the observed increased CV coarticulation in devoicing cases suggests that articulatory overlap between C and V may be the main cause of this vowel shortening, a mechanism of truncation of the vocalic gestures should be considered as well. Articulatory data are needed to elucidate this question.

Finally, given the high frequency of phrase-medial devoicing in French, one may speculate to what extent it has been or is being conventionalized as a proper phonetic target, or whether, on the other hand, it is merely the byproduct of the mechanical and phonetic factors mentioned above. In any case, it appears that the observed pattern of vowel devoicing in French may lay the foundation for future phonologizations, which, for instance, might lead to the emergence of a class of voiceless vowels or affricated consonant allophones (e.g.  $t \rightarrow tS \_i$ ) in this language.



# Realization of voiceless stops and vowels in conversational French and Spanish

---

## Chapter 6

This chapter is a reformatted version of:  
Francisco Torreira and Mirjam Ernestus (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Journal of Laboratory Phonology*, 2:331–353.

The present study compares the realization of intervocalic voiceless stops and vowels surrounded by voiceless stops in conversational Spanish and French. Our data reveal significant differences in how these segments are realized in each language. Spanish voiceless stops tend to have shorter stop closures, display incomplete closures more often, and exhibit more voicing than French voiceless stops. As for vowels, more cases of complete devoicing and greater degrees of partial devoicing were found in French than in Spanish. Moreover, all French vowel types exhibit significantly lower F1 values than their Spanish counterparts. These findings indicate that the extent of reduction that a segment type can undergo in conversational speech can vary significantly across languages. Language differences in coarticulatory strategies and “base-of-articulation” are discussed as possible causes of our observations.

## 6.1 Introduction

### 6.1.1 *Phonetic reduction in conversational speech and purpose of this research*

In conversational speech, segments often lack some or all of the acoustic cues that researchers are accustomed to finding in other, clearer, speech styles. For instance, intervocalic voiceless stops may lack their characteristic period of silence and burst (e.g. Lewis, 2001; Warner, 2005), while vowels may be produced with undershot formant frequencies (Koopmans-van Beinum, 1985; Moon and Lindblom, 1994, among others), and be partially or completely devoiced under the coarticulatory influence of neighboring voiceless consonants (e.g. Davidson, 2006; Torreira and Ernestus, 2010b, see Chapter 5). In extreme cases of reduction, segments may be acoustically absent (e.g. Browman and Goldstein, 1990; Johnson, 2004).

Reductions of this kind are often regarded as the consequence of the interaction between principles of economy of effort and universal articulatory and perceptual constraints (Lindblom, 1990). Under this hypothesis, reduction patterns should be largely the same in all languages. Common cross-linguistic coarticulatory patterns and recurrent diachronic lenitions (e.g. fronting of /k/ before front vowels, which lead to /tʃ/ in Romance, Slavic and Bantu languages; Recasens and Espinosa (2009)) support this view to a large extent. On the other hand, there are reasons to believe that at least some aspects of the reduction patterns present in connected speech are language-specific. For instance, it is now widely accepted that the strength and exact nature of coarticulatory processes often appears to differ among languages when closely examined (e.g. Öhman, 1966; Boyce, 1990; Manuel, 1990; Solé, 1995; Manuel, 1999; Beddor et al., 2002), and that the exact outcome of diachronic weakening processes is never predictable.

The present work directly compares the realization of intervocalic voiceless stops and vowels surrounded by voiceless stops in conversational Spanish and French in order to determine to what extent reduction patterns can differ between similar segmental sequences in two different languages. In order to assess reduction patterns within each language, we examine consonants and vowels in accented and unaccented syllables. According to manuals of Spanish and French pronunciation (e.g. Navarro Tomás, 1977; Tranel, 1987), the phonemes /p t k/ in these two languages should be pronounced similarly: as unaspirated voiceless stop consonants. However, previous studies carried out separately on French and Spanish raise the question of whether the speakers of these languages follow different production strategies when confronted with voiceless stops occurring between two vowels. Spanish voiceless stops in intervocalic position are frequently realized as voiced approximants in spontaneous speech (Lewis, 2001, and references therein), but to our knowledge no such

tendency has been reported for French voiceless stops. On the other hand, unaccented vowels in French can be extremely short and are often partially and sometimes even completely devoiced in the case of high vowels (Meunier et al., 2008; Torreira and Ernestus, 2010b, see Chapter 5), while vowel devoicing of this kind has not been reported for Spanish. These observations lead us to hypothesize that while French speakers privilege the articulation of voiceless stops and demonstrate greater vowel reduction, reduction in Spanish affects stops more than vowels. The present study addresses this hypothesis by directly contrasting French and Spanish voiceless stops and vowels extracted from corpora of spontaneous speech.

A key aspect of the present study is that it uses corpora of conversational speech collected in a similar way for each language: the Nijmegen Corpus of Casual French (Torreira et al., 2010, see Chapter 2) and the Nijmegen Corpus of Casual Spanish (Torreira and Ernestus, 2010a). Importantly, these corpora contain casual, spontaneous conversations held by groups of acquainted speakers. We believe that data of this kind is needed to reveal the extent and nature of reduction phenomena in conversational speech.

In the following subsections, we introduce the main reduction patterns that we consider in this study: loss of total closures and the occurrence of voicing in intervocalic voiceless stops, and devoicing and formant undershoot for vowels.

### **6.1.2 Reduction of voiceless stops**

According to standard descriptions of Spanish and French, voiceless stops in these languages should be realized similarly: with voiceless oral closures and no aspiration (e.g. Navarro Tomás, 1977; Tranel, 1987). Given the articulatory characteristics of these segments, in intervocalic position we can expect them to be weakened in two main ways. These two reduction patterns respectively involve the supraglottal and glottal components of the speech production system. First, at the supraglottal level, the stop closure can be temporally and spatially reduced, resulting in a stop with a short or an incomplete closure (e.g. Spanish *físico* /fisiko/ [fisiyo] ‘physical’). This kind of reduction has been reported to occur frequently in several Spanish dialects including Colombian and Northern Peninsular Spanish (Lewis, 2001) and Majorcan Spanish (Hualde et al., to appear). The present study further documents this phenomenon for spontaneous speech in Madrid Spanish, and examines the extent to which it occurs in standard French.

A second way in which intervocalic voiceless stops can be lenited involves voicing at the glottis. In intervocalic position, an optimal realization of /p/, /t/ or /k/ requires an abrupt cessation of glottal vibrations at the onset of the stop closure. However, due to the well-known tendency of contiguous speech gestures to overlap and blend with each other, or simply to reduce the size of the devoicing gesture (Browman and Gold-

stein, 1990), glottal vibrations from a preceding vowel may sometimes persist well into the stop closure, and even throughout its whole duration. Studies have shown that this reduction pattern is pervasive in several Spanish dialects. Lewis (2001), for Northern Peninsular and Colombian Spanish, and Hualde et al. (to appear) for Majorcan Spanish, found that phonologically voiceless intervocalic stops in these dialects are often significantly voiced, and that this tends to occur more often in conversational than in read speech, suggesting that the occurrence of voicing in intervocalic voiceless stops is a characteristic of hypoarticulated speech. The voicing of intervocalic voiceless stops has also been observed for the Spanish varieties spoken in Andalusia (Salvador, 1968), the province of Toledo (Torreblanca, 1976), Barcelona (Machuca-Ayuso, 1997), Cuba (Ruiz Hernández, 1984), Panama (Quilis and Graell, 1992), and the Canary Islands (Oftedal, 1985; Trujillo, 1970). The present study investigates the occurrence of this reduction phenomenon in Madrid Spanish, and examines the extent to which it occurs in French.

### **6.1.3 Vowel reduction**

#### **Temporal and spectral reduction**

Vowels can be reduced in terms of duration and spectral characteristics (Lindblom, 1963; van Son and Pols, 1990, 1992; Moon and Lindblom, 1994, among many others). Reduction in duration is straightforward, but the relation between reduction and the spectral characteristics of vowels is less so. Traditionally, vowel reduction has been understood as a process of vowel centralization by which vowel types become less distinct from each other and more similar to a central schwa-like vowel (Stetson, 1951; Delattre, 1969). However, under the coarticulatory influence of consonants requiring a constricted vocal tract, as in our case, not only low vowels but also high vowels may be expected to become more closed in their articulation, leading not to centralization, but rather to a raising of the vowel space. For instance, Lindblom (1963) found that undershoot in Swedish /bɪb/, /dɪd/ and /gɪg/ sequences did not lead to higher F1 values, as predicted by the centralization hypothesis, but to lower F1 values instead.

#### **Vowel devoicing**

As mentioned above, we also aim to investigate the occurrence of vowel devoicing in French and Spanish vowels. European French is known for its use of vowel devoicing in phrase-final position as a prosodic and discourse marker (Fagyal and Moisset, 1999; Smith, 2003). Because many phrase-final devoiced vowels in this language are long and sustained fricative noises preceded by voiced vocalic portions, rather than shortened or elided vowels, we cannot consider all cases of French phrase-final



devoicing as instances of reduction. For this reason, phrase-final vowel devoicing is not considered in this study.

We will focus instead on phrase-medial devoicing (e.g. *tu peux* /ty pø/ [t<sub>y</sub>pø] ‘you can’), another kind of devoicing reported to occur in European French (Martin, 2004; Meunier et al., 2008). Meunier et al. (2008) extracted short vowels (< 30 ms) from a corpus of conversational French, and found complete devoicing in 20% of the high vowels examined, and in a few tokens of non-high vowels as well. In the vast majority of cases, the devoiced vowels were preceded, but not necessarily followed, by voiceless consonants. In a study based on the same French corpus used here, (Torreira and Ernestus, 2010b, see Chapter 5) further noted that devoicing is more likely the shorter the vowel, and provided some acoustic evidence of increased consonant-to-vowel coarticulation in syllables with fully devoiced vowels. These findings indicate that phrase-medial devoicing can be seen as a reduction process, since it appears to involve articulatory readjustments typical of such phenomena (e.g. shortening, increased coarticulation). In the present study, we compare the extent to which phrase-medial vowel devoicing occurs in Spanish and French vowels preceded and followed by voiceless stops.

## 6.2 Method

### 6.2.1 Materials

All of the speech data used in this study was extracted from the Nijmegen Corpus of Spontaneous Spanish (NCCSp; Torreira and Ernestus, 2010a) and the Nijmegen Corpus of Spontaneous French (NCCFr; Torreira et al., 2010). The corpora contain recordings of 52 Spanish speakers (27 female and 25 male) from Madrid and 46 French speakers (24 female and 22 male) from Central and Northern France, all engaged in conversations with friends. The NCCSp contains 20 conversations, while the NCCFr contains 23. Each of the recorded conversations had an approximate duration of 90 minutes. Except for two female speakers in the NCCFr, all speakers were university students in their late teens or twenties. The corpora were collected in sound-attenuated rooms in Madrid and Paris. The recording equipment consisted of an Edirol R-09 solid-state stereo recorder, Samson QV head-mounted unidirectional microphones and a stereo microphone preamplifier. The microphones were placed at an average distance of 5 cm from the left corner of the speakers’ lips. The sampling rates used were 44.1 kHz for the NCCSp and 48 kHz for the NCCFr. Importantly for our purposes, the NCCFr and NCCSp contain casual and spontaneous speech. (Torreira et al., 2010, see Chapter 2) showed that the NCCFr contains significantly more informal words, swear words, hesitation sounds and other indicators of casualness than the ESTER corpus, a corpus of French radio broadcasts (Galliano et al.,

2005). Since the NCCSp and the NCCFr were collected according to the same procedure, and contain recordings of speakers of the same socioeconomic background, we believe that the materials in the two corpora are of highly similar characteristics in terms of speech style.

Intervocalic voiceless stops and vowels between two voiceless stops were randomly extracted from the NCCSp and the NCCFr. Segments were always part of an open syllable in a content word, with stops always in onset position. We only analyzed sequences with vowel types shared by the Spanish and French vowel inventories (/a/, /e/, /i/, /o/ and /u/). For Spanish, therefore, all vowel types in its inventory were considered, while for French these were only a subset of the 14 to 16 vowel types generally recognized by linguists (e.g. Fougeron and Smith, 1993; Coveney, 2001; Fagyal et al., 2006). For Spanish, all lexically stressed syllables were considered as accented, since in this language lexically stressed syllables usually carry a pitch accent (except in rare cases of deaccentuation, as in parenthetical utterances; Prieto and Ortega-Llebaria, 2006). For French, only syllables carrying a primary accent were considered as accented. These syllables occurred at the end of prosodic units one level below the utterance level, corresponding to the Accental Phrase in Jun and Fougeron (2002), the rhythmic unit in Di Cristo (1999) and the *intonème mineur* proposed by Rossi (1999). Syllables in non-phrase-final position were considered as unaccented, excepting a few cases in which F0 signs of a prenuclear accent were found. These cases were excluded from analysis.

French is said to have mid-high (/e/, /o/) and mid-low vowels (/ɛ/, /ɔ/), but the existence of a phonological contrast between unaccented mid-high and mid-low vowels in this language is controversial, especially in open syllables, where more closed realizations are preferred according to the *loi de position* (see Nguyen and Fagyal, 2008, for a more detailed discussion). For this reason, in unaccented position we merged these vowel categories into single categories that we denote as /e/ and /o/. In accented syllables, on the other hand, we only considered mid-high vowels (/e/ and /o/), taking as a reference *Le Petit Robert*, a widely-used dictionary of standard French.

The dataset contains a total of 1298 stops and 727 vowels for Spanish, and 856 stops and 497 vowels for French. Speakers contributed from two to 58 tokens, with an average contribution of 21.9 tokens and a standard deviation of 12.7. Table 6.1 presents examples of unaccented and accented voiceless stops and vowels in our dataset. Tables 6.2 and 6.3 show the numbers of segments in the dataset broken down by accent, type, and language. Remember that all analyzed voiceless stops are in onset position and take the accent value of the following vowel.

Spanish	
Unaccented	Accented
<i>diputado</i> /d <u>i</u> putado/ ‘deputy’	<i>ataca</i> /ataka/ ‘attacks’
<i>protocolo</i> /pro <u>t</u> okolo/ ‘protocol’	<i>que lo quites</i> /kelok <u>i</u> tes/ ‘that you remove it’
<i>lo poquito</i> /lo <u>p</u> okito/ ‘the little bit’	<i>un poquito</i> /unpo <u>k</u> ito/ ‘a little bit’

French	
Unaccented	Accented
<i>va couper</i> /va <u>k</u> upe/ ‘going to cut’	<i>va couper ta</i> /vakupe]ta/ ‘going to cut your’
<i>du côté</i> /dy <u>k</u> ote/ ‘on the side’	<i>au cas par cas tu</i> /okapaRka]ty/ ‘case by case you’
<i>critiquer</i> /kRi <u>t</u> ike/ ‘criticize’	<i>critiquer ta</i> /kRitike]ta/ ‘criticize your’

Table 6.1: Examples of words and word sequences including target voiceless stops and vowels (underlined) extracted from the NCCSp and NCCFr. Square brackets represent accentual phrase boundaries in the examples of French accented vowels.

	Unaccented			Accented		
	/p/	/t/	/k/	/p/	/t/	/k/
French	136	63	175	186	208	88
Spanish	267	141	163	163	375	189

Table 6.2: Numbers of unaccented and accented voiceless stop consonants in the dataset for each language.

	Unaccented					Accented				
	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/
French	129	32	55	106	52	26	39	21	15	22
Spanish	152	120	91	158	50	19	26	34	42	35

Table 6.3: Numbers of unaccented and accented vowels in the dataset for each language.

### 6.2.2 Measurements

In order to investigate the realization of voiceless stops and vowels in French and Spanish, we measured stop closure, vowel and voicing durations, intensity velocity maxima during the consonant-to-vowel (CV) transitions, and formant frequencies in the vowels of interest. The durations of the target stop closures and vowels were measured manually. Following Turk et al. (2006), acoustic segment durations were defined on the basis of the onsets and releases of consonantal constrictions rather than on voicing. Therefore, consonant duration included only the stop closure interval, while vowel duration included the interval extending from the oral release to the upcoming oral closure. Stop closures were annotated as being incomplete if uninterrupted acoustic energy was present in the spectrograms above the frequency band characteristic of voicing between the vowels flanking the stop. No duration values were assigned to stops with an incomplete closure nor to their contiguous vowels, since the boundaries between these segments cannot be determined on the basis of robust acoustic landmarks.

The presence of voicing within the target stops (intervocalic voicing) and vowels was identified automatically with the pitch detection function provided in Praat set to default parameters except for a time step of 2.5 ms (Boersma, 1993; Boersma and Weenink, 2009). Voiceless vowel duration (VVD) was recorded as the interval from the stop release up to the first pitch point detected in the vowel, or, in the case of completely devoiced vowels, up to the following stop closure. Figures 1 to 3 illustrate our temporal measures.

As an additional measure of consonantal strength, we measured the intensity velocity maxima in CV transitions. Stops with full closures and salient bursts should be characterized by abrupt rises in intensity and should therefore yield highly positive values of this measure, while weakened stops should exhibit values close to zero. In order to factor out most of the effects of voicing (fundamental frequency and first few harmonics), the signals were first high-passed filtered at 1 kHz. Intensity contours were calculated using a 30 ms window and a time step of 8 ms. We then identified the points of minimum intensity in the stop closure and of maximum intensity in the following vowel, and calculated intensity velocity maxima from the first derivative of the intensity curve between these two points. Derivatives were obtained by smoothing the contours via cubic interpolation and by calculating local differences in intensity using a time step of 1 ms.

Formant frequencies were estimated through the target vowels with the Burg method as implemented in Praat (with default settings except for the ceiling parameter, see below). Formant measurements were taken at the point of maximum F1 within the vowel, which usually occurred around its midpoint. In order to reduce the number of formant tracking errors, we adjusted the ceiling frequency of the formant tracker

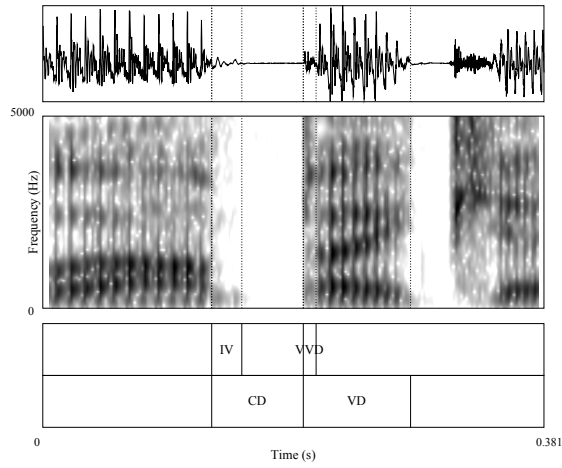


Figure 6.1: Illustration of temporal measures featuring the syllable /pa/ in the French word *paquet* ‘packet’. IV: intervocalic voicing; VVD: voiceless vowel duration; CD: closure duration; VD: vowel duration.

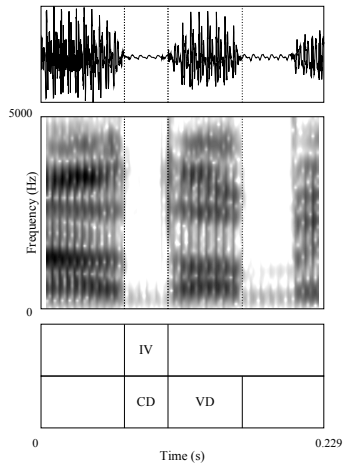


Figure 6.2: Illustration of temporal measures featuring the syllable /to/ in the Spanish word *autopista* ‘highway’. IV: intervocalic voicing; CD: closure duration; VD: vowel duration.

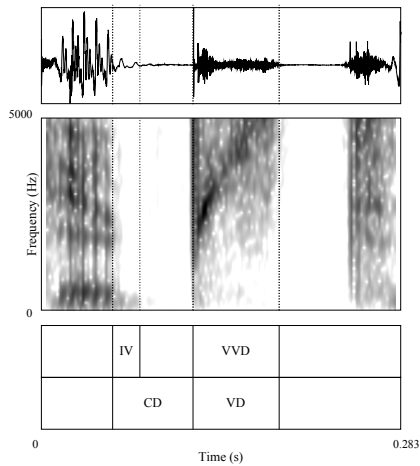


Figure 6.3: Illustration of temporal measures featuring the syllable /pi/ in the French word *capitale* ‘capital’. IV: intervocalic voicing; VVD: voiceless vowel duration; CD: closure duration; VD: vowel duration.

for each vowel type and gender following median optimal values provided in Escudero et al. (2009) for Portuguese vowels. Applying these ceilings resulted in a clear improvement in formant detection, as indicated by a drastic reduction in unlikely formant values (e.g. F2 values with typical F1 values). Inspection of the remaining unlikely values revealed that they mostly belonged to extremely short and sometimes devoiced vowels with either very weak or absent formants (F1 in particular). These cases (143 out of 1224) were excluded from the analysis of formant values. Since there are no reasons to suspect that French and Spanish speakers systematically differ in vocal tract length, and given the large number of speakers as well as the similar number of men and women in our dataset, we decided to use raw F1 values rather than to apply any specific procedure of formant normalization.

The direction of the effect of coarticulation on the F2 and F3 of a specific vowel crucially depends on its particular segmental context (Lindblom, 1963). For instance, central vowels can be expected to show higher F2 in the context of segments with an anterior place of articulation, and lower F2 in the context of segments with a posterior place of articulation. Since our unbalanced dataset does not allow us to compute reliable statistical estimates for each combination of segments under examination, we exclusively focus on F1 when assessing the degree of coarticulation between target vowels and the surrounding stop consonants. In this case, we can make the general prediction that more reduced and coarticulated vowels will display lower F1

values regardless of their segmental context (see Section 6.1.3).

A check was performed in order to assess the reliability of the measurements and annotations which had been done manually by the first author (i.e. closure and vowel durations, and the presence of a complete closure). One hundred and fifty stops and vowels (75 stops and vowels for each language) were randomly extracted from the dataset, and independently reanalyzed by an assistant unaware of the purposes of our study. No major disagreement was found between our measurements and those of the independent transcriber (93.5% of closure duration measurements and 87.7% of vowel duration measurements differed by 10 ms or less between the two annotators, and the annotation of complete and incomplete closures coincided in 93.3% of the cases).

## 6.3 Analysis and results

### 6.3.1 Analysis

In this section we analyze the presence of incomplete closures, intensity velocity maxima in CV transitions, closure duration, and intervocalic voicing for voiceless stops, and vowel duration, voiceless vowel duration, and F1 values for vowels. Apart from language (Spanish, French), consonant type (/p/, /t/, /k/) and vowel type (/a/, /e/, /i/, /o/, /u/), we consider accent (accented vs. unaccented) as predictors. Position in the word (word-initial vs. word-medial) was included in the analysis of consonants in order to control for variance, since this factor has been reported to condition the strength with which consonants are articulated (see Keating, 2006, and references therein). In the analysis of incomplete closures, we also consider the height of the surrounding vowels, since incomplete closures can be expected to occur more often between open vowels (/a/) than between closed vowels (/i/ and /u/) for coarticulatory reasons.

For all analyses, we use mixed-effects linear regression with contrast coding as implemented in the *lmer* function in R (Bates and Sarkar, 2006). For the presence of incomplete closures, a binary dependent variable, we use logistic mixed-effects linear regression. In all analyses, speaker is considered as a random factor. Word type was not included in the analyses as a random factor, since many of the used word types had only one or two tokens in the dataset.

A regression table is provided for each analysis (except for the analyses of F1 values, see below). Each table contains an intercept corresponding to a baseline combination of levels indicated in the caption (e.g. French unaccented /k/) and additional terms corresponding to levels not represented by the intercept (e.g. /p/, Spanish, accented). The regression coefficients for each term ( $\beta$ ) indicate deviations from the intercept. These coefficients are in ms for temporal dependent variables, in dB/s for

Term	Level	$\beta$	$z$	$p$
Intercept		-3.36		
Language	Spanish	3.17	6.77	< .0001
Stop type	/p/	-0.96	-2.32	< .05
Stop type	/t/	-0.79	-1.89	
Accent	accented	-0.69	-4.37	< .0001
Position	medial	1.56	3.61	< .0005
Language * Position	Spanish, medial	-1.12	-2.47	< .05
Language * Stop type	Spanish, /p/	-0.47	-1.04	
Language * Stop type	Spanish, /t/	-0.98	-2.14	< .05

Table 6.4: Regression coefficients with  $z$  and  $p$  values for the model predicting the likelihood of incomplete closures. The intercept corresponds to French, unaccented, word-initial /k/ consonants.

maximum intensity velocities, and in Hz for the models predicting F1. Only statistically significant predictors (with all their levels, whether significant or not) are kept in the models and shown in the tables. For categorical predictors with more than two levels, pairwise comparisons not involving the baseline level were done by rerunning the model with new baselines. These comparisons are not provided in the tables, but are mentioned in the text whenever they are statistically significant.

### 6.3.2 Voiceless stops

In the analysis of voiceless stops, we first investigate the likelihood of incomplete closures, and intensity velocity maxima in the CV transitions, then closure duration for those consonants with a complete closure, and finally the occurrence of intervocalic voicing.

#### Incomplete closures

In order to quantify differences in the likelihood of incomplete closures in Spanish and French voiceless stops, we fitted a logistic regression model with closure degree as response (complete vs. incomplete), speaker as random factor, and language (Spanish vs. French), stop type (/p/ vs. /t/ vs. /k/), accent (unaccented vs. accented), position in the word (initial vs. medial), previous vowel height, and following vowel height (high vs. mid vs. low) as fixed factors. Table 6.4 shows the regression results for this model.



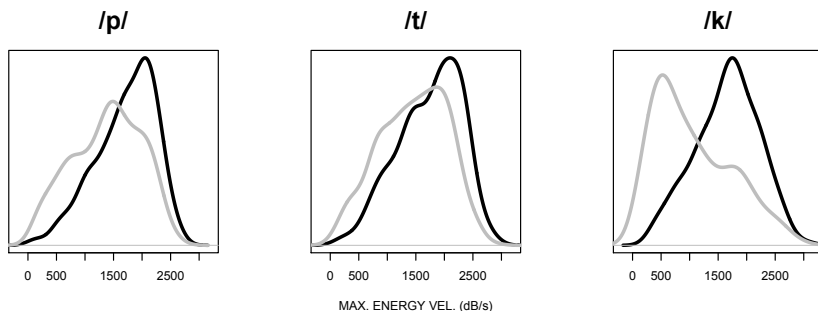


Figure 6.4: Density plots of maximum intensity velocity in CV transition for French (black) and Spanish (grey) stop consonants.

Spanish tended to have significantly more incomplete closures than French (25.5% vs. 5.4%;  $p < .0001$ ). This result confirms our expectation that Spanish voiceless stops are more likely to have reduced closures than French voiceless stops (see Section 6.1.1). This language effect interacted with consonant type in that the difference between the two languages was not so marked for consonant /t/ (15.1% vs. 5%;  $p < .05$ ). Language also interacted with position in the word: French tended to have more incomplete closures in word-medial than in word-initial position ( $p < .0005$ ), but no such tendency was observed for Spanish (in this language incomplete closure occurred equally often in word-initial and word-medial positions). Finally, in both languages, incomplete closures were less common for consonant /p/ than for consonant /k/ ( $p < .05$ ), and in accented than in unaccented syllables ( $p < .0001$ ).

### Maximum intensity velocity in CV transition

We then investigated language differences in maximum intensity velocity during the transition of the voiceless stop into the following vowel. Reduced stops should be characterized by values close to zero (i.e. by very gradual rises in intensity during the CV transition), while stops with clear closures and bursts should yield highly positive values. Figure 6.4 shows kernel density plots<sup>2</sup> of this measure for each stop type and language. It can be seen that Spanish stops tend to have lower values

<sup>2</sup>Kernel density plots display the estimated probability density function (y-axis) of a continuous random variable (x-axis), and have a purpose similar to that of histograms. However, whereas histograms group observations into a discrete number of bins, kernel density plots provide a continuous estimate of the distribution of a variable. The kernel density plots shown here were computed using the *density* function in R with default parameters. Note that the area under each curve integrates to 1. For further details, see the R manual (R Development Core Team, 2008) and Sarkar (2008).

and therefore less abrupt intensity changes in their CV transitions than French stops. Most noticeably, a majority of the Spanish instances of /k/ in our data appear to be particularly weak relative to their French counterparts.

Importantly, it can also be seen from these plots that, even in the case of /k/, a significant portion of Spanish stops had values similar to those most common of French stops. Thus, our data indicate that Spanish stops are not simply more reduced than French stops in general, but rather show more variability than French stops in that they become weakened significantly more often.

In order to examine whether these differences in our intensity velocity measures were statistically significant and not caused by effects of accent and position in the word, we fitted a regression model with maximum intensity velocity as response, speaker as random factor, and language (Spanish vs. French), stop type (/p/ vs. /t/ vs. /k/), accent (unaccented vs. accented) and position in the word (initial vs. medial) as fixed factors.

The resulting model is given in Table 6.5. A main effect of language and an interaction between language and stop type confirmed the impression given by Figure 6.4 that Spanish and French differ in maximum intensity velocity and that this difference is bigger for /k/ than for the other two consonant types<sup>3</sup>. This is an agreement with the results of the previous subsection, which indicated that the two languages differed especially in the realization of /k/ closures. Both accent and position in the word had the expected effects on maximum intensity velocity: accented and word-initial stops tended to have more abrupt intensity rises than unaccented and word-medial stops (respectively  $< .0001$  and  $< .005$ ).

### Closure duration

In order to test for language differences in closure duration, we ran a linear regression model with closure duration as response, speaker as random factor, and language, stop type, accent, and position in the word as fixed factors. Only consonants with a full closure ( $n = 1734$ ) were included in this analysis.

Table 6.6 shows the results of this model. Spanish tended to have shorter closures than French ( $p < .0001$ ). Spanish stops are not only more reduced than French stops in constriction degree, as shown in the two previous subsections, but also in the temporal dimension. As in the analysis of incomplete closures (Section 6.3.2), an interaction was found between language and position in the word. Consonants

<sup>3</sup>The editors note that this difference might be due to an unbalanced vocalic context in the datasets of the two languages. For instance, if /ku/ were more frequent, and /ki/ less frequent, in the Spanish than in the French dataset, the observed difference might be due to the sample rather than to language-specific reduction patterns. Inspection of the data showed that in fact, /ku/ was more frequent, and /ki/ less frequent, in the French than in the Spanish dataset. We therefore do not think that an unbalanced sample is the cause of the observed difference between the two languages

Term	Level	$\beta$	$t$	$p$
Intercept		1647.52		
Language	Spanish	-590.13	-8.43	< .0001
Stop type	/p/	63.96	1.24	
Stop type	/t/	51.97	0.9	
Accent	accented	176.1	4.9	< .0001
Position	medial	-121.81	-3.21	< .005
Language * Stop type	Spanish, /p/	262.66	3.84	< .0005
Language * Stop type	Spanish, /t/	329.96	4.82	< .0001

Table 6.5: Regression coefficients with  $t$  and  $p$  values for the model predicting intensity velocity maxima in CV transitions. The intercept corresponds to French, unaccented, word-initial /k/ consonants.

Term	Level	$\beta$	$t$	$p$
Intercept		50.6		
Language	Spanish	-11.62	-6.07	< .0001
Stop type	/p/	14.26	14.43	< .0001
Stop type	/t/	5.90	5.85	< .0001
Accent	accented	5.08	5.56	< .0001
Position	medial	-7.15	-5.31	< .0001
Language * Position	Spanish, medial	7.19	4.43	< .0001

Table 6.6: Regression coefficients with  $t$  and  $p$  values for the model predicting closure duration. The intercept corresponds to French, unaccented, word-initial /k/ consonants.

Term	Level	$\beta$	$t$	$p$
Intercept		10.96		
Closure duration		0.19	9.18	< .0001
Language	Spanish	2.11	1.22	
Stop type	/p/	3.9	3.09	< .005
Stop type	/t/	-2.05	-1.6	
Language * Stop type	Spanish, /p/	8.69	4.88	< .0001
Language * Stop type	Spanish, /t/	7.7	4.32	< .0001

Table 6.7: Regression coefficients with  $t$  and  $p$  values for the model predicting intervocalic voicing duration. The intercept corresponds to French, unaccented, word-initial /k/ consonants.

tended to be shorter in word-medial than in word-initial position in French ( $p < .0001$ ), but not in Spanish. We also found effects of stop type and accent affecting both languages in the same degree. Stop closures were longer for /t/ than for /k/ ( $p < .0001$ ), as well as for /p/ than for /t/ ( $\beta = 8.35, t = 9.41, p < .0001$ ). Finally, accented stops had longer closures than unaccented stops in both languages ( $p < .0001$ ).

### Intervocalic voicing

We then examined whether and how the interval of voicing during the realization of phonologically voiceless stops differed between French and Spanish. A first look at the data showed that intervocalic voicing was more pervasive in Spanish than in French voiceless stops. Closures were completely voiced in 32.7% of the Spanish stops, and only in 8.5% of the French stops. Moreover, in Spanish, 61.8% of stop closures were voiced for at least 50% of their duration or more, while in French this was the case only for 31.8% of the stop closures.

We then investigated in more detail which variables predict the duration of intervocalic voicing. Since intervocalic voicing duration could not be measured for incomplete closures, we limited our statistical analysis to stops with complete closures. We focused on duration rather than on the proportion of intervocalic voicing within the closure, since the latter did not have a normal distribution. We considered closure duration as a covariate, and language, stop type, accent, and position in the word as potential predictors. Speaker was included as a random factor. Table 6.7 shows the results of this model. Spanish tended to have longer durations of intervocalic voicing than French, but only for consonants /p/ and /t/, as shown by the absence of a main effect and the statistical significance of the two interaction terms ( $p < .0001$  in both cases). In both languages, intervocalic voicing was longer for

/p/ than for /k/ ( $p < .005$ ) and /t/ ( $\beta = 5.96, t = 4.82, p < .0001$ ), presumably due to the aerodynamic characteristics of bilabial stops (i.e. longer build-ups of intraoral pressure; Ohala and Riordan, 1979).

We finally counted how many incomplete closures were voiceless at some point in their realization. For French, we inspected all stops with an incomplete closure in the dataset, while for Spanish we examined a random sample of 200. Twenty-five out of the 45 incomplete closures (55.5%) were voiceless at some point for French, while only two cases (1%) were not voiced throughout for Spanish. Taken together, these results show that the occurrence of voicing is significantly more common in Spanish than in French intervocalic voiceless stops.

### 6.3.3 Vowels

As explained in the Method section, the interval that we marked as a vowel includes the stretch of signal between two stop closures, corresponding to the interval during which the vocal tract was not completely constricted. The stop release and following voiceless stretch of signal preceding the onset of voicing are therefore included in this interval. We examine language differences in vowel duration, in the occurrence and duration of voicelessness within this vowel interval, and finally in F1.

#### Duration

In order to investigate language differences in vowel duration, we fitted a regression model with vowel duration as response, speaker as random factor, and language, vowel type, and accent as fixed predictors. Since the vowel interval contained the release of the preceding stop, the identity of the preceding consonant was also included in the model as a predictor. The results of this model are shown in Table 6.8. In general, Spanish vowels tended to be shorter than French vowels ( $p < .0001$ ), especially /i/ vowels ( $p < .005$ ). As expected, accented vowels tended to be longer than unaccented vowels for both languages ( $p < .0001$ ), but this difference was more pronounced in French than in Spanish, as shown by the significant interaction term in Table 6.8 ( $p < .0001$ ). Finally, vowels tended to be shorter if preceded by consonant /p/ in French ( $p < .005$ ), but not in Spanish. These findings indicate that Spanish vowels tend to be slightly shorter than French vowels. At the temporal level, then, our initial expectation that French would exhibit more vowel reduction than Spanish is not borne out by the data.

#### Voicing

We then investigated whether the extent of voicelessness within the interval marked as a vowel (which included the consonantal release) differed between Spanish and

Term	Level	$\beta$	$t$	$p$
Intercept		92.46		
Language	Spanish	-17.56	-5.02	< .0001
Vowel type	/e/	-7.38	-2.47	< .05
Vowel type	/i/	-0.79	-0.26	
Vowel type	/o/	-6.17	-2.53	< .05
Vowel type	/u/	-11.97	-4.18	< .0001
Accent	accented	28.61	12.81	< .0001
Stop type	/p/	-9.6	-4.37	< .0001
Stop type	/t/	-1.98	-0.79	
Language * Vowel type	Spanish, /e/	-5.98	-1.46	
Language * Vowel type	Spanish, /i/	-13.55	-3.18	< .005
Language * Vowel type	Spanish, /o/	-2.19	-0.61	
Language * Vowel type	Spanish, /u/	3.21	0.75	
Language * Accent	Spanish, accented	-18.46	-6.07	< .0001
Language * Stop type	Spanish, /p/	10.89	3.29	< .005
Language * Stop type	Spanish, /t/	1.87	0.52	

Table 6.8: Regression coefficients with  $t$  and  $p$  values for the model predicting vowel duration. The intercept corresponds to French, unaccented /a/ vowels preceded by a /k/ consonant.

Term	Level	$\beta$	$t$	$p$
Intercept		8.34		
Vowel duration		0.19	7.83	< .0001
Language	Spanish	-5.43	-2.78	< .001
Stop type	/p/	-4.29	-3.72	< .0005
Stop type	/t/	0.73	0.58	
Vowel type	/e/	-0.2	-0.1	
Vowel type	/i/	24.13	10.41	< .0001
Vowel type	/o/	0.04	0.02	
Vowel type	/u/	11.49	5.68	< .0001
Language * Vowel type	Spanish, /e/	-0.93	-0.33	
Language * Vowel type	Spanish, /i/	-23.5	-7.51	< .0001
Language * Vowel type	Spanish, /o/	0.3	0.12	
Language * Vowel type	Spanish, /u/	-11.28	-3.80	< .0005

Table 6.9: Regression coefficients with  $t$  and  $p$  values for the model predicting voiceless vowel duration. The intercept corresponds to French /a/ vowels preceded by a /k/ consonant.

French. We found that in French, 24 out of 451 vowels were completely devoiced. Closer inspection revealed that the majority of these vowels were high unaccented vowels ( $n = 21$ , 19.6% of high unaccented vowels in the French dataset). In Spanish, on the other hand, we found only one case of complete devoicing.

Differences in the extent of partial devoicing could not be investigated with the proportion of voicelessness within the vowel interval, since this measure did not present a normal distribution. Alternatively, we fitted a regression model with voiceless vowel duration (the absolute duration of the voiceless stretch within the vowel interval) as response, speaker as random factor, vowel duration as a covariate, and language, vowel type, preceding stop type, and accent as predictors (c.f. the analysis of intervocalic voicing in Section 6.3.2). Importantly, the use of vowel duration as a covariate allows us to test for relative differences in the duration of voicelessness within the vowel interval. The results of this model are shown in Table 6.9. In accordance with the cross-linguistic tendency for /k/ to have longer Voice Onset Time (VOT) than /p/ and /t/ (Cho and Ladefoged, 1999), voiceless vowel duration, a measure related to VOT, was longer for consonant /k/ than for consonant /p/ ( $p < .0005$ ). More importantly, there was a main effect of language and an interaction between language and vowel type. French had longer voiceless vowel durations than Spanish, especially for the high vowels /i/ and /u/ (/i/:  $p < .0001$ ; /u/:  $p < .0005$ ).

Together, the counts of cases of complete devoicing and the analysis of voiceless vowel duration indicate that vowel devoicing occurs almost exclusively in high French vowels.

### Formant values

We finally investigated whether French and Spanish vowels surrounded by voiceless stops differed in F1. A regression model was fitted with F1 as the response variable, language, speaker gender, accent, and vowel type as predictors, and speaker as a random factor. A four-way interaction between language, accent, vowel type, and gender was found for F1 ( $F(4, 1041) = 2.64, p < .05$ ). In order to interpret this interaction, we split the dataset by vowel type and fitted models with language, accent, and gender as predictors. Instead of providing the details of these five models in separate tables, we summarize our main findings. First, all vowel types had statistically higher F1 in Spanish than in French (/a/:  $\beta = 89.32, t = 4.26, p < .0001$ ; /e/:  $\beta = 111.36, t = 10.03, p < .0001$ ; /i/:  $\beta = 108.16, t = 9.76, p < .0001$ ; /o/:  $\beta = 147.89, t = 16.88, p < .0001$ ; /u/:  $\beta = 121.19, t = 9.79, p < .0001$ ). This trend can be seen clearly in Figures 6.5 and 6.6, which shows boxplots of F1 values for females and males separately.

Second, accented vowels tended to have higher F1 values than unaccented vowels both in French and Spanish, but this effect was statistically significant only for the

following combinations of language, vowel and gender: /e/ and /o/ vowels for both genders and languages (/e/:  $\beta = 32, t = 3.64, p < .0005$ ; /o/:  $\beta = 147.89, t = 16.88, p < .0001$ ), /a/ vowels in French male and female speakers, and for Spanish females (main effect of accent:  $\beta = 53.08, t = 2.45, p < .05$ ; three-way interaction of accent, accent and gender:  $\beta = -136.50, t = -2.93, p < .005$ ), and French high vowels (main effect of accent for /i/:  $\beta = 54.57, t = 4.1, p < .0001$ ; interaction of language and accent for /i/:  $\beta = -61.61, t = -3.78, p < .0001$ ; main effect of accent for /u/:  $\beta = 38.59, t = 2.87, p < .01$ ; interaction of language and accent for /u/:  $\beta = -34.63, t = -2.12, p < .05$ ). It is worth noting that the differences observed between Spanish and French are not merely statistically significant, but actually of a considerable size. For instance, the boxplots<sup>4</sup> in Figures 6.5 and 6.6 show that numerous tokens of French /a/ and /e/ vowels had F1 values respectively comparable to those of many Spanish /e/ and /i/ vowels.

In summary, we found that all French vowels had significantly lower F1 values than Spanish vowels, suggesting that they tend to be articulated with a more constricted vocal tract than their Spanish counterparts. Accent showed a consistently similar effect in both languages whenever it reached statistical significance.

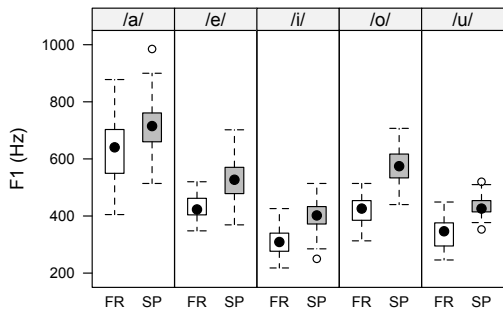


Figure 6.5: Boxplots of F1 values for French (white) and Spanish (grey) vowels produced by female speakers.

<sup>4</sup>The rectangle in a boxplot contains the middle 50% of the data points (the interquartile range, IQR). The whiskers include data points beyond this range, extending up to a length of 1.5 the IQR. Data points which are beyond 1.5 times the IQR are represented with circles.



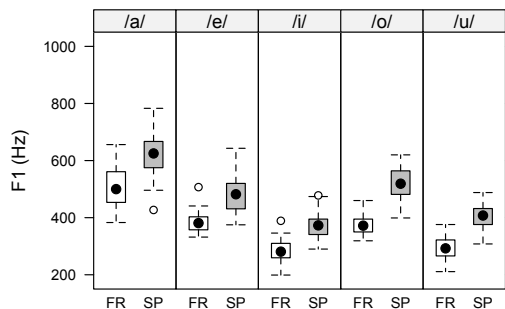


Figure 6.6: Boxplots of F1 values for French (white) and Spanish (grey) vowels produced by male speakers.

## 6.4 Discussion and conclusion

The present study has revealed significant differences between French and Spanish in the realization of intervocalic voiceless stops and vowels surrounded by voiceless stops. In intervocalic position, Spanish voiceless stops tend to have shorter stop closures, display incomplete closures more often, and exhibit more voicing than French voiceless stops. In Spanish, two major characteristics of voiceless stops according to standard phonetic descriptions (oral occlusion and absence of glottal vibration) are often compromised in intervocalic position: a quarter of the examined stops lacked a complete closure (i.e. were realized as approximants), and if stops with complete and incomplete closures are taken together, practically half of the examined Spanish stops are voiced throughout their entire closure. These reduction phenomena were also found in French voiceless stops, but to a much lesser extent. Regarding vowels, we found more cases of complete devoicing and greater degrees of partial devoicing in French than in Spanish, in spite of the fact that French vowels were generally longer than their Spanish counterparts. We also found that vowels tend to have significantly lower F1 values in French than in Spanish, suggesting that, at least in the context of stop consonants, French vowels are generally produced with a more constricted vocal tract than Spanish vowels.

In general, accent appeared to affect French and Spanish voiceless stops and vowels in the same way. In both languages, incomplete closures are rarer in accented than in unaccented stops. Similarly, for both languages stop closures tend to be longer and intensity velocity maxima during CV transitions tend to be higher in accented than in unaccented syllables. These results are in agreement with numerous studies showing

that consonants in accented syllables tend to be realized with greater articulatory effort and enhanced acoustic cues (Beckman and Edwards, 1994; Cho and McQueen, 2005; Cole et al., 2007; Keating, 2006, among many others). Finally, in both languages unaccented vowels tended to have lower F1 values than accented vowels, in agreement with the findings of Lindblom (1963) for Swedish. Thus, vowel reduction in the examined context did not result in centralization, but rather into less open articulations. The fact that accent had a roughly similar effect for French and Spanish vowels and stops indicates that the direction of the examined reduction processes is not essentially different within each segment type and language.

Position in the word appeared to affect voiceless stops more consistently in French than in Spanish. An effect of position in the word on maximum intensity velocities was found for both languages, but only French voiceless stops tended to be longer and maintain a full closure more often in word-initial than in word-medial position. This may be attributed to the different prosodic systems of these two languages. Since French lacks lexical stress, speakers of this language may opt to mark word boundaries with strengthened articulation in order to facilitate word segmentation for hearers. Spanish speakers, on the other hand, might tend to focus articulatory effort on segments in lexically stressed syllables, since strengthening word-initial unstressed syllables might wrongly signal a stressed syllable and compromise word recognition. Further research would be needed to test this hypothesis.

A particular aspect of the sequences of voiceless stops and vowels analyzed in this study is that they involve antagonistic articulatory gestures at the supraglottal and glottal levels. While voiceless stops require an absence of glottal vibrations and a maximally constricted vocal tract, vowels require the production of voicing and an unconstricted vocal tract. In continuous speech, supraglottal and glottal gestures in these sequences must be blended in some way, and crosslinguistic differences in the “blending strength” (Fowler and Saltzman, 1993) of each of these gestures could appear. Interestingly, our data are consistent with the hypothesis that French and Spanish speakers use asymmetrical coarticulatory strategies when confronted with these segmental sequences. The degree of constriction and voicing specifications of voiceless stops are generally realized faithfully in French, but French vowels are characterized by low F1 values and often become devoiced. On the other hand, Spanish voiceless stops appear to be greatly coarticulated to adjacent vowels in terms of degree of constriction and voicing. We may hypothesize then that, in terms of constriction degree and voicing, voiceless stops have a higher degree of coarticulatory resistance in French than in Spanish, and, conversely, that vowels have a higher degree of coarticulatory resistance in Spanish than in French.

We can also try to explain our findings by invoking general differences in base of articulation (Honikman, 1964; Disner, 1983; Bradlow, 1995) or specific differences

in the phonetic targets and dynamics of stops and vowels in French and Spanish. French vowels and stops might all be characterized by more constricted articulatory targets than Spanish vowels and stops, regardless of the context and the speech register in which they are produced. In support of this hypothesis, Meunier et al. (2003) found that vowels spoken in isolated syllables (e.g. /si/, /mi/, /li/) tended to have lower F1 in French than in Spanish and English. In an older study, however, Delattre (1969) did not find significant differences in F1 between the Spanish and French vowel space. The results of both of these studies, however, should be considered with caution, since they used a small number of speakers.

As for voiceless stops, no explicit comparison of careful speech between French and Spanish has been performed to our knowledge. Nevertheless, it can be noted that the distributions of velocity maxima in CV transitions shown in Figure 4 suggest that, at least at the acoustic level, an important proportion of Spanish stops are similar to their French counterparts. It should be also noted that an explanation entirely based on interlanguage differences in the specification of consonantal constriction degree and vowel openness leaves unanswered our findings regarding the realization of voicing in phonologically voiceless stops. While the extremely constricted nature of French high vowels might plausibly contribute to their frequent devoicing for aerodynamic reasons, the observed interlanguage differences in the occurrence of voicing in phonologically voiceless stops can only be explained by also positing differences in the size of glottal devoicing gestures. For all these reasons, further research is needed to assess the extent to which language-specific bases of articulation or phonetic targets can explain our observations.

In summary, our findings indicate that the extent of reduction that a given segment type can undergo in conversational speech can vary *significantly* across languages, even in the case of two related languages such as French and Spanish. In terms of constriction degree and voicing, Spanish intervocalic voiceless stops appear to be assimilated to adjacent vowels in a greater degree than French voiceless stops. On the other hand, French vowels show more signs of coarticulatory influence from neighboring voiceless stops than Spanish vowels. Based on these observations, we propose that asymmetrical reduction patterns may result from the fact that the supraglottal and glottal gestures of voiceless stops and vowels are attributed different degrees of coarticulatory resistance in each language. Language differences in general “basis-of-articulation” properties or in the exact specification of phonetic targets for voiceless stops and vowels also need to be considered as possible causes of the observed differences in reduction patterns.



# Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish

---

## Chapter 7

This chapter is a reformatted version of:  
Francisco Torreira and Mirjam Ernestus. Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. Submitted.

This study describes the weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish, and investigates the role of several of its potential conditioning factors, including morphological, lexical and probabilistic variables. This sort of weakening of intervocalic /s/ is characterized by the presence of voicing, the loss of high-frequency frication and temporal reduction. In order to model the variability observed in our vowel-/s/-vowel sequences, three acoustic parameters were examined: the occurrence of uninterrupted voicing, the difference in high-band (4-8 KHz) intensity between /s/ and the following vowel, and the duration of the dip in low-band (0-1.5 KHz) intensity corresponding to /s/. We observed that over a third of the /s/ consonants exhibited uninterrupted voicing. All three examined parameters were highly sensitive to temporal factors, and less consistently, to prosodic factors such as word-position and stress. Furthermore, we found that /s/ suffixes in redundant morphosyntactic contexts were more prone to weakening than other word-final /s/ segments. Several high-frequency words were particularly prone to weakening, but no general probabilistic effects were observed for factors previously claimed to favor reduction such as word frequency, contextual predictability or grammatical class. These findings demonstrate the complex variability characteristic of reduction phenomena in spontaneous speech.

## 7.1 Introduction

### 7.1.1 *Speech reduction and goal of the present study*

The reduction of consonants and vowels is pervasive in spontaneous conversational speech, as attested in a rising number of corpus studies (e.g. Kohler, 2000; Ernestus, 2000; Johnson, 2004; Adda-Decker et al., 2005; Schuppler et al., 2011; Adda-Decker and Snoeren, 2011; Bürki et al., 2011, among others). In spite of the increasing interest in speech reduction phenomena (e.g. recent special issue of *Journal of Phonetics*, vol. 39), much remains to be explored about their phonetic characteristics and conditioning factors. For instance, many corpus studies have focused on the effects of probabilistic, morphological, and lexical factors on word and segmental durations, but studies examining the effects of such factors on other phonetic parameters and on specific reduction phenomena are scarce.

The present work describes the weakening of intervocalic /s/ in Spanish, an almost undocumented phenomenon despite its high frequency in conversational speech, and investigates its conditioning factors. During the initial inspections of the Nijmegen Corpus of Casual Spanish (NCCSp) (Torreira and Ernestus, 2010a), a recently created corpus of Madrid Spanish, we observed that intervocalic /s/ consonants, although always realized as sibilants, often exhibit uninterrupted voicing and little frication noise. This is illustrated in Figures 7.1 to 7.3, which show utterances extracted from this corpus displaying different degrees of weakening. In contrast to the aspiration and deletion of syllable-final /s/ ( $s > h > \emptyset$ ) in coastal Latin American and southern Spanish dialects, a reduction process which has been the object of numerous studies (see extensive bibliography in Hernández-Campoy and Trudgill (2002)), the sort of /s/ reduction that we investigate here has only been described impressionistically in the philological literature for a few rural dialects of Castilian Spanish (Torreblanca, 1986). This sort of /s/ reduction does not involve a complete loss of the oral constriction as in /s/-aspiration, and is referred to as /s/-weakening from now on for the sake of simplicity. We describe it as it occurs in Madrid Spanish, using quantitative acoustic phonetic methods.

In Study 1, we provide an initial overview of the phenomenon by estimating the frequency with which it occurs, and by examining the extent to which it is influenced by factors known to condition coarticulatory and reduction phenomena such as speech rate, stress, position in the word, number of syllables in the word and the segmental context. In Studies 2 and 3, we explore the roles of morphosyntactic, probabilistic and lexical factors after controlling for relevant conditioning factors identified in Study 1. Since /s/ is the most frequent consonant in Spanish (Moreno Sandoval et al., 2008), occurs in a wide variety of function and content words, and often carries morphosyntactic information as a plural marker or verbal suffix (e.g.

*las casas* ‘the houses’; *vas* ‘you go’), /s/-weakening provides an ideal case for investigating the roles of these variables in speech reduction.

So far, most studies investigating potential conditioning factors of reduction phenomena (e.g. content versus function word distinction, lexical frequency and contextual predictability) have used global measures of reduction such as word, morpheme or segmental durations, on the basis that they should capture most kinds of reduction in a dataset (e.g. deletions, vowel reduction, reduced articulatory strength). In the present study we adopt a different strategy, and investigate whether the effects of such variables can also be observed on more fine-grained acoustic measures in a specific reduction phenomenon. In all three studies, we examine three acoustic parameters related to the articulation of /s/: the occurrence of uninterrupted voicing, the difference in high-band (4-8 KHz) intensity between /s/ and the following vowel, and the duration of the dip in low-band (0-1.5 KHz) intensity corresponding to /s/.

### **7.1.2 Realizations of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish**

This subsection illustrates the variability in the realization of intervocalic /s/ present in the Corpus of Casual Spanish (NCCSp; see Section 7.1.6 below) with examples extracted from this corpus. Figure 7.1 shows canonical and slightly weakened realizations of intervocalic /s/. The example at the top illustrates a canonical pronunciation containing the acoustic features expected for an intervocalic /s/. Voicing is interrupted for around 80 ms, as shown by the lack of a voicing bar and periodic energy in the time region of the spectrogram corresponding to the fricative consonant. In this time region energy drops in the lower part of the spectrum, reflecting the formation of an oral constriction, and it sharply increases above 3.5 KHz, reflecting the creation of turbulence at the constriction. It is this prominent concentration of aperiodic energy in mid-high regions of the spectrum that gives /s/ its characteristic sibilant quality. The bottom example in Figure 7.1 displays a similar pattern in the mid-high frequencies. However, in this example voicing is present throughout much of the realization of the fricative, and gradually disappears towards its mid point, before picking up as the transition into the upcoming vowel approaches. Moreover, one can also see a less sharp decrease in low-frequency energy than in the top example, resulting perhaps from a less narrow oral constriction and the presence of voicing.

The examples in Figure 7.2 show further weakened realizations of intervocalic /s/. Frication noise in the mid-high frequencies is less prominent than in the examples in Figure 7.1, and in both cases voicing continues uninterrupted throughout the /s/ consonants with no visible signs of decay. Moreover, the duration of the consonants, as estimated informally from the amplitude envelope in the waveforms and the occurrence of aperiodic energy in the spectrograms, appears to be significantly shorter

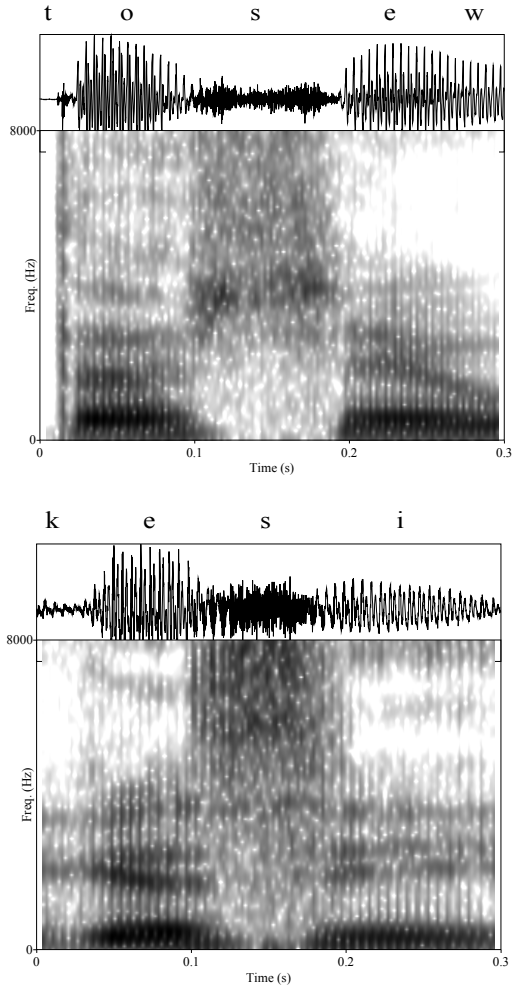


Figure 7.1: Waveforms and spectrograms of utterances containing canonical (top) and slightly weakened (bottom) realizations of intervocalic /s/ in spontaneous Madrid Spanish. The example at the top has been extracted from the utterance *ochocientos euros* ‘eight hundred euros’. The second example has been extracted from the utterance *que sí* ‘definitely yes’.



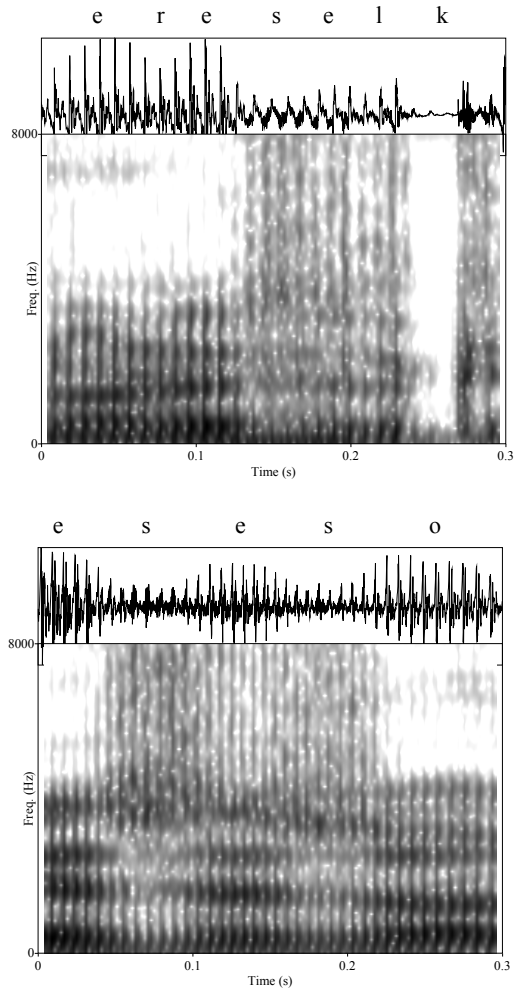


Figure 7.2: Waveforms and spectrograms of utterances containing weakened intervocalic /s/ in spontaneous Madrid Spanish. The example at the top has been extracted from the utterance *tú no eres el que [...]* ‘you are not the one who [...]’. The second example has been extracted from the utterance *es eso* ‘that’s it’.

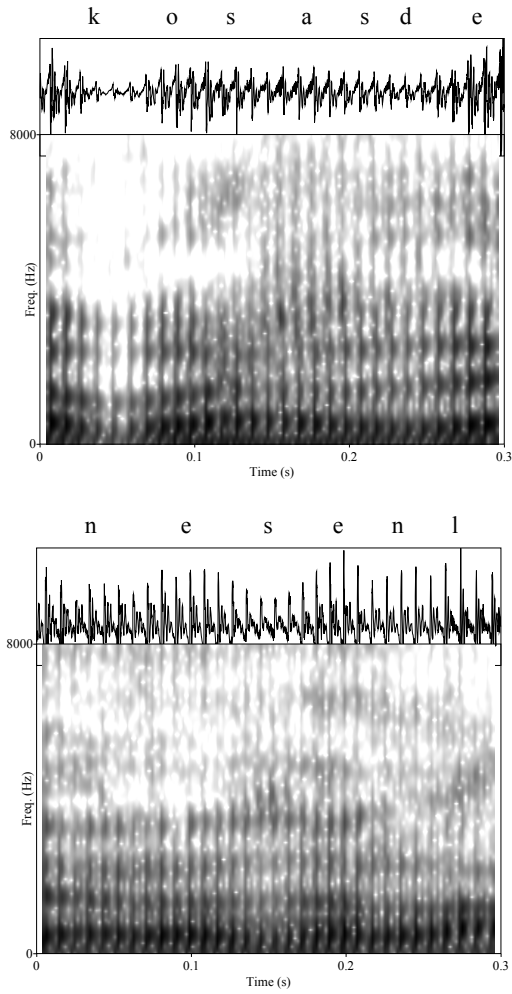


Figure 7.3: Waveforms and spectrograms of utterances containing extremely weakened realizations of intervocalic /s/ in spontaneous Madrid Spanish. The example at the top has been extracted from the utterance [...] *mazo cosas de* [...] ‘[...] a lot of things about [...]’. The second example has been extracted from the utterance *corrupción es en la* [...] ‘[...] corruption is in the [...]’.

than that of the examples in Figure 7.1. The examples in Figure 7.3 illustrate cases of extreme weakening. Frication noise in these examples is considerably weaker than in the previous ones, to the point that it is hardly visible. Voicing, as in the examples in Figure 7.2, continues uninterrupted throughout the vowel-/s/-vowel sequences.

These examples make it apparent that the familiar correlates of a voiceless fricative (aperiodic energy in the high frequencies and voicing offsets and onsets) are not always present in our data, and that, if present, they may be very weak. For this reason, segmental durations and related measures, such as the proportion of the fricative containing voicing, cannot be made in a consistent way across the tokens in our data. The study of /s/-weakening in spontaneous Madrid Spanish, therefore, must be based on other acoustic measures. Section 7.2.1 presents the measures that we employed, namely the occurrence of uninterrupted voicing, the difference in high-band (4-8 KHz) intensity between /s/ and the following vowel, and the duration of the dip in low-band (0-1.5 KHz) intensity corresponding to /s/.

### **7.1.3 Possible articulatory mechanisms behind /s/-weakening**

The voicing and weak frication in Spanish intervocalic /s/ illustrated in the previous subsection may result from different mechanisms. Given that the segments surrounding /s/ require the production of voicing and that the intervening /s/ requires interrupting it, it is plausible that voicing in intervocalic /s/ is a weakening phenomenon. Generally speaking, voicing in intervocalic /s/ is likely to arise from the conflicting demands of adjacent segments on glottal articulation, especially in cases in which the glottis must adopt opposite configurations back and forth in a reduced amount of time (i.e. at fast speech rates).

The high-frequency frication characteristic of /s/ consonants has its source in a jet of air passing through a narrow coronal constriction. This jet of air hits the incisors, creating an additional turbulent noise source, and enters a small front cavity between the constriction and the lips with a high-frequency resonance. The weakened frication observed in many of our /s/ tokens (Figures 7.2 and 7.3) can be attributed to at least two mechanisms. First, the supraglottal constriction may not be narrow enough for the creation of turbulence. This could be caused by the contextual influence of adjacent vowel sounds with constriction targets specifying an open vocal tract. Second, air volume velocity at the supraglottal constriction may not be sufficient for the creation of strong turbulent noise. This could be the consequence of the glottal weakening discussed in the previous paragraph. During the production of intervocalic /s/, the glottis may not always reach a fully abducted position under the coarticulatory influence of adjacent vowels. This will impede airflow and hinder the creation of turbulence at the downstream supraglottal constriction.

In the present work we do not intend to investigate the exact articulatory mechanisms (e.g. gestural blending, truncation, carry-over or anticipatory strategies, or even the suppression of the glottal abduction gesture) leading to the frequent weakening of intervocalic /s/ in Madrid Spanish. To do this we would need articulatory and aerodynamic data collected under more highly controlled conditions. In Study 1, where we present an exploratory overview of the phenomenon, we will just examine the extent to which /s/-weakening is sensitive to factors known to condition coarticulation, and more generally, speech reduction. We expect /s/-weakening to be favored at higher speech rates, in temporally reduced realizations, and in prosodically weaker positions (i.e. unstressed position, in non-word-initial position, in longer words).

### 7.1.4 Morphological effects on speech reduction

Recent studies indicate that reduction does not affect all words and morphemes to the same degree. For instance, it has been claimed that segments which carry semantic information by themselves, such as one-segment suffixes, tend to be less reduced than segments which are part of a stem. For example, several studies have reported higher rates of deletion for English word-final /t/ and /d/ consonants when they are part of a stem (as in *mist*) than when they are a verbal suffix (as in *missed*; Guy, 1994; Bybee, 2002; Labov, 2004). Such differences could arise either from morphological effects on phonological encoding percolating down to articulation, or, indirectly, via mechanisms such as those proposed by van Son and Pols (2003), Pluymaekers et al. (2005), and Aylett and Turk (2004), in which the speaker modulates articulatory effort according to the informativeness of the speech signal.

Regarding /s/-weakening, it should be noted that word-final /s/ in Spanish can occur as a plural marker in determiners, nouns, and adjectives (e.g. *las niñas bonitas* ‘the pretty girls’), as a verbal suffix (e.g. *come* ‘she eats’ versus *comes* ‘you eat’), and as part of a larger morpheme, either a root or a suffix (*miércoles* ‘Wednesday’; *vamos* ‘we go’, where the final /s/ is part of the suffix *-amos*). In line with the morphological effects on final /t/ and /d/ deletion in English mentioned above, it could be hypothesized that, in Spanish, /s/ suffixes (e.g. *comes* ‘you eat’) are less prone to reduction than /s/ segments that are part of a larger morpheme (e.g. *miércoles* ‘Wednesday’). However, this hypothesis has not been confirmed in several variationist studies on the realization of /s/ in several Latin American Spanish dialects (Poplack, 1980; Hundley, 1987; Lafford, 1989). Instead, several of these studies have found the opposite: word-final /s/ is aspirated and deleted more often when it is a suffix than when it is part of a stem. A possible explanation for this is that /s/ suffixes in Spanish are often redundant based on the morphological, semantic, or syntactic information in the utterance (e.g. *cuatro cosas* ‘four things’). Poplack (1980), for Puerto Rican Spanish, and Hundley (1987), for Peruvian Spanish, found support for this explanation in that

/s/ suffixes in these dialects are more likely to be aspirated or deleted when they are morphosyntactically redundant given the context. Similarly, Erker (2010) found that the duration of sibilant realizations of /s/ in Dominican Spanish tends to decrease when the target word is preceded by other words with -s plural suffixes in the same noun phrase.

In Study 2, we test if the weakening of word-final /s/ consonants in Madrid Spanish is affected by morphological factors such as the ones discussed above. In particular, we will examine if /s/-weakening in Madrid Spanish is affected by morphosyntactic contextual redundancy or if it is sensitive to the morphological structure of words in a more direct way.

### **7.1.5 Probabilistic and lexical effects on speech reduction**

A series of studies on English and Dutch have shown that there is a tendency towards more reduction in highly frequent and predictable words, in function words compared to content words, and in suffixes highly predictable from their stems (e.g. Bell et al., 2003, 2009; Aylett and Turk, 2004, 2006; Pluymaekers et al., 2005; Ernestus et al., 2006; Baker et al., 2011). These studies have examined a range of dependent variables, such as segmental transcriptions of full versus reduced forms (Bell et al., 2003), word duration (Bell et al., 2009; Baker et al., 2011), syllable duration (Aylett and Turk, 2004), vowel duration and spectral characteristics (Aylett and Turk, 2006), suffix duration and number of realized segments (Pluymaekers et al., 2005), and the duration and voicing of consonant clusters (Ernestus et al., 2006). Regarding Spanish, word frequency has been claimed to favor the aspiration and deletion of /s/ in several Latin American dialects (Brown and Cacoullos, 2003; Minnick Fox, 2006; File-Muriel, 2007; Brown, 2009). These probabilistic and lexical effects on articulation have been interpreted as the consequence of different speech production mechanisms. These include the repetition of frequent articulatory routines (Bybee, 2001), a modulation of articulatory effort sensitive to informativeness (Aylett and Turk, 2004; Pluymaekers et al., 2005), and lexical access, which is claimed to operate differently for function words and content words (Bell et al., 2009).

On the other hand, several studies suggest that probabilistic and lexical effects on speech reduction may not be ubiquitous. Recent work on the realization of /s/ in Caleño Spanish (File-Muriel and Brown, 2011) report longer /s/ durations the higher the lexical frequency, and higher spectral center of gravity values and percentages of voicelessness the higher the word bigram frequency of the target word and its following word (see regression coefficients in their Tables 7.1 and 7.2). Similarly, Kuperman et al. (2007) found that Dutch -s- interfixes were longer the more probable they were given the compound in which they occurred and its constituents. Finally, (Torreira and Ernestus, 2010a, see Chapter 5) investigated phrase-medial vowel de-

voicing in spontaneous French, and found that devoicing was as common in content words as in high-frequency function words. These observations, all of them on specific segments, suggest that the roles of factors such as lexical class and the frequency and predictability of linguistic units may be more complex than is often claimed in the literature. Since Spanish /s/ occurs in function as well as in content words representing a wide range of lexical frequencies, it offers an ideal case for further testing the role of these factors on articulation. Study 3 examines the influence of these factors on /s/-weakening.

### **7.1.6 The Nijmegen Corpus of Casual Spanish**

All of the speech data used in the studies presented in this paper come from the Nijmegen Corpus of Casual Spanish (NCCSp). This recently created corpus consists of recordings of 52 Spanish speakers (27 female and 25 male) engaged in casual conversations with friends. All speakers were university students in their late teens or twenties and had been raised in the Madrid region. The corpus, which features around 30 hours of high-quality audio and video recordings, has been orthographically annotated by professional transcribers at Verbio Speech Technologies S.L. The transcriptions of the corpus contain around 340,000 word tokens and 16,500 word types (distinct lexical items).

The NCCSp was collected in a sound-attenuated room at the Universidad Politécnica in Madrid during March 2009. In each recording session, which lasted around ninety minutes, we recorded a group of three friends of the same sex: two naive speakers and one confederate. The role of the confederate was to recruit the other two participants (the naive speakers), and to make the conversation more lively whenever it approached a dead end. Eight naive speakers took part in subsequent recordings as confederates. All participants wore a Samson QV head-mounted unidirectional microphone. The microphones were placed at an average distance of 5 cm from the left corner of the speakers' lips. The naive speakers were captured in separate channels of a stereo signal with an Edirol R-09 solid-state recorder. The confederate was directly recorded on a computer via a dedicated sound card. The sampling rate used was 44.1 KHz, while quantization was set to 16 bits.

The recording procedure used for the NCCSp was similar to that previously used for the Nijmegen Corpus of Casual French (Torreira et al., 2010, see Chapter 2). Each recording session consisted of three parts. In Part 1 we told the participants that the confederate's microphone did not work properly and asked the confederate to leave the room. While waiting, the naive speakers spontaneously engaged in free conversation, which was recorded unbeknownst to them. This part lasted around 20 minutes on average. Part 2 consisted of free conversation between the confederate and their friends, which lasted around 35 minutes. Part 3 required participants to

choose three questions from a list of general interest questions, and to negotiate a common position for their group in around 35 minutes. At the end of the recording, we disclosed our procedures to the participants and paid 30 euros to each of the speakers and 45 euros to the confederate as a compensation for their time. All of the participants signed a consent form agreeing to the use of the recordings for academic and scientific purposes. More detailed information about the corpus can be found online at <http://mirjamernestus.ruhosting.nl/Ernestus/NCCSp>, and in Torreira and Ernestus (2010a).

### **7.1.7 Outline of this article**

Sections 2-4 present three corpus studies carried out to investigate the occurrence of /s/-weakening in spontaneous Spanish. Study 1 provides an exploratory overview of this reduction phenomenon. We estimate its frequency and to what extent it is sensitive to several factors known to condition coarticulation and speech reduction in general. In Study 2, we examine the potential role of morphosyntactic factors on /s/-weakening. We investigate if word-final /s/ is more weakened when it is part of a stem than when it is a suffix, and, within the group of suffixes, if grammatically redundant suffixes show a bias towards more weakening. Finally, Study 3 investigates the potential role of lexical and probabilistic variables on /s/-weakening, including grammatical class (content versus function words), lexical frequency and contextual predictability, and word-specific effects.

## **7.2 Study 1: Exploratory overview of /s/-weakening**

### **7.2.1 Method**

#### **Materials**

As a first step, we randomly extracted from the NCCSp 1100 intervocalic /s/ consonants not disrupted by laughter, disfluencies, or intrusive overlapping speech. Then we sampled additional tokens meeting the same conditions until every speaker in the corpus was represented by at least 20 tokens. This dataset contained 1257 tokens and 593 word types. The average number of word tokens by word type was 3.3, with a standard deviation of 9.9 and a range extending from 1 to 93.

In order to roughly locate vowel-/s/-vowel sequences in the corpus, a forced segmental alignment of the NCCSp was performed with HTK automatic speech recognition software (Young et al., 2002). Using the orthographic transcription of the corpus and a dictionary of canonical pronunciations, a monophone system consisting of 25 Spanish phones and a silence phone was trained on the NCCSp. Each monophone

was represented by a 3-state left-to-right HMM model. The frame length of each state was 10 ms. Therefore, phones in the alignment were always assigned a duration of minimally 30 ms. This alignment was also used to estimate speech rate, an important variable in our analyses (see Section 7.2.1).

As explained in Section 7.1.2, the usual correlates of a voiceless fricative (aperiodic energy in the high frequencies and voicing offsets and onsets) were present in such varying degrees across tokens that they could not be used in a consistent way to determine the start and end of /s/ consonants. For this reason, we did not perform temporal measures using traditional segmentation procedures. Instead, we based our measurements on the energy profile of the vowel-/s/-vowel sequence. The following measures were made:

*Low-band (0-1.5 KHz) intensity dip duration:* During the initial inspection of the data, we noticed that all tokens exhibited a dip in energy in the lower frequencies in the part of the signal corresponding to /s/. Since this dip can be reasonably related to the formation and release of an oral constriction in all tokens, its duration is a useful measure with which we can estimate the temporal reduction of /s/ consonants. This duration measure cannot be considered as a *segmental* duration, and, accordingly, it cannot be used to compute related measures, such as the amount of voicing within the /s/ segment. This is not a technical limitation of our work, but a direct consequence of the fact that segments in spontaneous speech often do not present clear boundaries and exhibit more variable realizations than in controlled speech.

In order to measure the duration of the low-band intensity dip in the vowel-/s/-vowel sequence, we used the following procedure (illustrated by the solid curves and dotted vertical lines in the intensity panels of Figure 7.2): first, we low-pass filtered the signals below 1.5 KHz; second, we placed marks at the midpoints of the automatically segmented vowels, and then checked if the interval between them included the intensity velocity minima and maxima in the vowel-to-/s/ and /s/-to-vowel. In those cases in which this was not the case, we manually corrected the placement of the marks (but not the automatic segmentation). It should be noted that the exact location of these marks did not affect our measures as long as the interval between them included the aforementioned intensity velocity minima and maxima in the vowel-to-/s/ and /s/-to-vowel transitions; third, we identified the locations of the intensity velocity minimum and maximum in the vowel-to-/s/ and /s/-to-vowel transitions. These points, located where intensity below 1.5 KHz was decreasing and increasing most rapidly, were considered as the start and the end of the dip. The low-band intensity dip duration was then derived from these time points.

*Voicing:* The occurrence of voicing in the low-band intensity dip was estimated with the pitch detection algorithm available in Praat set to default parameters. We observed that most tokens showed either uninterrupted voicing, or a stretch of voice-



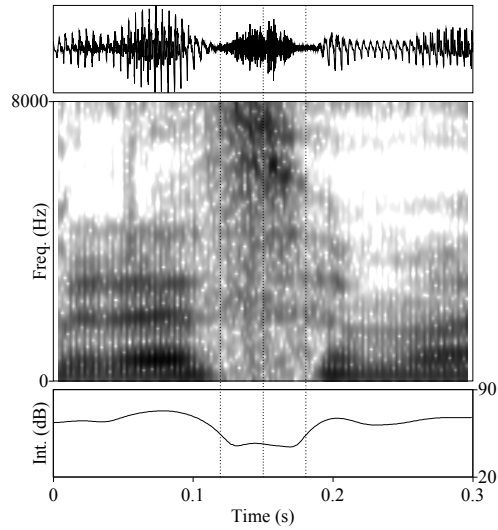
lessness encompassing the intensity dip. For this reason, the amount of voicing within the dip cannot be used to study gradient changes in voicing duration within /s/. We therefore created a categorical variable encoding the presence of uninterrupted voicing: each /s/ token was labeled as *voiced* when no interruption of pitch periods was observed throughout the dip in low-frequency intensity, and as *voiceless* otherwise. Unvoiced /s/ tokens can therefore present variable degrees of voicing throughout their realization (in parts of the signal other than the low-frequency intensity dip), resulting from variable degrees of weakening at the glottal level. Voiced tokens, on the other hand, can all be regarded as instances of substantial weakening.

*High-band (4-8 KHz) intensity difference:* Clearly articulated /s/ consonants typically exhibit significantly more acoustic energy than vowels in the high frequencies, resulting mainly from the presence of aperiodic energy. In order to estimate the strength of the frication noise characteristic of /s/, we defined two 30 ms windows, one centered at the midpoint of the low-band intensity dip, and the other one left-aligned immediately after it (corresponding roughly to the initial part of the following vowel). Then we calculated the difference in intensity in a 4-8 KHz band between the intensity peak in the low-band intensity dip window and the lowest intensity value in the vowel window. During the initial inspection of the data, this measure appeared to be a better correlate of the intensity of high-frequency frication than more widely used measures such as the spectral center of gravity. It is referred to as high-band intensity difference from now on. Realizations of /s/ with strong frication noise should exhibit high values of this measure, as opposed to consonants with weak frication.

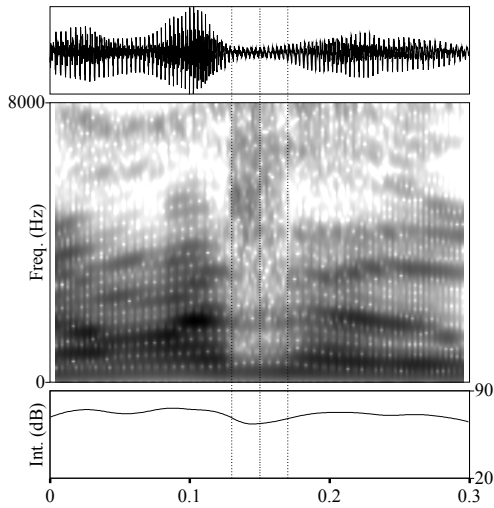
## Predictors

For each intervocalic /s/ token, we coded a number of variables as potential conditioning factors of the measures of /s/-weakening that we performed (low-band intensity dip duration, voicing, and high-band intensity difference). These variables were chosen on the basis of common knowledge about the conditioning factors of coarticulatory and reduction phenomena, and of the ease with which they could be coded automatically. These included speech rate, the immediate segmental context in which /s/ occurred (preceding and following vowel), the position of /s/ in the word (initial versus medial versus final), the number of syllables in the word, and whether /s/ was the onset of a stressed syllable. We also included speaker gender in our predictors, since this variable could potentially affect voicing and spectral measures due to physiological differences between females and males.

In order to estimate speech rate in the utterances in which target /s/ consonants were produced, we used the following procedure. First, we calculated the mean and standard deviation of the log duration of every phone category in the forced



(a)



(b)

Figure 7.4: Waveforms, spectrograms, and intensity contours of examples of clearly articulated (a) and weakened (b) /s/ consonants. The solid curves in the intensity panels correspond to the intensity contours of signals low-passed filtered at 1.5 KHz. The onsets, offsets, and midpoints of the low-band (0-1.5 KHz) intensity dip are represented by vertical dotted lines (see text for more details on how these time points were determined).

	/a/	/e/	/i/	/o/	/u/
$S_{i-1}$	272 (21.6%)	468 (37.2%)	102 (8.1%)	398 (31.6%)	17 (1.3%)
$S_{i+1}$	334 (26.5%)	403 (32.1%)	247 (19.6%)	198 (15.7%)	75 (5.9%)

Table 7.1: Number of tokens for each segment type preceding and following the target /s/ ( $S_{i-1}$  and  $S_{i+1}$ ).

alignment of the NCCSp. Then we computed the log duration of the three phones preceding and following, but not including, the target /s/ consonant, and then calculated how many standard deviations each phone log duration was from the mean log duration of its phone category. Our estimate of speech rate was defined as the average value of these deviations. Negative values of this estimate were characteristic of fast speech rates (i.e. as a consequence of phone durations shorter than average in the vicinity of the target /s/), while positive values were characteristic of slow speech rates.

We investigated the role of the segmental context by coding the identity of the surrounding vowels. These segments are referred to as  $S_{i-1}$ ,  $S_{i+1}$ , where  $S_i$  is the target /s/ consonant. Table 7.1 shows the number of tokens for each of this variable.

As for lexical stress, we coded if /s/ was the onset of a stressed syllable or not. In order to determine whether a syllable was stressed, we followed the guidelines provided in Hualde (2009). In the case of word-final /s/, we assume that resyllabification always takes place (e.g. *los otros* is pronounced as [lo.so.tros] ‘the others’). Word-final /s/ consonants therefore take the stress value of the first syllable of upcoming words. This was the case for 115 out of the 362 word-final /s/ consonants in the dataset. It should be noted that if our resyllabification assumption is not valid, interactions between position in word and stress should be observed in our statistical analyses. Following this procedure, 635 /s/ consonants were coded as stressed, and 622 as unstressed.

### Statistical analysis

In all analyses, we use mixed-effects linear regression with contrast coding as implemented in the *lmer* function in R (Bates and Sarkar, 2006). For the occurrence of voicing throughout /s/, a binary dependent variable (voiceless versus voiced), we use logistic mixed-effects linear regression. Speaker and word type are modeled as random factors. The other predictors (e.g. speech rate, position in word, number of syllables) are treated as fixed predictors. Only statistically significant predictors are retained in the reported models. Similarly, only statistically significant interactions between predictors are retained. Since determining the number of degrees of free-

dom in mixed-effects models is a controversial issue,  $p$  values are obtained through Monte Carlo Markov Chain simulations (Baayen, 2008).

## 7.2.2 Results

### Voicing

Over a third (34%,  $n = 434$ ) of the randomly extracted vowel-/s/-vowel sequences contained uninterrupted voicing. Voicing of intervocalic /s/ in casual Spanish therefore appears to be a common phenomenon. In order to investigate the different sources of variability behind the voicing of intervocalic /s/, we fitted a regression model considering the variables in Section 7.2.1 (i.e. speech rate, preceding and following vowel, position in the word, number of syllables in the word, stress) as fixed predictors, and speaker and word type as random factors. Low-band intensity dip duration was added to the other predictors in the model because we wanted to investigate if weakening at the glottal level co-occurred with the temporal reduction of the consonant.

In order to facilitate the interpretation of regression coefficients, the continuous predictors speech rate and low-band intensity dip duration were standardized into  $z$  scores. Since these continuous predictors were moderately correlated with each other ( $r = .33$ ), we orthogonalized these two variables by replacing low-band intensity dip duration with the residuals of a linear model in which low-band intensity dip duration was predicted by speech rate. This variable captures the information in low-band intensity dip duration that cannot be attributed to its linear relationship with speech rate, and allows us to assess the role of low-band intensity dip duration independently from that of speech rate. Since low-band intensity dip duration appeared to be slightly associated with position in the word (word-final /s/ consonants tend to be shorter, see subsection on low-band intensity dip duration below), we decided to run, in addition to our regression models, an ANOVA test in which the role of word position was assessed sequentially after controlling for the effects of duration.

Speaker exhibited a standard deviation of 0.88, indicating that the average difference in the occurrence of voicing among speakers was approximately 22%. This inter-speaker variability is illustrated in the kernel density plot in Figure 7.3, which shows the distribution of percentages of voicing for each speaker. This figure shows that most speakers in our data had a voicing rate below 20%, including a few who rarely voiced their /s/ consonants, and one who never did. High rates of voicing were not uncommon in the data either, with 8 of the 52 speakers exhibiting voicing rates over 50%, and reaching 80% for two speakers. The random factor word was not significant.

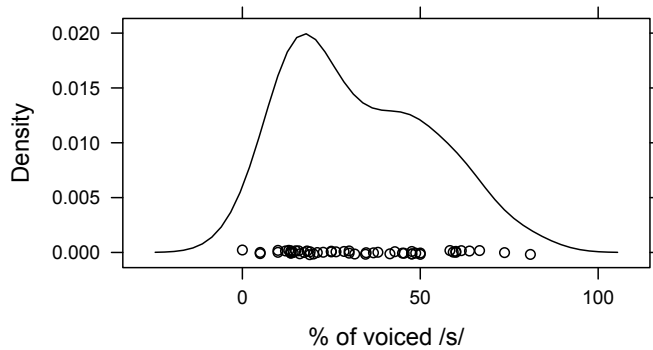


Figure 7.5: Kernel density plot of percentages of voicing of intervocalic /s/ per speaker. Each speaker is represented by a circle. Larger areas under the density curve correspond to greater concentrations of circles along the x-axis.

As for the fixed effects, we found that only speech rate, low-band intensity dip duration and word position were statistically significant predictors of the likelihood of voicing. These effects are illustrated in Figure 7.5. Voicing tended to occur more often the higher the speech rate ( $\beta = -0.43, z = -5.85, p < 0.0001$ ), and the shorter the low-band intensity dip ( $\beta = -0.71, z = -8.3, p < 0.0001$ ). With respect to word position, /s/ tended to be voiced more often in word-final position than in word-initial and medial positions (word-initial:  $\beta = -0.46, z = -2.87, p < 0.005$ ; word-medial:  $\beta = -0.52, z = -3.01, p < 0.005$ ). An ANOVA test confirmed that word-position conditioned the occurrence of voicing after controlling for the effects of duration ( $\chi^2 = 11.06, p < 0.005$ ).

### High-band intensity difference

As explained above, we used the maximal difference in intensity between the midpoint of /s/ and the beginning of the following vowel in a 4-8 KHz band as an indirect measure of the strength of frication in /s/. We first examined the distribution of this high-band intensity difference measure. It exhibited a mean of 17.88 dB, a standard deviation of 8.21 dB and a range extending from -3.75 to 44.2 dB. Around 17% of the tokens had values below 10 dB, and 6% below 5 dB. As an illustration of the incidence of weak frication in our data, consider that the examples of /s/ realization with little or no weakening in Figure 7.1 have high-band intensity differences of 19.34 and 18.6 dB, which is slightly above average in our data, while the weakened

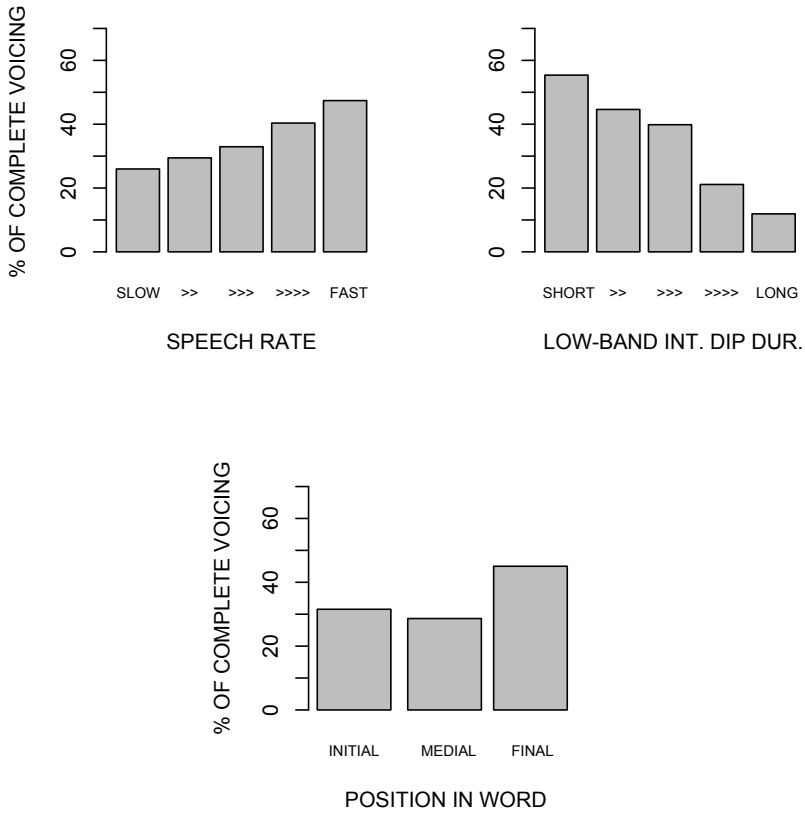


Figure 7.6: Percentage of voicing in intervocalic /s/ consonants as a function of speech rate, low-band intensity dip duration (discretized in five intervals of equal size), and position in the word.

examples in Figures 7.2 and 7.3 had much lower values ranging from -0.1 dB to 4.75 dB. These numbers indicate that weak frication such as that seen in Figures 7.2 and 7.3 is not as frequent as the complete voicing of /s/, which affected over a third our tokens, although it does occur in a significant number of cases.

We then fitted a regression model by considering all the predictors presented in Section 7.2.1. Since voicing greatly affects the strength of frication, as it diminishes airflow at the consonantal constriction, we included voicing as a predictor in the regression model. Moreover, we also included low-band dip duration in the model in order to assess if the weakening of frication was associated to the temporal reduction of the consonant.

Speaker was a statistically significant random factor ( $\sigma = 3.26$ ). As for the fixed effects, we found that high-band intensity difference was affected by speech rate, voicing, and low-band dip duration: voiced tokens of /s/ had lower high-band intensity difference values the higher the speech rate ( $\beta = -0.65, t = -3.53, p < 0.0005$ ), and an interaction between voicing and low-band intensity dip duration was observed. Frication was stronger the longer the duration of the low-band intensity dip, but only for voiceless realizations of /s/ ( $\beta = 1.57, t = 3.43, p < 0.001$ ). No other statistical effects were observed.

### Low-band intensity dip duration

Finally, we examined the duration of the low-band intensity dip in the vowel-/s/-vowel sequences. Although this measure cannot be used as the segmental duration of /s/, it can be used as an estimate of its temporal reduction (see Section 7.2.1). The dependent variable exhibited a range extending from 16 to 134 ms, with a mean of 50 ms and a standard deviation of 21 ms. As an illustration of the amount of temporal reduction of /s/ in our data, consider that around 16% of our tokens had values below 30 ms, which was roughly the durational value of the weakened tokens in Figures 7.2 and 7.3. As for the weakening of frication, therefore, temporal reduction of the kind illustrated by Figures 7.2 and 7.3 appeared to affect a significant number of cases, although it is less common than the complete voicing of /s/.

Regression modeling revealed that speaker and word type were statistically significant random factors, although their effects were rather small ( $\sigma = 6$  ms for speaker, and  $\sigma = 5$  ms for word). As for the fixed effects, we found that only speech rate, position in the word, and stress were statistically significant predictors, with longer low-band intensity dips at slower speech rates ( $\beta = 5.44, t = 10.15, p < 0.0001$ ), in word-initial and medial positions compared to final position (word-initial:  $\beta = 4.82, t = 2.75, p < 0.01$ ; word-medial:  $\beta = 4.74, t = 2.83, p < 0.005$ ), and in stressed syllables ( $\beta = 3.55, t = 2.52, p < 0.05$ ).

## Discussion

The analyses presented in Study 1 show that the voicing of intervocalic /s/ in Madrid Spanish is a frequent phenomenon. Roughly a third of the intervocalic /s/ tokens in our data showed no interruption of voicing. Signs of extreme weakening in terms of loss of frication and temporal compression were also observed, although these phenomena appeared to be somewhat less common than uninterrupted voicing. Regression modeling revealed that voicing was more frequent at faster speech rates, in word final position, and in temporally reduced consonants. High-band intensity difference, an estimate of the strength of frication, was lower the higher the speech rate, and, within the group of voiceless realizations of /s/, the shorter the consonant. Low-band intensity dip duration, an acoustic estimate of the temporal reduction of intervocalic /s/, was affected by speech rate and position in the word in the same way as voicing, and also showed an effect of stress, with shorter durations in unstressed positions.

Our observations therefore indicate that the weakening of intervocalic /s/ in Spanish is sensitive to several factors known to condition coarticulatory and reduction phenomena, in particular to temporal variables. In our view, this is consistent with the possibility that /s/-weakening arises under the coarticulatory pressure of adjacent vowel segments. Our results also indicate that the examined acoustic measures (the presence of uninterrupted voicing, high-band (4-8 KHz) intensity difference and low-band (0-1 KHz) intensity dip duration) were successful in capturing acoustic variability associated to /s/-weakening.

### 7.3 Study 2: Morphological factors

As explained in the Introduction, Spanish word-final /s/ can function as a plural marker or a verbal suffix. This raises questions regarding possible morphological effects on the realization of /s/ in word-final position, which, interestingly, tends to be more weakened than in other positions in the word (see Study 1 above). In Study 2, we address two hypotheses regarding the role of morphology on articulation. First, segments that carry morphosyntactic information by themselves are hypothesized to undergo less reduction than segments that do not. For this reason, /s/ suffixes are predicted to be less prone to weakening than /s/ segments which are part of another morpheme (e.g. *martes* ‘Tuesday’, *es* ‘is’). Second, within the group of suffixes, we hypothesize that weakening should occur more in /s/ suffixes when they are grammatically or semantically redundant than when they carry new information. For instance, the /s/ plural marker of the word *cosas* ‘things’ is predicted to be more weakened when it is preceded by a word like *las* ‘the’, an article also containing a



plural marker, than when it is preceded by a word with no grammatical information regarding number (e.g. *quiero* ‘I want’).

In order to investigate these issues with sufficient statistical power, we created a dataset of word-final /s/ consonants consisting of all word-final /s/ consonants in the dataset of Study 1 plus 572 word-final /s/ consonants additionally sampled from the NCCSp. The additional tokens were measured and annotated in the same way as the tokens from Study 1. This dataset, exclusively composed of word-final /s/ consonants, contained 930 tokens, of which 559 corresponded to /s/ suffixes (222 plural markers in nouns, 127 plural markers in articles, determiners and pronouns, 59 plural markers in adjectives, and 151 verbal suffixes). Of these suffix /s/ tokens, 234 were redundant based on the morphosyntactic or semantic properties of the preceding word (e.g. *las casa*<sub>s</sub> ‘the houses’, *cuatro año*<sub>s</sub> ‘four years’). Most of these redundant suffixes were found in nouns ( $n = 154$ ) and adjectives ( $n = 43$ ).

Regression models predicting /s/ voicing, high-band intensity difference, and low-band intensity dip duration were fitted with a morphological predictor with three levels (non-suffix /s/, redundant suffix, and non-redundant suffix) plus the predictors used in Study 1. No differences were observed between any levels of the morphological predictor in the models predicting high-band intensity difference and low-band intensity dip duration. In contrast, the voicing model indicated that non-suffix /s/ tends to be voiced less often than redundant suffixes (48% versus 56%;  $\beta = 0.39, t = 2.01, p < 0.05$ ), and with a similar likelihood as non-redundant suffixes (these were voiced in 50% of the cases;  $\beta = 0.01, t = 0.06, p = 0.94$ ). A post-hoc comparison between redundant and non-redundant suffixes yielded a marginally statistically significant difference ( $p = 0.06$ ). Although the size of the observed differences were relatively small, these results provide some evidence for the hypothesis that grammatical redundancy can lead to increased reduction. On the other hand, the hypothesis that /s/ suffixes are more resistant to weakening than non-suffix /s/ consonants in general is not supported by our data.

We then investigated to what extent the word-position effects on voicing found in Study 1 (voicing was more frequent in word-final position) were attributable to morphological redundancy, since many of the word-final consonants in the dataset used in this study were morphologically redundant /s/ suffixes. In order to do this, we reran the voicing model of Study 1 on a subset of the randomly sampled dataset of Study 1 from which morphologically redundant /s/ tokens had been excluded. This model showed that, when redundant /s/ consonants are not taken into account, the coefficients of the factor word position remain largely unchanged (word-initial:  $\beta = -0.39, z = -2.04, p < 0.05$ ; word-medial:  $\beta = -0.51, z = -2.74, p < 0.01$ ). The coefficients of the other predictors in the model did not change significantly either. Therefore, the frequent voicing of /s/ in word-final position cannot be solely

attributed to the fact that it often carries redundant morphosyntactic information, since significant differences between final and non-final /s/ was also observed in a dataset from which morphosyntactically redundant /s/ suffixes had been removed.

## 7.4 Study 3: Probabilistic and lexical factors

We now investigate the potential role of lexical and probabilistic variables on /s/-weakening, including grammatical class (content versus function words), lexical frequency and contextual predictability, and word-specific effects. In particular, we assess the extent to which reported effects of lexical and probabilistic factors on global measures of reduction such as word or syllable duration computed across large speech corpora (e.g. Bell et al., 2003, 2009; Aylett and Turk, 2004, 2006) apply in the case of a specific reduction phenomenon such as intervocalic /s/-weakening in Madrid Spanish.

In order to address this question with sufficient statistical power, we added tokens of highly frequent words to the dataset used in Study 1. These words were those which had more than 15 tokens in the randomly-sampled dataset used in Study 1. The extended dataset contained 1615 /s/ consonants, of which more than half ( $n = 928$ ) occurred in one of the 12 following high-frequency words: *entonces* ‘then’ ( $n = 51$ ), *es* ‘is’ ( $n = 137$ ), *eso* ‘that’ (pronoun) ( $n = 96$ ), *la* (feminine article) ( $n = 54$ ), *lo* (masculine article) ( $n = 65$ ), *nosotros* ‘we’ ( $n = 52$ ), *sea* ‘to be’ (3rd person singular of the present subjunctive) ( $n = 57$ ), *pues* ‘so’ ( $n = 52$ ), *sabes* ‘you know’ ( $n = 103$ ), *se* (pronominal particle) ( $n = 132$ ), *si* ‘if’ ( $n = 66$ ) and *sé* ‘I know’ ( $n = 50$ ).

Lexical frequency was computed based on the counts of orthographic forms in the transcriptions of the NCCSp. We assessed the validity of our frequency counts by computing correlations with frequency counts from two other corpora of spoken Spanish, the Spanish Call Home Corpus and the Corpus Oral del Español (Garrett et al., 1997). In both cases high correlations were observed ( $r = 0.94$  and  $r = 0.89$ ). As a measure of contextual predictability, we computed two word bigram frequencies from the transcriptions of the NCCSp:  $C(w_{i-1}w_i)$  (the bigram frequency of the word preceding the target word and of the target word) and  $C(w_iw_{i+1})$  (the bigram frequency of the target word and of the word following the target word). All frequency measures were log-transformed, since they exhibited highly skewed distributions. Finally, we considered nouns, adjectives, adverbs and verbs as content words, and all other word types as function words.

The computed probabilistic variables and grammatical class were correlated with each other in several cases. Grammatical class and lexical frequency were moderately associated ( $r = .33$ ; function words tended to be of higher lexical frequency), and lexical frequency and the two word bigram frequency measures also displayed

moderate correlations ( $r = .47$  and  $r = .55$ ). Since multicollinearity among predictors pose serious obstacles to regression modeling, we decided to run separate analyses for each variable first, and then proceed by fitting pairs of orthogonalized variables using the residuals method (see Section 7.2.2 above).

Contrary to our expectations, intervocalic /s/ consonants did not appear to be significantly more prone to voicing in function words than in content words (39.9% for function words versus 37.3% for content words). This was confirmed by a regression model in which the conditioning factors of voicing identified in Section 7.2.2 were controlled for ( $p > .4$ ). Similarly, no effects of word class were found on high-band intensity difference ( $p > .35$ ) nor on low-band intensity dip duration ( $p > .2$ ). These results suggest that /s/-weakening in Madrid Spanish is not conditioned by the function versus content word distinction.

We then investigated the roles of lexical frequency and contextual predictability. Again, we started from models including the conditioning factors identified in Study 1, to which we added the probabilistic predictors separately, and in orthogonalized pairs. No effect of any of the probabilistic variables was found in any of the analyses predicting voicing ( $p > .25$  in all cases), nor high-band intensity difference ( $p > .3$  in all cases). For low-band intensity dip duration, on the other hand, we found a statistically significant effect of word bigram frequency  $C(w_i w_{i+1})$ , that is of the bigram frequency of the target word and its following word. Contrary to our expectations, intervocalic /s/ consonants tended to have longer low-band intensity dip durations the higher the  $C(w_i w_{i+1})$  value ( $\beta = 0.6, t = 2.2, p < .05$ ). It should be noted, however, that the size of this effect was small. As an illustration, consider that the range of  $\log C(w_i w_{i+1})$  in our data extended from 0 to 6.3, and that the change in low-band intensity dip duration by  $\log C(w_i w_{i+1})$  unit was only of 0.6 ms. In our data, therefore, the maximum predicted difference in low-band intensity dip duration attributable to  $C(w_i w_{i+1})$  is only 3.78 ms. Further research would be needed to establish whether this is a robust trend.

Finally, we investigated word-specific effects. Our goal was to test if specific high-frequency words exhibited more /s/-weakening than predicted by the regression models developed in Study 1. We defined a factor *word type* with 13 levels, one for each of the 12 highly frequent words, plus a default level corresponding to all other word types in the corpus. Given that the factors word position and word type were not independent, we split the dataset into subsets with word-final and non-word-final /s/, and ran separate analyses on each of them. The results are summarized in Table 7.2.

For voicing, we found statistically significant effects for *nosotros* (word-medial:  $\beta = 0.92, z = 2.68, p < .001$ ) and *sabes* (word-initial /s/:  $\beta = 0.73, z = 2.2, p < .05$ ). The word-medial /s/ in *nosotros* and the initial /s/ in *sabes* tended to be voiced more

<i>Word-initial and medial position</i>						
Word	N	% voiced	Voicing	Int. Dif.	Duration	
<i>eso</i>	96	21%				
<i>nosotros</i>	52	57%	< .001	< .0001	< .05	
<i>sea</i>	57	29%		< .0001	< .0001	
<i>sabes</i>	52	55%	< .05	< .05	< .05	
<i>se</i>	132	35%		< .05	< .05	
<i>si</i>	66	30%				
<i>sé</i>	50	32%				

<i>Word-final position</i>						
Word	N	% voiced	Voicing	Int. Dif.	Duration	
<i>entonces</i>	51	49%				
<i>es</i>	137	43%				
<i>las</i>	54	51%				
<i>los</i>	65	33%				
<i>pues</i>	52	59%				
<i>sabes</i>	51	49%				

Table 7.2: High-frequency words: number of tokens analyzed (N), percentage of voiced /s/ (% voiced), and statistically significant  $p$  values obtained for the difference between each high-frequency word and low-frequency words in the models predicting voicing, high-band intensity difference and low-band intensity dip duration (see Section 7.2.1 for details).

often than /s/ consonants in other words (the voicing rate was 58% for medial /s/ in *nosotros* and 56% for the initial /s/ in *sabes*, versus 31% and 26% for medial and initial /s/ in low-frequency words).

For high-band intensity difference, effects were found again for the word-medial /s/ in *nosotros* ( $\beta = -5.25, t = -5.05, p < 0.0001$ ), *sea* ( $\beta = -3.95, t = -4.08, p < 0.0001$ ) and word-initial /s/ in *sabes* ( $\beta = -2.23, t = -2.12, p < .05$ ). High-band intensity difference values tended to be lower for these words than for the group of low-frequency words.

As for low-band intensity dip duration, effects were observed for the words *nosotros* (word-medial:  $\beta = -6.57, t = -2.33, p < 0.05$ ), *sea* ( $\beta = -14.82, t = -5.61, p < 0.0001$ ), *se* ( $\beta = -5.23, t = 2.27, p < .05$ ), and for the word-initial /s/ in *sabes* ( $\beta = -6.51, t = -2.25, p < .05$ ). Low-band intensity dip duration tended to be shorter in these words than in low-frequency words.

In conclusion, our data do not seem to support the hypothesis that /s/ consonants in highly frequent and predictable words and function words generally exhibit more weakening than when they occur in non-high-frequency words and content words.

However, word-specific trends towards more reduction could be found in some high-frequency words, in particular in the words *nosotros* and *sabes*.

## 7.5 General discussion

The studies presented above have investigated the weakening of intervocalic /s/ consonants in the Nijmegen Corpus of Casual Spanish (NCCSp), a newly created corpus of informal conversations among speakers of Madrid Spanish. Since acoustic segmentation following traditional procedures proved inappropriate for our data, we devised three acoustic parameters related to the articulation of /s/ that could be measured semi-automatically in a relatively large corpus of spontaneous speech such as the NCCSp. The analysis of such parameters (high-band intensity difference, low-band intensity dip duration and the occurrence of uninterrupted voicing) have allowed us to assess the pervasiveness of /s/-weakening in this variety of Spanish, and the relevance of some of its conditioning factors.

In Study 1, we have observed that over a third of the /s/ consonants randomly sampled from the NCCSp exhibit uninterrupted voicing throughout their realization, and that the frequency of this phenomenon is highly variable among speakers. Judging from the distributions of high-band intensity difference and low-band intensity dip duration values, it appears that extreme reduction in terms of frication and the durational properties of intervocalic /s/ is less common than the occurrence of uninterrupted voicing.

In Study 1 we have also examined to what extent /s/-weakening is affected by several variables known to affect coarticulatory and reduction phenomena. All three acoustic parameters were consistently affected by speech rate. As expected, we observed more weakening at higher speech rates. In order to further investigate the extent to which the voicing and loss of frication in intervocalic /s/ is sensitive to temporal reduction, we also examined the relationship between these dependent variables and low-band intensity dip duration. In both cases, more weakening was observed in temporally reduced consonants. In our view, this is consistent with a view of /s/-weakening as a coarticulatory phenomenon.

Voicing and low-band intensity dip duration were also affected by position in the word, with more weakening in word-final position than in word-initial and medial positions. Although the sort of /s/-weakening in Madrid Spanish and /s/-aspiration in other dialects are not the same phenomena – in our data all /s/ tokens were clearly produced as sibilants – it is interesting to note that the aspiration of intervocalic /s/ in /s/-aspirating dialects occurs almost exclusively in word-final position. Word-final position therefore appears as a recurrent locus of different kinds of /s/ reduction across different Spanish dialects.

Stress had a small effect on low-band intensity dip duration. We found that /s/ had slightly longer durations in stressed syllables than in unstressed syllables. Interestingly, no effects of stress were observed for voicing and high-band intensity difference. Although this finding may seem counterintuitive, it parallels our previous observations for intervocalic /p/, /t/ and /k/ consonants in the NCCSp (Torreira and Ernestus, 2011, see Chapter 6). The realization of these consonants in spontaneous Spanish showed an effect of stress on duration and constriction degree, but, as in the case of intervocalic /s/, not on the presence of intrusive voicing from adjacent vowels. It therefore appears that stress does not play an important role in the occurrence of intervocalic voicing in voiceless obstruents in Spanish, although it does affect other aspects of their articulation.

Together with previous data on the weakening of Spanish intervocalic voiceless stops (e.g. Machuca-Ayuso, 1997; Lewis, 2001; Martínez Celdrán, 2009; Torreira and Ernestus, 2011; Hualde et al., to appear), our findings suggest that the kind of /s/-weakening studied here is probably an instance of a more general lenition process affecting all Spanish intervocalic voiceless obstruents. We now have evidence that several Spanish voiceless obstruents (/p/, /t/, /k/ and /s/) are often realized with uninterrupted voicing when occurring in intervocalic position. Moreover, we have evidence that voiceless stops are often produced without complete oral closures, and that /s/ consonants may exhibit very weak frication noise. In addition, preliminary data from an ongoing study indicates that other Spanish voiceless fricative types in the NCCSp are also subject to frequent intervocalic voicing (25.9% for /f/, 28.7% for /θ/ and 15.5% for /x/). It is possible that all of these cases of consonantal weakening share a common set of causes. For instance, it is possible that the glottal and supraglottal targets of voiceless obstruents in Spanish have all shifted towards more lenited positions in phonetic space. This may be due to more global changes in the basis-of-articulation of this language, for instance towards a more sonorant overall quality (e.g. a preponderance of wider constrictions and voicing). Another possibility is that it is a change in the coarticulatory resistance of voiceless obstruents to adjacent segments such as vowels, not a change in their phonetic targets, that is mainly responsible for the observed weakening of intervocalic obstruents. A combination of corpus, experimental and quantitative modeling work would be required to investigate these issues.

One of the goals of the present work was to examine the potential influence of morphological factors on /s/-weakening. We hypothesized that /s/ segments which are part of a larger morpheme (e.g. *martes* ‘Tuesday’) are more prone to weakening than /s/ segments that carry grammatical information by themselves as a plural or verbal suffix. We also hypothesized that /s/ suffixes that are redundant based on the linguistic context (e.g. *las casas* ‘the houses’) are more prone to weakening than other

word-final /s/ segments. We found some support only for the second hypothesis, with 56% of voicing for redundant /s/ suffixes versus 48% for non-suffix /s/ and 50% for non-redundant suffixes. Our findings are therefore in line with Poplack (1980), Hundley (1987) and Erker (2010), who found that the reduction of /s/ in Puerto Rican, Peruvian and Dominican Spanish is not conditioned directly by morphological factors, but rather by the contextual information available in the utterance.

Regarding the roles of probabilistic and lexical factors, our data do not support the hypotheses that highly frequent and predictable words and function words systematically exhibit more reduction than low-frequency, unpredictable words and content words. In order to test for lexical frequency and word class effects, we used an extended dataset containing numerous tokens of highly frequent lexical items, many of them function words. In spite of the increase in statistical power allowed by this dataset, lexical frequency and word-class did not approach statistical significance in the analyses of voicing and high-band intensity difference. In the case of low-band dip intensity duration, our estimate of temporal reduction, we found a small effect of predictability based on the following word (word bigram frequency  $C(w_i w_{i+1})$ ), but this effect, as for two different acoustic correlates of /s/ in Caleño Spanish (File-Muriel and Brown, 2011), was in the opposite direction of that predicted by the probabilistic reduction hypothesis. Given the small size of this effect, further research is required to establish whether it is recurrent across different datasets.

One possible reason why our findings on the role of lexical class, frequency and contextual predictability contrast with much of the previous literature is that our dataset, while allowing for reasonable statistical power, did not contain a large number of word tokens and word types. In this sense, for instance, our dataset is not comparable to those used in influential studies like Bell et al. (2009) and Aylett and Turk (2004). A second possible explanation is that we focused on three acoustic parameters of a very specific reduction phenomenon, rather than on global measures of reduction such as word or syllable duration. This would suggest that the frequency and word class factors claimed to favor reduction in previous research may well exhibit general effects on global phonetic parameters measured across large-scale corpora, but not necessarily on all reduction phenomena when considered individually. Under the latter hypothesis, the reasons why some reduction phenomena are more sensitive to lexical and probabilistic factors than others would need to be investigated. In our case, intervocalic /s/-weakening appeared to be very frequent in conversational Madrid Spanish. It is possible that under such circumstances ceiling effects inhibit lexical class and probabilistic factors from affecting articulation.

A third possibility is that the general grammatical class and probabilistic effects reported by previous research are largely carried by a number of specific words and word sequences. This would be consistent with our findings of word-specific effects

for several lexical items, particularly *nosotros* ‘we’ and *sabes* ‘you know’. These words exhibited more /s/-weakening (more frequent voicing, lower high-band intensity differences, and shorter low-band intensity dip durations) than the bulk of low-frequency words in the data. Since *nosotros* is a very long high-frequency function word, it could be that frequency or word class do condition /s/-weakening, but only in interaction with word length. As for *sabes*, its status as a very frequent discourse marker may explain its special behavior. In our view, the fact that a number of word-specific effects could be identified, in contrast to general effects of lexical class and frequency, suggests that the influence of specific lexical properties (e.g. grammatical class, frequency, contextual predictability) on articulation is more complex than is usually acknowledged, and that further investigation on the role of these factors in articulation is still needed.

In conclusion, this article has provided a description of a reduction phenomenon little studied in the past despite its pervasiveness in conversational Spanish. Statistical analysis of three acoustic correlates of intervocalic /s/-weakening has provided insights into the role of several of its conditioning variables. The weakening of /s/ was affected by temporal factors and prosodic factors, suggesting that it may be caused by coarticulation with adjacent vowels. We did not find that /s/ suffixes are more resistant to weakening than word-final /s/ consonants in monomorphemic words. In fact, we found that /s/ suffixes tend to be more weakened than other word-final /s/ tokens when they occur in morphosyntactically redundant contexts. Regarding the roles of lexical class, frequency and predictability, often reported in the literature as important conditioning factors of speech reduction, we have observed a more complex pattern of variability than initially expected. Further phonetic research of the kind presented here is needed to better understand how all these factors come to condition speech reduction.



Despite the fact that speech reduction is prevalent in everyday verbal communication, it has received little attention from researchers in speech science. In particular, phonetic studies on speech reduction in languages other than English, German, and Dutch are still scarce. The aim of this dissertation was to partially remedy this situation by investigating speech reduction in French and Spanish. It has presented two new large corpora of French and Spanish spontaneous speech, the Nijmegen Corpus of Casual French (Chapter 2) and the Nijmegen Corpus of Casual Spanish (Chapter 7), as well as a series of research studies on different reduction phenomena in these languages (Chapters 3-7). In this chapter, I discuss the contributions of this dissertation to the study of reduction phenomena in spontaneous speech, and suggest some directions for future research.

## **8.1 The Nijmegen Corpus of Casual French and the Nijmegen Corpus of Casual Spanish**

Spontaneous speech data collected in a naturalistic conversational setting provide the best medium for observing and quantifying the pervasiveness of speech reduction phenomena in everyday language. Given the lack of accessible corpora of spontaneous French and Spanish, I created two new corpora in order to study speech reduction in these languages: the Nijmegen Corpus of Casual French (NCCFr) and the Nijmegen Corpus of Casual Spanish (NCCSp). A novel recording protocol was devised to elicit highly casual and naturalistic verbal interactions. Chapter 2 showed that this protocol was successful, in that the NCCFr contains significantly more informal words, swear words, hesitations, and other indicators of casualness than the ESTER corpus, a corpus of journalistic French (Galliano et al., 2005). The NCCFr and the NCCSp contain orthographically transcribed high-quality audio and video data and will be publicly available soon from the European Language Resources Association. For these reasons, these corpora will be of great value for researchers interested in spontaneous and casual speech, and more particularly in speech reduction.

The research studies presented in this dissertation have confirmed that the NCCFr and NCCSp are rich sources of data for investigating reduction phenomena. For instance, in the NCCFr more than half of the occurrences of the word *c'était* /setɛ/ 'it was' contain few or no acoustic traces of a vowel between [s] and [t] (Chapter 4), and over a third of phrase-medial high vowels preceded by a voiceless consonant are totally devoiced (Chapter 5). In Spanish, voiceless stops are often produced with an incomplete closure (Chapter 6), and intervocalic voiceless obstruents are often realized with uninterrupted voicing (Chapter 6 and 7).

## 8.2 The nature of reduction phenomena

Chapters 4 and 5 investigated whether specific reduction phenomena are categorical (e.g. deletion of a segment during phonological encoding) or if they result from continuous articulatory reduction. These chapters studied the elision of the vowel /e/ in the French word *c'était* and the devoicing of phrase-medial high vowels in French. They have shown that quantitative analysis of acoustic data extracted from a corpus of spontaneous speech can provide valuable information on the gradient versus categorical nature of speech reduction, in particular, by observing the distribution of acoustic parameters (e.g. duration, voicing) and identifying their conditioning factors. For instance, in Chapter 4 I examined if the vowel /e/ in the French word *c'était* /setɛ/ is elided gradiently or categorically. I observed that the distribution of [s(e)] durations is bimodal, and that the modes of this bimodal distribution correspond to the groups of *c'était* words with elided and non-elided /e/ vowels. It was also found that the duration of [e] (when not elided) is not conditioned by the same factors as [e]-elision, in support of the idea that the temporal compression of [e] vowels and the elision of [e] are qualitatively different phenomena. In Chapter 5, which investigated vowel devoicing in French, I found that devoicing is more frequent in short or coarticulated vowels, and that complete and partial devoicing are conditioned by the same variables. This indicates that vowel devoicing in French is a gradient phenomenon. Based on the results of such quantitative methods, I concluded that the temporal properties of *c'était* pronunciations with an elided /e/ vowel differ from canonical pronunciations in a way that suggests categorical reduction, whereas French high-vowel devoicing presents the temporal characteristics expected to arise from gradient articulatory reduction.

At the same time, these chapters illustrate some of the limitations of corpus research for the study of speech production. In Chapter 4, for instance, I found that although /e/-elision in the word *c'était* involves some form of discreteness, the spectral characteristics of elision and non-elision [st] clusters suggested that the possibility of gradient articulatory reduction in [e] elision should not be totally discarded. This is-

sue could not be properly addressed due to the low number of canonical [st] clusters in our corpus, and to the lack of direct observations of articulation (as those provided by electromagnetic articulography or ultrasound techniques). In the spectral analysis of consonants /s/, /t/ and /k/ in the French words *si*, *tu* and *qui* presented in Chapter 5, I found some quantitative signs that these consonants might be more coarticulated when the following vowel is devoiced than when it is not. However, in the case of *tu* and *qui*, statistical significance was not achieved, probably due to an insufficient number of tokens of these words in my corpus.

Since eliciting reduced speech in the lab poses a major challenge to researchers, my findings show that corpora of spontaneous speech can be valuable tools for conducting an initial investigation of undocumented and understudied reduction phenomena. However, as noted above, the analysis of uncontrolled acoustic data has intrinsic limitations that can only be addressed with experimental data. For these reasons, corpus studies will often need to be complemented by more controlled experimentation in order to address hypotheses about the production mechanisms underlying reduction phenomena.

### 8.3 Probabilistic, lexical, and morphological factors in speech reduction

Chapters 3, 5, and 7 have addressed the question of whether speech reduction in Romance languages such as French and Spanish is conditioned by probabilistic and lexical factors, as has been shown in the past for Germanic languages like English and Dutch. Chapter 3 showed that French [t] duration is sensitive to word bigram frequency. This chapter therefore found support for the existence of probabilistic effects on articulation in a Romance language. Interestingly, the nature of the observed probabilistic effects depended on the sources from which lexical frequencies had been estimated (from a corpus of spontaneous speech or written texts), and on the type of speech being analyzed (spontaneous vs. journalistic speech). For these reasons, a conclusive interpretation of the results in terms of speech production mechanisms was not provided. In Chapter 5, I investigated if phrase-medial vowel devoicing in French was more common in three high-frequency function words (*tu* ‘you’, *qui* ‘who’ and *si* ‘if’) than in content words. No evidence for such a difference between function words and content words was found. In Chapter 7, I investigated if intervocalic word-final /s/ consonants in Madrid Spanish were more weakened in highly frequent and predictable words and function words compared with low-frequency, unpredictable words and content words. No evidence was found for a general effect of grammatical class or lexical frequency, although word-specific effects were observed for a small number of high-frequency words, in particular the

words *sabes* ‘you know’ and *nosotros* ‘we’. Regarding predictability, I found, contrary to the predictions of the probabilistic reduction hypothesis, that /s/ consonants were slightly longer the higher the bigram frequency of the target word and its following word.

I have therefore found only partial support for the kind of probabilistic and lexical effects found for Germanic languages in previous literature. One possible explanation for this is that, while previous research has tended to focus on coarse phonetic measures of reduction such as word duration and the number of realized segments in a word, I have examined finer-grained acoustic parameters in specific reduction phenomena (e.g. high-vowel devoicing in French, voicing of intervocalic /s/ in Spanish). Another possibility is that the Germanic languages studied in previous research and the two Romance languages examined in this dissertation have properties that may render them subject to reduction in different ways. For instance, Romance languages such as French and Spanish are said to have a syllable-timed rhythm (syllables within an utterance tend to have similar durations), while Germanic languages are said to be stress-timed (the intervals between stressed syllables tend to have similar durations). Since Romance languages exhibit less variability in syllable duration than Germanic languages, in which unstressed syllables are subject to extreme temporal reduction, it could be argued that probabilistic effects should be less easily observed in the former language family than in the latter. This might explain the discrepancies between my findings on French and Spanish and previous studies on English and Dutch. Further research is needed to investigate this question.

Chapter 7 studied the role of morphological factors in speech reduction in Spanish. In this chapter, I investigated if /s/ segments which are part of a larger morpheme (e.g. *jueves* ‘Tuesday’, *es* ‘is’) are more prone to reduction than /s/ segments that carry grammatical information by themselves as a plural or verbal suffix. I also investigated if /s/ suffixes that are redundant based on the linguistic context (e.g. *las mesas* ‘the tables’, where the final /s/ of *mesas* can be predicted on the basis of the determiner *las*) are more reduced than other word-final /s/ segments. In line with the findings of studies on /s/-aspiration and deletion in Latin American Spanish dialects, I only found support for the second hypothesis. This indicates that functional factors can affect the articulation of speech in its phonetic detail. On the other hand, my findings disagree with claims about the deletion of word-final /t/ and /d/ in English, said to delete more often in monomorphemic words (e.g. *mist* /mɪst/) than when they occur as a suffix (e.g. *missed* /mɪst/).

## 8.4 Language differences in speech reduction

Chapter 6 explicitly addressed the question of language specificity in speech reduction patterns. It compared the realization of sequences of voiceless stops and vowels in French and Spanish (as in Spanish *la tapa* ‘the lid’). Significant differences were observed between the two languages. Spanish voiceless stops tend to have shorter stop closures, exhibit incomplete closures more often, and display more voicing than French voiceless stops. Regarding vowels, more cases of complete devoicing and greater degrees of partial devoicing were found in French than in Spanish. Moreover, all the vowel types examined (/aeiou/) appeared to be less open in French than in Spanish. These findings indicate that the extent of reduction that specific segment types can undergo in conversational speech can vary substantially across languages.

In the discussion section of Chapter 6, I pointed to differences in coarticulatory strategies and in basis-of-articulation as possible causes of the differences observed between French and Spanish. Although the idea that languages may differ in these terms is by no means an original proposal, the importance of my findings lies in the size of the cross-linguistic differences observed in my corpus data. Small cross-linguistic phonetic differences have been repeatedly observed in laboratory studies, but, to my knowledge, no study has reported differences of the size of those presented in Chapter 6, with Spanish voiceless stops often being realized as voiced approximants and French stops only rarely. My research therefore indicates that the study of phonetic diversity can benefit greatly from corpus studies of spontaneous speech.

## 8.5 Concluding remarks

This dissertation has presented two new corpora of French and Spanish spontaneous speech, the Nijmegen Corpus of Casual French and the Nijmegen Corpus of Casual Spanish, and a series of studies on different reduction phenomena in these languages. Apart from providing two new speech corpora to the research community, it has made three main contributions to the study of speech reduction in spontaneous speech. First, it has shown that corpus-based research on acoustic data can be used to investigate the production mechanisms underlying specific reduction phenomena. Second, it has found only partial support for a series of probabilistic, lexical and morphological effects on speech reduction previously reported for Germanic languages. Finally, it has shown that the reduction of specific segments can vary greatly among languages, suggesting that the phonetic causes of small cross-linguistic differences observed in previous research, such as different patterns of coarticulation

and basis-of-articulation, may have far wider consequences in spontaneous speech than in laboratory speech.

## Summary

---

In everyday spontaneous conversation, speech sounds, syllables and words are often pronounced less clearly than in more formal speaking styles. In Spanish, for instance, the sounds /p/, /t/ and /k/ are often pronounced as voiced approximant sounds ([β], [ð] and [ɣ]). Studies of conversational speech have shown that reduction phenomena such as this one are much more common than is generally thought. However, despite its pervasiveness, speech reduction has received little attention from researchers in speech science until recently. In particular, phonetic studies on speech reduction in languages other than English, German, and Dutch are still scarce. The aim of this dissertation is to remedy this situation by investigating speech reduction in French and Spanish. It presents two new large corpora of French and Spanish spontaneous speech, the Nijmegen Corpus of Casual French and the Nijmegen Corpus of Casual Spanish, as well as a series of research studies on different reduction phenomena in these languages (Chapters 3-7). Chapters 3, 5, and 7 address the question of whether speech reduction in Romance languages such as French and Spanish is conditioned by probabilistic and lexical factors, as has been shown in the past for Germanic languages like English and Dutch. In Chapters 4 and 5, we investigated whether specific reduction phenomena are categorical (e.g. deletion of a segment during phonological encoding) or if they result from continuous articulatory reduction. Chapter 6 explicitly addresses the question of language specificity in speech reduction by comparing voiceless stops and vowels in casual French and Spanish.

### ***Chapter 2: The Nijmegen Corpus of Casual French***

Chapter 2 describes the preparation, recording and orthographic transcription of a new speech corpus, the Nijmegen Corpus of Casual French (NCCFr). The motivation for creating this corpus was to provide large amounts of high-quality recordings of casual speech suitable for phonetic analysis. The corpus contains a total of over 36 hours of recordings of 46 French speakers engaged in conversations with friends.

Casual speech was elicited during three different parts, which together provided around ninety minutes of speech from every pair of speakers. While Parts 1 and 2 did not require participants to perform any specific task, in Part 3 participants negotiated a common answer to general questions about society.

Comparisons with the ESTER corpus of journalistic speech show that the two corpora contain speech of considerably different registers. A number of indicators of casualness, including swear words, casual words, *verlan*, disfluencies and word repetitions, are more frequent in the NCCFr than in the ESTER corpus, while the use of double negation, an indicator of formal speech, is less frequent. In general, these estimates of casualness are constant through the three parts of the recording sessions and across speakers. Based on these facts, it is concluded that the NCCFr is a rich resource of highly casual speech, and that it can be effectively exploited by researchers in language science and technology.

### **Chapter 3: Probabilistic effects on French [t] duration**

A number of recent corpus studies on English and Dutch have shown that probabilistic factors such as the frequency of occurrence and contextual predictability of specific linguistic units (e.g. words, syllables, segments) may affect articulation (Pluymaekers et al., 2005; Bell et al., 2003, 2009; Aylett and Turk, 2004). For instance, Pluymaekers et al. (2005) found a shortening effect of the frequency of a word on the duration of several affixes and affix-internal segments drawn from a corpus of spontaneous Dutch. Similarly, Bell et al. (2009) reports that English content and function words tend to be shorter the higher their lexical frequency and conditional probability (i.e. probability of a word given its preceding or following words).

Chapter 3 investigates whether [t] consonants in French are affected by probabilistic factors, in spontaneous as well as in journalistic speech. In Study 1, which used casual speech materials from the NCCFr, an effect of the bigram frequency of the target word and its preceding word was identified when the probabilistic estimates were computed from a large corpus of journalistic texts. When probabilistic estimates computed from casual speech materials were used, a marginal effect of the bigram frequency of the target word and its following word was found. In Study 2, which used journalistic speech materials and probabilistic estimates, we found an effect of bigram frequency of the target word and its following word. In all cases, the duration of [t] closures tended to be shorter the higher the word bigram frequency. Given these findings, we conclude that French is also subject to the sort of probabilistic effects on articulation which have been identified for Germanic languages.

### **Chapter 4: Vowel elision in casual French: the case of vowel /e/ in the word *c'était***

Chapter 4 investigates the reduction of vowel /e/ in the French word *c'était* /setɛ/ 'it was'. A total of 450 tokens of the word *c'était* were extracted for analysis from the NCCFr. In order to compare [st] clusters in reduced *c'était* pronunciations with un-



derlying /st/ clusters, we further extracted tokens of words beginning with underlying /st/ clusters (e.g. *stage*, *stéréotype*).

We found that the reduction of vowel /e/ in the word *c'était* is highly frequent, as more than half of the occurrences of this word in the NCCFr contain few or no acoustic traces of a vowel between the consonants [s] and [t]. All my durational analyses clearly supported a categorical absence of vowel /e/ in a subset of *c'était* tokens. This interpretation was also supported by the finding that the occurrence of complete elision and [e] duration in non-elision tokens were conditioned by different factors. However, spectral measures were consistent with the possibility that a highly reduced /e/ vowel is still present in elision tokens. We discuss how these findings can be reconciled, and conclude that acoustic analysis of uncontrolled materials can provide valuable information about the mechanisms underlying reduction phenomena in casual speech.

### **Chapter 5: Phrase-medial vowel devoicing in spontaneous French**

Chapter 5 documents and investigates the occurrence of phrase-medial vowel devoicing in French (e.g. /ty pø/ [typø] 'you can'). A total of 550 words containing an unaccented high vowel between a voiceless consonant and a voiceless stop were extracted for analysis from the NCCFr.

It is shown that phrase-medial vowel devoicing occurs frequently in connected French. Over a third of the studied high vowels are completely devoiced, and an important number of tokens are partially devoiced in a considerable degree, both in function and content words. Devoicing occurs for syllables of all durations, but it tends to be more frequent in syllables with temporally reduced vowels. The spectral characteristics of the consonant preceding the target vowel provide some evidence of increased coarticulation between these segments in cases of complete devoicing, especially for syllables with fricative onsets. Finally, it was found that the occurrence of complete and partial devoicing are conditioned by similar variables (speech rate, manner of articulation and distance to the upcoming AP boundary). No differences in the likelihood and extent of devoicing were found between function and content words. In view of these findings, it is proposed that phrase-medial vowel devoicing in French is a gradient phenomenon arising mainly from the temporal compression of vocalic articulatory gestures.

### **Chapter 6: Realization of voiceless stops and vowels in conversational French and Spanish**

Chapter 6 compares the realization of intervocalic voiceless stops and vowels surrounded by voiceless stops in conversational Spanish and French. We study voiceless

stops surrounded by two vowels and vowels surrounded by two voiceless stops. The analyzed dataset contains a total of 1298 stops and 727 vowels for Spanish, and 856 stops and 497 vowels for French. The data were extracted from the Nijmegen Corpus of Casual French (NCCFr) and the Nijmegen Corpus of Casual Spanish (NCCSp). The NCCSp was collected following the same procedure as the NCCFr (see Chapter 2).

Significant differences were observed between French and Spanish. In intervocalic position, Spanish voiceless stops tend to have shorter stop closures, display incomplete closures more often, and exhibit more voicing than French voiceless stops. In Spanish, two major characteristics of voiceless stops according to standard phonetic descriptions (oral occlusion and absence of glottal vibration) are often compromised in intervocalic position: a quarter of the examined stops lacked a complete closure (i.e. were realized as approximants), and if stops with complete and incomplete closures are taken together, practically half of the examined Spanish stops are voiced throughout their entire closure. These reduction phenomena were also found in French voiceless stops, but to a much lesser extent. Regarding vowels, we found more cases of complete devoicing and greater degrees of partial devoicing in French than in Spanish, in spite of the fact that French vowels were generally longer than their Spanish counterparts. We also found that vowels tend to have significantly lower F1 values in French than in Spanish, suggesting that, at least in the context of stop consonants, French vowels are generally produced with a more constricted vocal tract than Spanish vowels.

These findings indicate that the extent of reduction that a segment type can undergo in conversational speech can vary significantly across languages, perhaps much more so than indicated by laboratory studies on language differences in coarticulatory strategies and “base-of-articulation”.

### ***Chapter 7: Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish***

Chapter 7 describes and investigates the conditioning factors of the weakening of intervocalic /s/ in Madrid Spanish, an almost undocumented phenomenon in which /s/ is realized with uninterrupted voicing and weak frication noise. Three phonetic parameters are examined, which were related to the occurrence of voicing through /s/, the intensity of frication noise and the temporal properties of the consonant.

Over a third of the /s/ consonants in a large random sample from the Nijmegen Corpus of Casual Spanish exhibited uninterrupted voicing. Weakening in terms of loss frication and temporal compression was also observed, but it was less frequent than the occurrence of uninterrupted voicing. All three acoustic parameters were consistently affected by speech rate. Voicing and low-band intensity dip duration

were also affected by position in the word, with more weakening in word-final position than in word-initial and medial positions. Finally, we observed a small effect of stress on the temporal properties of the consonant.

Subsequent analyses showed that /s/ suffixes in a redundant morphosyntactic context were more prone to reduction than other word-final /s/ segments. Several high-frequency words exhibited more reduction than low-frequency words, but no general effects of word frequency or grammatical class (content versus function word) were observed. This suggests that /s/-weakening in Madrid Spanish is subject to a more complex pattern of variability than initially expected from previous research on the role of lexical, probabilistic and morphological factors in speech reduction.

### ***Concluding remarks***

This dissertation presents two new corpora of French and Spanish spontaneous speech, the Nijmegen Corpus of Casual French and the Nijmegen Corpus of Casual Spanish, as well as a series of studies on different reduction phenomena in these languages. Apart from providing two new speech corpora to the research community, it makes three main contributions to the study of speech reduction in spontaneous speech. First, it shows that corpus-based research on acoustic data can contribute to the investigation of the production mechanisms underlying specific reduction phenomena. Second, it finds only partial support for a series of probabilistic, lexical and morphological effects on speech reduction previously reported for Germanic languages. Finally, it shows that the reduction of specific segments can vary greatly among languages, suggesting that the phonetic causes of small cross-linguistic differences observed in previous research, such as different patterns of coarticulation and basis-of-articulation, may have far wider consequences in spontaneous speech than in laboratory speech.



# Samenvatting

---

Spraakklanken, lettergrepen en woorden worden vaak minder duidelijk uitgesproken in alledaagse spontane conversaties dan in formelere spreekstijlen. In het Spaans bijvoorbeeld worden de stemloze klanken /p/, /t/ en /k/ vaak uitgesproken als de stemhebbende klanken [β], [ð] en [ɣ]. Onderzoek naar spontane spraak heeft aangetoond dat reducties, zoals in dit voorbeeld, vaker voorkomen dan over het algemeen verondersteld wordt. Ook al is spraakreductie een zeer frequent fenomeen, er is tot voor kort weinig aandacht voor geweest. Ook fonetische studies naar spraakreductie in andere talen dan Engels, Duits en Nederlands zijn nog steeds zeldzaam. Het doel van dit proefschrift is om daar verandering in te brengen door spraakreductie in het Frans en het Spaans te bestuderen. Er worden twee nieuwe grote corpora met spontaan Frans en Spaans beschreven, namelijk het Nijmegen Corpus Informeel Frans en het Nijmegen Corpus Informeel Spaans. Verder bevat dit proefschrift enkele onderzoeken naar verschillende reductiefenomenen in deze twee talen. De hoofdstukken 3, 5 en 7 onderzoeken de vraag of spraakreductie in Romaanse talen als het Frans en het Spaans wordt beïnvloed door de eigenschappen en voorspelbaarheid van het woord, zoals dat ook is aangetoond voor Germaanse talen als het Engels en het Nederlands. In de hoofdstukken 4 en 5 worden onderzoeken beschreven naar de vraag of specifieke reductiefenomenen categoriaal zijn (bijv. de volledige afwezigheid van een klank in het productieproces) of dat ze veroorzaakt worden door co-articulatie of door het verkleinen van de bewegingen van de mond en de tong. In hoofdstuk 6 richten we ons op de vraag of effecten in spraakreductie specifiek kunnen zijn voor één taal, door stemloze plofklanken (zoals /p/, /t/ en /k/) en klinkers in informeel Frans en Spaans te vergelijken.

## ***Hoofdstuk 2: Het Nijmegen Corpus Informeel Frans***

In Hoofdstuk 2 worden de voorbereidingen, de opnames en de orthografische transcripties van een nieuwe verzameling van opgenomen spraak beschreven. Dit corpus, het Nijmegen Corpus Informeel Frans (NCCFr), bevat meer dan 36 uur aan informele spraak geproduceerd door 46 sprekers. De opnames bestaan uit gesprekken met vrienden en zijn door hun hoge geluidskwaliteit zeer geschikt voor fonetische analyses.

In elke opnamesessie, van negentig minuten, werden conversaties tussen twee sprekers opgenomen. Elke sessie bestond uit drie delen en in elk deel werd op een andere manier informele spraak uitgelokt. Voor deel 1 en 2 werd geen specifieke taak gebruikt, terwijl in deel 3 de sprekers gevraagd werd om een antwoord te formuleren op een algemene vraag over de samenleving.

We hebben het NCCFr vergeleken met het ESTER-corpus, dat spraak van journalisten bevat. Uit deze analyse bleek dat de twee corpora duidelijk verschillende spreekstijlen bevatten. Aan de ene kant bevat het NCCFr meer indicatoren voor informaliteit, zoals vloekwoorden, informele woorden (bijv. *pote* 'maat' i.p.v. het formelere *ami* 'vriend'), *verlan*, haperingen en herhalingen. Aan de andere kant bevat het ESTER-corpus meer indicatoren voor formele spraak, zoals dubbele ontkenningen (bijv. *je ne veux pas* i.p.v. het informelere *je veux pas*). Binnen het NCCFr waren de indicatoren over het algemeen gelijk voor de drie onderdelen en voor de verschillende sprekers. We concluderen hieruit dat het NCCFr een rijke bron aan informele spraak is en dat het uitstekend door onderzoekers uit zowel de taalwetenschap als de spraaktechnologie gebruikt kan worden.

### **Hoofdstuk 3: Voorspelbaarheidseffecten op de duur van de Franse [t]**

Recente corpusstudies hebben aangetoond dat voorspelbaarheidsfactoren, zoals de frequentie waarmee een woord voorkomt en de contextuele voorspelbaarheid van specifieke taalkundige eenheden (bijv. woorden, lettergrepen of segmenten), invloed kunnen hebben op de articulatie van die eenheden in het Engels en Nederlands (Pluymaekers et al., 2005; Bell et al., 2003, 2009; Aylett and Turk, 2004). Pluymaekers et al. (2005) bijvoorbeeld hebben een studie gedaan op basis van het Corpus Gesproken Nederlands en vonden dat de frequentie waarmee een woord voorkomt invloed heeft op de duur van verschillende voor- en achterzetsels (zoals *ge*, *ont* en *lijk*) en de duur van de klanken binnen een voor/achtervoegsel. Vergelijkbare resultaten zijn gevonden door Bell et al. (2009). Zij onderzochten Engelse inhoudswoorden (zoals *bureau* en *weekend*) en functiewoorden (zoals *de*, *maar* en *wij*). Beide woordtypen bleken korter te zijn wanneer zij vaker voorkomen en bij een hogere contextuele voorspelbaarheid (d.w.z. een hogere voorspelbaarheid van een woord gegeven het volgende of voorgaande woord).

In het derde hoofdstuk worden twee studies beschreven naar de vraag of de Franse [t] in spontane en journalistieke spraak beïnvloed wordt door de voorspelbaarheid van het woord. Het eerste onderzoek is gebaseerd op de informele spraak uit het NCCFr. We vonden een effect van de bigramfrequentie van het targetwoord en het voorgaande woord, wanneer de voorspelbaarheidsfactoren berekend werden op basis van het journalistieke ESTER-corpus. Verder was er een marginaal effect van de bigramfrequentie van het targetwoord en het volgende woord, wanneer we de

voorspelbaarheidsfactoren op basis van informele spraak berekenden. In het tweede onderzoek bestudeerden we de Franse [t] in journalistieke spraak en baseerden we ook de voorspelbaarheidsfactoren daarop. Dit onderzoek toonde een effect van de bigramfrequentie van het targetwoord en het volgende woord. Voor alle bigrameffecten gold dat de duur van [t] korter was wanneer de bigramfrequentie hoger was. Op basis van deze resultaten concluderen we dat de articulatie in het Frans beïnvloed wordt door voorspelbaarheidsfactoren, zoals dat ook voor Germaanse talen gevonden is.

#### **Hoofdstuk 4: Klinkerdeletie in informeel Frans: de klinker /e/ in het woord *c'était***

In Hoofdstuk 4 onderzoeken we de reductie van de klinker /e/ in het Franse woord *c'était* /setɛ/ 'het was'. Voor de analyse hebben we 450 tokens van de Franse expressie *c'était* uit het NCCFr gehaald. Om de [st]-clusters van gereduceerde *c'était* tokens te kunnen vergelijken met ongereduceerde /st/-clusters in andere woorden, hebben we ook andere woorden die beginnen met /st/-clusters uit het corpus gehaald (bijv. *stage* of *stéréotype*).

Uit de analyses bleek dat de klinker /e/ in *c'était* vaak gereduceerd wordt. Akoestisch gezien was de klinker tussen [s] en [t] in meer dan de helft van de geanalyseerde tokens helemaal of bijna helemaal verdwenen. Als de /e/ afwezig was, had de spreker waarschijnlijk ook geen moeite gedaan om de /e/ uit te spreken. Dit blijkt onder andere uit de duur van de aanwezige /e/-klinkers en doordat de duur van [e] door andere factoren beïnvloed wordt dan de complete afwezigheid van /e/. Niettemin suggereerden de spectrale analyses dat een /e/ die afwezig lijkt te zijn, mogelijk toch nog steeds aanwezig is. Daarom bespreken we in dit hoofdstuk hoe deze bevindingen met elkaar in overeenstemming te brengen zijn en concluderen we uiteindelijk dat akoestische analyses van ongecontroleerd materiaal waardevolle informatie kunnen verschaffen over de mechanismen die reductiefenomenen in informele spraak veroorzaken.

#### **Hoofdstuk 5: Het stemloos uitspreken van klinkers in het midden van Franse zinnen**

Hoofdstuk 5 beschrijft een studie naar klinkers in het midden van Franse zinnen die stemloos uitgesproken zijn (dit is enigszins vergelijkbaar met het fluisteren van klinkers, bijv. ty pø/ typø 'jij kunt'). Voor dit onderzoek hebben we in totaal 550 woorden uit het NCCFr gehaald die een ongeaccentueerde hoge klinker bevatten tussen een stemloze medeklinker en een stemloze plofklank.

De resultaten toonden dat stemloos geworden zinsmediale klinkers vaak voorkomen in zowel Franse inhouds- als functiewoorden. Van alle geanalyseerde hoge klinkers was meer dan een derde compleet stemloos uitgesproken en een aanmerkelijk aantal tokens waren voor een groot deel stemloos uitgesproken. Het stemloos worden kwam voor in lettergrepen van alle lengtes, maar vaker in lettergrepen met een kort uitgesproken klinker. Uit analyses van de spectrale kenmerken van de medeklinker vóór de targetklinker bleek dat de co-articulatie tussen de medeklinker en klinker groter is wanneer de klinker compleet stemloos uitgesproken is. Dit effect was het grootst in lettergrepen die met een wrijfklank (bijv. [s] en [z]) beginnen. Verder werd het compleet stemloos worden van segmenten door dezelfde factoren beïnvloed als het gedeeltelijk stemloos worden van die segmenten, namelijk spreeknelheid, manier van articulatie en afstand tot de volgende frase. Bovendien vonden we geen verschillen tussen inhouds- en functiewoorden in hoe vaak ze stemloos worden en in de mate waarin dat gebeurt. In de conclusie stellen we voor dat het stemloos worden van Franse zinsmediale klinkers een gradueel proces is, dat voornamelijk veroorzaakt wordt door het verkleinen van de bewegingen van de mond en de tong.

### ***Hoofdstuk 6: De uitspraak van stemloze plofklanken en klinkers in Franse en Spaanse conversaties***

In Hoofdstuk 6 vergelijken we de uitspraak van klanken in Franse en Spaanse conversaties. De focus ligt op stemloze plofklanken uitgesproken tussen klinkers (d.w.z. intervocale plofklanken) en op klinkers die omgeven worden door stemloze plofklanken. De geanalyseerde dataset bevat in totaal 1298 Spaanse en 856 Franse plofklanken en 727 Spaanse en 497 Franse klinkers. De data zijn uit het Nijmegen Corpus Informeel Frans (NCCFr) en het Nijmegen Corpus Informeel Spaans (NCCSp) gehaald. Het laatste corpus was verzameld met dezelfde procedure als het eerste (zie hoofdstuk 2).

De analyse toonde aan dat Spanjaarden een kortere afsluiting in hun mond maken tijdens het uitspreken van stemloze plofklanken in intervocale posities dan Fransen. Verder zijn de Spaanse afsluitingen vaker incompleet en meer stemhebbend. Volgens de standaard fonetische beschrijvingen zijn er twee belangrijke kenmerken van Spaanse stemloze plofklanken, namelijk de afsluiting en de afwezigheid van stembandtrilling. In intervocale positie zijn deze kenmerken vaak niet duidelijk aanwezig: In onze analyse vonden we dat bij een kwart van de plofklanken een complete afsluiting ontbreekt (d.w.z. de plofklank wordt uitgesproken als een approximant). Verder is de afsluiting van praktisch de helft van alle Spaanse plofklanken volledig stemhebbend uitgesproken. Voor het Frans vonden we dezelfde reductiefenomenen, maar in veel mindere mate.



De klinkers lieten zien dat zij vaker compleet stemloos uitgesproken worden in het Frans dan in het Spaans. Bovendien is een groter gedeelte van de klinker stemloos uitgesproken in het Frans dan in het Spaans. Dit vonden we ondanks het feit dat Franse klinkers over het algemeen langer zijn dan hun Spaanse tegenhangers. Verder waren de F1-waarden van de Franse klinkers significant lager dan die van de Spaanse klinkers. Dit suggereert dat Franse klinkers in de context van plofklanken over het algemeen met een meer vernauwde keel-mondholte geproduceerd worden.

Deze bevindingen geven aan dat de mate waarin een bepaald segment in informele spraak gereduceerd kan worden, significant kan verschillen tussen talen. Misschien zelfs veel meer dan aangetoond kan worden in laboratoriumonderzoeken naar verschillen tussen talen in co-articulatie strategieën en in de basis-articulatie.

### ***Hoofdstuk 7: Verzwakking van intervocale /s/ in het Nijmegen Corpus Informeel Spaans***

Het zevende hoofdstuk bevat een onderzoek naar de factoren die invloed hebben op de verzwakking van de Spaanse intervocale /s/. Dit is een bijna ongedocumenteerd fenomeen, waarbij /s/ volledig uitgesproken wordt met stembandtrilling en een zwakke frictie heeft. We hebben drie fonetische parameters onderzocht, namelijk het voorkomen van ononderbroken stemhebbendheid door de /s/ heen, de intensiteit van de frictie en een maat die gerelateerd is aan de duur van /s/.

Voor de analyse hebben we een grote random steekproef van /s/-medeklinkers uit het Nijmegen Corpus Informeel Spaans gehaald. Meer dan een derde van deze /s/-tokens was geproduceerd met ononderbroken stemhebbendheid. Afwezigheid van frictie en verkorting kwam ook voor, maar lang niet zo vaak als ononderbroken stemhebbendheid. Alle drie de fonetische variabelen waren gevoelig voor spreek-snelheid. Ononderbroken stemhebbendheid en verkorting waren ook gevoelig voor de positie van /s/ in het woord, zodat /s/ in woord-finale positie meer verzwakt werd dan /s/ in woord-initiële en woord-mediale positie. Ten slotte vonden we ook een klein effect van klemtoon op de verkorting van /s/.

Latere analyses toonden aan dat /s/-achterevoegsels in een redundante morfosyntactische context vatbaarder zijn voor reductie dan andere /s/-tokens aan het eind van een woord. Verder waren verschillende woorden die vaak voorkomen meer gereduceerd dan woorden die minder vaak voorkomen, maar we vonden geen algemeen effect van woordfrequentie of van woordtype (d.w.z. inhouds- versus functiewoorden). Dit doet vermoeden dat de verzwakking van /s/ een complexer fenomeen is dan eerder verondersteld werd op basis van voorgaande studies naar de rol van lexicale, morfologische en voorspelbaarheidsfactoren in spraakreductie.

**Slotopmerkingen**

Dit proefschrift presenteert twee nieuwe corpora met spontaan Frans en Spaans, namelijk het Nijmegen Corpus Informeel Frans en het Nijmegen Corpus Informeel Spaans. Daarnaast biedt het een aantal studies naar verschillende reductiefenomenen in deze twee talen. Naast het verschaffen van twee nieuwe corpora, draagt dit proefschrift op drie andere manieren bij aan de studie van reductie in spontane spraak. Ten eerste toont het dat studies, die gebaseerd zijn op akoestische data uit corpora, kunnen bijdragen aan het onderzoek naar spraakproductiemechanismen, die ten grondslag liggen aan specifieke reductiefenomenen. Ten tweede verschaft het slechts gedeeltelijk ondersteuning voor lexicale, morfologische en voorspelbaarheids-effecten op spraakreductie, zoals die eerder gerapporteerd werden voor Germaanse talen. Ten slotte laat dit proefschrift zien dat de reductie van specifieke segmenten sterk kan verschillen tussen talen. Dit laatste suggereert dat de fonetische processen, die kleine verschillen tussen talen veroorzaken (zoals in eerder onderzoek geobserveerde verschillen in co-articulatie en basis-articulatie), veel grotere consequenties kunnen hebben in spontane spraak dan spraak opgenomen in laboratoria.

## Acknowledgments

---

This dissertation would not have been possible without the help of many people. First of all, I would like to thank my *promotores* Mirjam Ernestus and Lou Boves for making it possible for me to come to Nijmegen when they hired me in 2007. During all these years, they were always available to provide feedback, answer questions and help throughout all stages of the dissertation process. From them I have learned how to turn research ideas into published articles. They provided a constant motivation to aim higher than I thought I could reach. I believe all of this has significantly improved the quality of my work and made me a better researcher. My thanks too to the European Science Foundation for funding Mirjam Ernestus' EURYI project on speech reduction in European languages, of which this dissertation forms a part. I would also like to thank José Ignacio Hualde and Jennifer Cole, my former professors at the University of Illinois at Urbana-Champaign. During my early years as a graduate student, they taught me the love of linguistic research and gave me the encouragement and confidence to pursue it.

Second, I would like to thank those in Paris and Madrid who helped me collect the corpora used in this dissertation. Cécile Fougeron in Paris and Manolo Pardo in Madrid welcomed me in their labs during the initial part of my dissertation. Along with colleagues, students and staff in their labs, they provided ideal settings for recording of my corpora of spontaneous French and Spanish, and helped with the technical aspects of the process as well. I would also like to thank Ton Wempe from the University of Amsterdam for providing assistance with the recording equipment, and Louis ten Bosch for helping with the forced alignment of the Spanish corpus. Luuk van der Knaap and the Technical Group at the MPI provided useful help with archiving the corpora.

Finally, I would like to thank my colleagues at the Max Planck Institute for Psycholinguistics and at the Radboud University in Nijmegen for creating a friendly atmosphere at the workplace during all these years. Special thanks go to my office-mates Marco, Lucie and Iris for listening to my silly stories and numerous complaints, and to Christophe, Kevin, Vitoria, Barbara, Michele, Giusy, Sarah, Ely, Kaoru, Jeremy, Christina, Steve, Eric, Maarten, Joost, Helen, Inge, Cristel, Simone, Dan and Jessamyn for so many fun moments in, around, and outside the office.



## Bibliography

- Adda-Decker, M., de Mareüil, P. B., Adda, G., and Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, 46(2):119–139.
- Adda-Decker, M. and Snoeren, N. D. (in press). Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics*.
- Adda-Decker, M. and Snoeren, N. D. (2011). Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics*, 39(3):261–270, 2011.
- Armstrong, N. and Smith, A. (2002). The influence of linguistic and social factors on the recent decline of French *ne*. *Journal of French Language Studies*, 12(1):23–41.
- Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56.
- Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119(5):3048–3058.
- Baayen, H. (2008). *LanguageR: Data sets and functions with Analyzing Linguistic Data: A practical introduction to statistics*. R package version 0.953.
- Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Engen, K. J. V., and Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, 39(1):1–17.
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1):1–22.
- Barnes, J. and Kavitskaya, D. (2002). Phonetic analogy and schwa deletion in French. In *Proceedings of the Twenty-Seventh Annual Meeting of the Berkeley Linguistic Society*, pages 39–50.

- Barras, C., Geoffrois, E., Wu, Z., and Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1-2):5–22.
- Bates, D. and Sarkar, D. (2006). lme4: Linear mixed-effects models using S4 classes. URL: <http://CRAN.R-project.org>, R package version 0.9975-10.
- Beckman, M. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In Keating, P., editor, *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, pages 7–33. Cambridge University Press.
- Beckman, M. E. and Shoji, A. (1984). Spectral and perceptual evidence for CV coarticulation in devoiced /si/ and /syu/ in Japanese. *Phonetica*, 41(2):61–71.
- Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4):591–627.
- Bell, A., Brenier, J., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, 113(2):1001–1024.
- Blanche-Benveniste, C. (1990). *Le français parlé, études grammaticales*. Éditions du CNRS, Paris.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, pages 97–110.
- Boersma, P. and Weenink, D. (2009). Praat: doing phonetics by computer (version 5.1.18).
- Boyce, S. E. (1990). Coarticulatory organization for lip rounding in Turkish and English. *Journal of the Acoustical Society of America*, 88(6):2584–2595.
- Bradlow, A. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97(3):1916–1924.

- Browman, C. and Goldstein, L. (1990). Tiers in Articulatory Phonology, with some implications for casual speech. In Kingston, J. and Beckman, M., editors, *Papers in Laboratory Phonology 1: between the grammar and the physics of speech*, pages 341–376. Cambridge University Press, Cambridge.
- Browman, C. and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49(3-4):155–180.
- Brown, E. and Cacoulios, R. T. (2003). Spanish /s/: A different story from beginning (initial) to end (final). In Núñez-Cedeño, R., López, L., and Cameron, R., editors, *A Romance perspective in language knowledge and use. Selected papers from the 31st linguistic symposium of Romance languages (LSRL)*, pages 22–38. John Benjamins.
- Brown, E. K. (2009). The relative importance of lexical frequency in syllable- and word-final /s/ reduction in Cali, Colombia. In Collentine, J., editor, *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, pages 165–178, Somerville, MA. Cascadilla Proceedings Project.
- Bürki, A., Fougeron, C., and Gendrot, C. (2007). On the categorical nature of the process involved in schwa elision in French. In *Proceedings of Interspeech 2007*.
- Bürki, A., Ernestus, M., and Frauenfelder, U. H. (2010). Is there only one “fenêtre” in the production lexicon? on-line evidence on the nature of phonological representations of pronunciation variants for french schwa words. *Journal of Memory and Language*, 62(4):421–437.
- Bürki, A., Fougeron, C., Gendrot, C. and Frauenfelder, U. H. (2011). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics*, 39(3):279–288, 2011.
- Bybee, J. (2001). *Phonology and language use*. Cambridge University Press, Cambridge.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14(3):261–290.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24(2):209–244.
- Cedergren, H. . L. S. (1985). La chute des voyelles hautes en français de Montréal: ‘As-tu entendu la belle syncope?’. In Lemieux, M. and Cedergren, H., editors, *Les*

- tendances dynamiques du français parlé à Montréal*, pages 57–145. Bibliothèque nationale du Québec.
- Cho, T. and Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27(2):207–229.
- Cho, T. and McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33(2):121–157.
- Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper and Row, New York.
- Clark, H. (1996). *Using Language*. Cambridge University Press.
- Clark, H. and Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3):201–242.
- Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech. *Journal of Phonetics*, 35(2):180–209.
- Côté, M.-H. and Morrison, G. S. (2007). The nature of the schwa-zero alternation in French clitics: experimental and non-experimental evidence. *Journal of French Language Studies*, 17(2):159–186.
- Coveney, A. (2001). *The Sounds of Contemporary French*. Elm Bank Publications, Exeter.
- Coveney, A. B. (1996). *Variability in Spoken French. A Sociolinguistic Study of Interrogation and Negation*. Elm Bank Publications, Exeter.
- Crystal, T. and House, A. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, 83(4):1553–1573.
- Dalby, J. (1986). *Phonetic Structure of Fast Speech in American English*. Indiana University Linguistics Club, Bloomington.
- Dauer, R. M. (1980). The reduction of unstressed high vowels in modern Greek. *Journal of the International Phonetic Association*, 10(1-2):17–27.
- Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica*, 63(2-3):79–112.
- Delattre, P. (1969). An acoustic study of vowel reduction in four languages. *International Review of Applied Linguistics*, 7(4):295–325.



- Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français: première partie. *French Language Studies*, 9(2):143–179.
- Disner, S. F. (1983). Vowel quality: The relation between universal and language-specific factors. *UCLA Working Papers in Phonetics*, 58.
- Duez, D. (1995). On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics*, 23(4):407–427.
- 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.
- Egging, S. and Slade, D. (1997). *Analysing Casual Conversation*. Cassell Academic, London.
- Ellis, L. and Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics*, 30(3):373–396.
- Erker, D. G. (2010) A subsegmental approach to coda /s/ weakening in Dominican Spanish. *International Journal of the Sociology of Language*, (203):9–26, 2010.
- Ernestus, M. (2000). *Voice assimilation and segment reduction in Dutch: A corpus-based study of the phonology-phonetics interface*. LOT, Utrecht, The Netherlands.
- Ernestus, M., Lahey, M., Verhees, F. and Baayen, H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120(2):1040–1051, 2006.
- Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America*, 126(3):1379–1393.
- Fagyal, Z. (1998). Le retour du e final en français parisien: changement phonétique conditionné par la prosodie. In Englebert, A., Pierrard, M., Rosier, L., and van Raemdonck, D., editors, *Vivacité et diversité de la variation linguistique*, volume 3, pages 151–160. Max Niemeyer Verlag.
- Fagyal, Z., Kibbee, D., and Jenkins, F. (2006). *French: a Linguistic Introduction*. Cambridge University Press.
- Fagyal, Z. and Moisset, C. (1999). Sound change and articulatory release: where and why are high vowels devoiced in Parisian French. In *Proc. of the 14th ICPHS*, pages 309–312, San Francisco.

- Farnetani, E. and Busà, M. G. (1994a). Consonant-to-consonant interactions in clusters: categorial or continuous processes? *Quaderni del Centro di Studio per le Ricerche di Fonetica CNR Padova*, 13:220–245.
- Farnetani, E. and Busà, M. G. (1994b). Italian clusters in continuous speech. *Quaderni del Centro di Studio per le Ricerche di Fonetica CNR Padova*, 13:248–251.
- File-Muriel, R. (2007). *A Study Of Lenition: The Role Of Lexical Frequency And Phonetic Context In The Weakening Of Lexical /s/ In The Spanish Of Barranquilla*. PhD thesis, Indiana University, Bloomington.
- File-Muriel, R. J. and Brown, E. K. (2011). The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change*, 23(223–243), 2011.
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29(2):109–135.
- Fougeron, C. and Jun, S.-A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics*, 26(1):45–69.
- Fougeron, C. and Smith, C. (1993). Illustration of the IPA: French. *Journal of the International Phonetic Association*, 23:73–76.
- Fougeron, C. and Steriade, D. (1997). Does deletion of French schwa lead to neutralization of lexical distinctions? In *Proceedings of Eurospeech97*, volume 2, pages 943–946.
- Fowler, C. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36(2-3):171–195.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34(6):709–738.
- Galliano, S., Geoffrois, E., Mostefa, D., Choukri, K., Bonastre, J.-F., and Gravier, J. (2005). ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast Newshase II Evaluation Campaign for the Rich Transcription of French Broadcast News. *Proceeding of Interspeech 2005*, pages 2453–2456.
- Garrett, S., Morton, T. and McLemore, C. (1997). *LDC Spanish Lexicon*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, 1997.

- Gauvain, J.-L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V., Lamel, L., and Schwenk, H. (2005). Where are we in transcribing French broadcast news? *Proceeding of Interspeech 2005*, pages 1665–1668.
- Grabe, E. and Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology Vol. 7*, pages 515–546. Mouton De Gruyter.
- Gubian, M., Torreira, F., Strik, H., and Boves, L. (2009). Functional Data Analysis as a tool for analyzing speech dynamics: a case study on the French word *c'était*. In *Proceedings Interspeech 2009*, pages 2199–2202. Causal Productions Pty Ltd.
- Guy, G. R. (1994). Contextual conditioning in variable lexical phonology. *Language Variation and Change*, 3(2):223–239.
- Hallé, P. and Adda-Decker, M. (2007). Voicing assimilation in journalistic speech. In *Proceedings of 16th International Conference of Phonetic Sciences*, pages 493–496.
- Hernández-Campoy, J. M. and Trudgill, P. (2002). Functional compensatio and Southern Peninsular Spanish /s/ loss. *Folia Linguistica Historica*, 23(1-2):31–57.
- Hirose, H. (1971). The activity of the adductor laryngeal muscles in respect to vowel devoicing in Japanese. *Phonetica*, 23(3):156–170.
- Honikman, B. (1964). Articulatory settings. In Abercrombie, D., editor, *In Honour of Daniel Jones*, pages 73–84. Longmans, London.
- Honorof, D. (1999). *Articulatory gestures and Spanish nasal assimilation*. PhD thesis, Yale University.
- Hualde, J. I. (2009). Unstressed words in Spanish. *Language Sciences*, 31(2-3):199–212.
- Hualde, J. I., Simonet, M., and Nadeu, M. (to appear). Consonant lenition and phonological recategorization. *Journal of Laboratory Phonology*.
- Hundley, J. E. (1987). Constraints on plural marker deletion in Peruvian Spanish. *Hispania*, 70(4):891–894.
- Jakobson, R. and Halle, M. (1968). Phonology in relation to phonetics. In Malmberg, B., editor, *Manual of Phonetics*, pages 411–449. North Holland, Amsterdam.
- Jannedy, S. (1995). Gestural phasing as an explanation for vowel devoicing in Turkish. *Ohio State University Working Papers in Linguistics*, 45:56–84.

- Johnson, K. (2004). Massive reduction in conversational American English. In Yoneyama, K. and Maekawa, K., editors, *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, pages 29–54, Tokyo, Japan. The National International Institute for Japanese Language.
- Jousse, V., Estève, Y., Béchet, F., Bazillon, T., and Linares, G. (2008). Caractérisation et détection de parole spontanée dans de larges collections de documents audio. In *Proceedings of JEP 2008*, Avignon.
- Jun, S.-A. and Fougeron, C. (2002). Realizations of accentual phrase in French. *Probus*, 14(1):147–172.
- Keating, P. (2006). Phonetic encoding of prosodic structure. In Harrington, J. and Tabain, M., editors, *Speech production: Models, phonetic processes, and techniques*, Macquarie Monographs in Cognitive Science, pages 167–186. Psychology Press, New York and Hove.
- Keating, P., Cho, T., Fougeron, C., and Hsu, C. S. (2003). Domain-initial strengthening in four languages. In Local, J., Ogden, R., and Temple, R., editors, *Phonetic interpretation: Papers in Laboratory Phonology VI*, pages 143–161. Cambridge University Press.
- Kemps, R., Ernestus, M., Schreuder, R., and Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90(1-3):117–127.
- Klatt, D. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research*, 17(1):51–63.
- Kochetov, A. and Pouplier, M. (2008). Phonetic variability and grammatical knowledge : an articulatory study of Korean place assimilation. *Phonology*, 25(3):399–431.
- Kohler, K. (2000). Investigating unscripted speech: Implications for phonetics and phonology. *Phonetica*, 57(2-4):85–94.
- Koopmans-van Beinum, F. (1985). *Vowel Contrast Reduction, An acoustic and Perceptual study of Dutch Vowels in Various Speech Conditions*. PhD thesis, University of Amsterdam.
- Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America*, 121(4):2261–2271.

- Kuzla, C., Cho, T., and Ernestus, M. (2007). Prosodic strengthening of German fricatives in duration and assimilatory devoicing. *Journal of Phonetics*, 35(3):301–320.
- Labov, W. (2004). Quantitative analysis of linguistic variation. In Ammon, U., Dittmer, N., Mattheier, K. J., and Trudgill, P., editors, *Sociolinguistics: An International Handbook of the Science of Language and Society*, pages 6–21. De Gruyter.
- Lafford, B. (1989). Is functionalism a fact? data from the Caribbean. *Hispanic Linguistics*, 3:49–74.
- Lamel, L., Rosset, S., Gauvain, J., Bennacef, S., Garnier-Rizet, M., and Prouts, B. (2000). The LIMSI ARISE system. *Speech Communication*, 31(4):339–354.
- Lewis, A. (2001). *Weakening of Intervocalic /p, t, k/ in Two Spanish Dialects: Toward the Quantification of Lenition Processes*. PhD thesis, University of Illinois at Urbana-Champaign.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(11):1773–1781.
- Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W. J. and Marchal, A., editors, *Speech Production and Speech Modelling*. Kluwer Academic Publishers.
- Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, 31(3-4):321–339.
- Local, J. (2007). Phonetic detail and the organisation of talk-in-interaction. In *Proceedings of the XVIth International Congress of Phonetic Sciences. Saarbruecken, Germany: 16th ICPhS Organizing Committee*.
- Machuca-Ayuso, M. J. (1997). *Las obstruyentes no continuas del español: relación entre las categorías fonéticas y fonológicas en habla esponánea*. PhD thesis, Universitat Autònoma de Barcelona.
- Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, 88(3):1286–1298.
- Manuel, S. (1992). Recovery of “deleted” schwa. In *Perilius: Papers from the Symposium on Current Phonetic Research Paradigms for Speech Motor Control*, pages 115–118), Stockholm. University of Stockholm.

- Manuel, S. (1999). Cross-language studies: Relating language-particular coarticulation patterns to other language-particular facts. In Hardcastle, W. J. and Hewlett, N., editors, *Coarticulation: Theory, Data and Techniques*, pages 179–98. Cambridge University Press.
- Martin, P. (2004). Dévoisement vocalique en français. *La Linguistique*, 40(2):3–22.
- Martínez Celdrán, E. (2009). Sonorización de las oclusivas sordas en una hablante murciana: Problemas que plantea. *Estudios de Fonética Experimental*, 18:255–271.
- Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Besnerais, M. L. (2003). Production and perception of vowels: does the density of the system play a role? In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 723–726, Barcelona, Spain.
- Meunier, C. M., Meynadier, Y., and Espesser, R. (2008). Voyelles brèves en parole conversationnelle. In *Actes, Journées d'Etude sur la Parole 2008*, pages 97–100.
- Minnick Fox, M. (2006). *Usage-based Effects in Latin American Spanish Syllable-final /s/ Lenition*. PhD thesis, University of Pennsylvania, Philadelphia.
- Mo, Y. (2007). Temporal, spectral evidence of devoiced vowels in Korean. In *Proceedings of ICPHS 2007, Saarbrücken*.
- Moon, S. and Lindblom, B. (1994). Interaction between duration, context, and speaking style in english stressed vowels. *The Journal of the Acoustical Society of America*, 96(1):40–55.
- Moore, R. K. (2003). A comparison of the data requirements of automatic speech recognition systems and human listeners. In *Proceedings of Eurospeech 2003*, pages 2581–2584, Geneva, Switzerland.
- Moore, R. K. (2005). Cognitive informatics: The future of spoken language processing? In *Proceedings of the 10th International Conference on Speech and Computer*, Patras, Greece.
- Moreno Sandoval, A., Toledano, D. T., de la Torre, R., Garrote, M., and Guirao, J. M. (2008). Developing a phonemic and syllabic frequency inventory for spontaneous spoken Castilian Spanish and their comparison to text-based inventories. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., and Tapias, D., editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, pages 1097–1100, Marrakech, Morocco. European Language Resources Association (ELRA).

- Navarro Tomás, T. (1977). *Manual de pronunciación española*. Publicaciones de la Revista de Filología Española. Consejo Superior de Investigaciones Científicas, Madrid, 19th edition.
- Nguyen, N. and Fagyal, Z. (2008). Acoustic aspects of vowel harmony in French. *Journal of Phonetics*, 36(1):1–27.
- Nolan, F. (1992). The descriptive role of segments. In Docherty, G. J. and Ladd, R., editors, *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pages 261–280. Cambridge University Press.
- Nolan, F., Holst, T., and and, B. K. (1996). Modelling [s] to [ʃ] accommodation in English. *Journal of Phonetics*, 24(1):113–137.
- Oftedal, M. (1985). *Lenition in Celtic and in Insular Spanish: the secondary voicing of stops in Gran Canaria*. Oslo University Press.
- Ohala, J. and Riordan, C. (1979). Passive vocal tract enlargement during voiced stops. In Wolf, J. J. and Klatt, D. H., editors, *Speech Communication Papers*, pages 89 – 92. Acoust. Soc. of Am.
- Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39:151–168.
- Plug, L. (2005). From words to actions: the phonetics of eigenlijk in two communicative contexts. *Phonetica*, 62(2-4):131–145.
- Pluymaekers, M., Ernestus, M., and Baayen, H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62(2-4):146–159.
- Poplack, S. (1980). Deletion and disambiguation in Puerto Rican Spanish. *Language*, 56(2):371–385.
- Prieto, P. and Ortega-Llebaria, M. (2006). Stress and accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance. In *Proceedings of Speech Prosody 2006*, Dresden.
- Quilis, A. and Graell, M. (1992). La lengua española en Panamá. *Revista de Filología Española*, 72(3-4):583–638.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, <http://www.R-project.org>.
- Ramsay, J. O. and Silverman, B. (1997). *Functional Data Analysis*. Springer-Verlag.

- Recasens, D. and Espinosa, A. (2009). Acoustics and perception of velar softening for unaspirated stops. *Journal of Phonetics*, 37(2):189–211.
- Rossi, M. (1999). *L'intonation: le système du français*. Orpheus, Gap.
- Hernández, J. V. (1984). *El consonantismo en Cuba*. Editorial de Ciencias Sociales, La Habana, Cuba.
- Salvador, G. (1968). Neutralización de G-/K- en español. In C.S.I.C., editor, *Actas del XI Congreso Internacional de Lingüística y Filología Románicas*, Madrid.
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer.
- Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(1):1–63.
- 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(1):96–109.
- Serpollet, N., Bergounioux, G., Chesneau, A., and Walter, R. (2007). A large reference corpus for spoken French: ESLO 1 and 2 and its variations. In *Proceedings from Corpus Linguistics Conference Series*, University of Birmingham.
- Shockey, L. (2003). *Sound Patterns of Spoken English*. Blackwell.
- Shriberg, E. (2001). To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association*, 31(1):153–169.
- Smith, C. L. (2003). Vowel devoicing in contemporary French. *Journal of French Language Studies*, 13(2):177–194.
- Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38(1):1–23.
- Stetson, R. (1951). *Motor Phonetics*. North-Holland, Amsterdam.
- Sun-Ah Jun, Mary Beckman, H.-J. L. (1998). Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents. *UCLA Working Papers in Phonetics*, 96:43–66.
- Torreblanca, M. (1976). La sonorización de las oclusivas sordas en el habla toledana. *Boletín de la Real Academia Española de la Lengua*, 56:117–146.
- Torreblanca, M. (1986). La 's' sonora prevocálica en el español moderno. *Thesaurus*, 41:59–69.



- Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52(3):201–221.
- Torreira, F. and Ernestus, M. (2010a). The Nijmegen Corpus of Casual Spanish. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Torreira, F. and Ernestus, M. (2010b). Phrase-medial vowel devoicing in spontaneous French. In *Proceedings of Interspeech 2010*, pages 2006–2009, Makuhari, Japan.
- Torreira, F. and Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Journal of Laboratory Phonology*, 2:331–353.
- Tranel, B. (1987). *The Sounds of French*. Cambridge University Press.
- Trujillo, R. (1970). *Resultado de dos encuestas dialectales en Masca*. Instituto de Estudios Canarios, La Laguna, Spain.
- Turk, A., Nakai, S., and Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In Sudhoff, S., editor, *Methods in Empirical Prosody Research*. Mouton De Gruyter.
- Valdman, A. (2000). La langue des faubourgs et des banlieues : de l'argot au français populaire. *The French Review*, 73(6):1179–1192.
- van Son, R. and Pols, L. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 88(4):1683–1693.
- van Son, R. and Pols, L. (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 92(1):121–127.
- van Son, R. and Pols, L. (2003). An acoustic model of communicative efficiency in consonants and vowels taking into account context distinctiveness. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.*, pages 2141–2144.
- van Son, R. and van Santen, J. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47(1-2):100–123.

- Waals, J. (1999). *An Experimental View of the Dutch Syllable*. LOT, Utrecht, The Netherlands.
- Warner, N. (2005). Reduction of flaps: speech style, phonological environment, and variability. *Journal of the Acoustical Society of America*, 118:2035.
- Yoshioka, H. (1981). Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica*, 38(4):236–251.
- Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (2002). *The HTK Book 3.2*. Entropic.
- Zsiga, E. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English. In Connell, B. and Arvaniti, A., editors, *Papers in Laboratory Phonology 4: phonology and phonetic evidence*, pages 282–302. Cambridge University Press, Cambridge.
- Zwicky, A. (1972). Note on a phonological hierarchy in English. In Stockwell, R. and Macaulay, R., editors, *Linguistic Change and Generative Theory*. Indiana University Press.

# Curriculum vitae

---

Francisco Torreira was born in Cádiz, Spain on March 29, 1978. From 1996 to 2000, he studied Translation at the Institut de Traducteurs et Interprètes in Brussels. From 2001 to 2004, he studied Musicology and Linguistics at the Université Libre de Bruxelles. In 2004, he started graduate studies at the University of Illinois at Urbana-Champaign, where he obtained a Masters Degree in Hispanic Linguistics. In September 2007, he began doctoral work at the Centre for Language and Speech Technology at the Radboud Universiteit Nijmegen. He is now a research staff member in the Language & Cognition group at the Max Planck Institute for Psycholinguistics in Nijmegen.

## List of publications

### *Journal articles*

Torreira F. and Ernestus, M. (submitted). Weakening of intervocalic /s/ in the Nijmegen Corpus of Casual Spanish.

Torreira, F. (to appear). Investigating the nature of aspirated stops in Western Andalusian Spanish. *Journal of the International Phonetic Association*.

Torreira, F. and Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Journal of Laboratory Phonology*, 2:331–353.

Torreira, F. and Ernestus, M. (2011). Vowel elision in casual French : the case of vowel /e/ in the word c'était. *Journal of Phonetics*, 39:50-58.

Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52:201-221.

Hualde, J.I., Lujanbio, O., and Torreira, F. (2008). Lexical tone and stress in Goizueta Basque, *Journal of the International Phonetic Association*, 38:1-24.

Hualde, J.I., Simonet, M., and Torreira, F. (2008). Postlexical contraction of mid vowels in Spanish, *Lingua*, 118:1906-1925.

Prieto, P. and Torreira, F. (2007). The segmental anchoring hypothesis revisited: syllable structure and speech rate effects on peak timing in Spanish, *Journal of Phonetics*, 35:473-500.

### **Refereed conference papers and book chapters**

Torreira, F. and Ernestus, M. (2010). Phrase-medial vowel devoicing in spontaneous French. *Proc. of Interspeech 2010*.

Torreira, F. and Ernestus, M. (2010). The Nijmegen Corpus of Casual Spanish. *Proc. of LREC 2010*.

Gubian, M., Torreira, F., Strik, H., and Boves, L. (2009). Functional Data Analysis as a tool for analyzing speech dynamics: a case study on the French word c'était. *Proc. of Interspeech 2009*.

Torreira, F. and Ernestus, M. (2009). Probabilistic effects on French [t] duration. *Proc. of Interspeech 2009*.

Torreira, F. (2007) Tonal realization of syllabic affiliation in Spanish, in *Proc. of the 16th International Congress of Phonetic Sciences*.

Torreira, F. (2007). Pre- and postaspirated stops in Andalusian Spanish, in P.Prieto, J.Mascaró, and M.J. Solé (ed.) *Segmental and Prosodic Issues in Romance Phonology*, John Benjamins.

Torreira, F. (2006). Coarticulation between aspirated-s and voiceless stops in Spanish: an interdialectal comparison, in *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, Cascadilla Press.

## MPI Series in Psycholinguistics

---

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. Miranda van Turenout
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. Niels O. Schiller
3. Lexical access in the production of ellipsis and pronouns. Bernadette M. Schmitt
4. The open-/closed-class distinction in spoken-word recognition. Alette Haveman
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. Kay Behnke
6. Gesture and speech production. Jan-Peter de Ruiter
7. Comparative intonational phonology: English and German. Esther Grabe
8. Finiteness in adult and child German. Ingeborg Lasser
9. Language input for word discovery. Joost van de Weijer
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. James Essegbey
11. Producing past and plural inflections. Dirk Janssen
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. Anna Margetts
13. From speech to words. Arie van der Lugt
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. Eva Schultze-Berndt

15. Interpreting indefinites: An experimental study of children's language comprehension. Irene Krämer
16. Language-specific listening: The case of phonetic sequences. Andrea Weber
17. Moving eyes and naming objects. Femke van der Meulen
18. Analogy in morphology: The selection of linking elements in Dutch compounds. Andrea Krott
19. Morphology in speech comprehension. Kerstin Mauth
20. Morphological families in the mental lexicon. Nivja H. de Jong
21. Fixed expressions and the production of idioms. Simone A. Sprenger
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). Birgit Hellwig
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. Fermín Moscoso del Prado Martín
24. Contextual influences on spoken-word processing: An electrophysiological approach. Daniëlle van den Brink
25. Perceptual relevance of prevoicing in Dutch. Petra M. van Alphen
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. Joana Cholin
27. Producing complex spoken numerals for time and space. Marjolein Meeuwissen
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. Rachèl J. J. K. Kemps
29. At the same time...: The expression of simultaneity in learner varieties. Barbara Schmiedtová
30. A grammar of Jalonke argument structure. Friederike Lüpke
31. Agrammatic comprehension: An electrophysiological approach. Marlies Wassenaar
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). Frank Seifart

33. Prosodically-conditioned detail in the recognition of spoken words. Anne Pier Salverda
34. Phonetic and lexical processing in a second language. Mirjam Broersma
35. Retrieving semantic and syntactic word properties. Oliver Müller
36. Lexically-guided perceptual learning in speech processing. Frank Eisner
37. Sensitivity to detailed acoustic information in word recognition. Keren B. Shatzman
38. The relationship between spoken word production and comprehension. Rebecca Özdemir
39. Disfluency: Interrupting speech and gesture. Mandana Seyfeddinipur
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. Christiane Dietrich
41. Cognitive cladistics and the relativity of spatial cognition. Daniel B.M. Haun
42. The acquisition of auditory categories. Martijn Goudbeek
43. Affix reduction in spoken Dutch. Mark Pluymaekers
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. Valesca Kooijman
45. Space and iconicity in German Sign Language (DGS). Pamela Perniss
46. On the production of morphologically complex words with special attention to effects of frequency. Heidrun Bien
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. Amanda Brown
48. The acquisition of verb compounding in Mandarin Chinese. Jidong Chen
49. Phoneme inventories and patterns of speech sound perception. Anita Wagner
50. Lexical processing of morphologically complex words: An information-theoretical perspective. Victor Kuperman
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. Claudia Wegener

52. Prosodic structure in speech production and perception. Claudia Kuzla
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. Sarah Schimke
54. Studies on intonation and information structure in child and adult German. Laura de Ruiter
55. Processing the fine temporal structure of spoken words. Eva Reinisch
56. Semantics and (ir)regular inflection in morphological processing. Wieke Tabak
57. Processing strongly reduced forms in casual speech. Susanne Brouwer
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. Miriam Ellert
59. Lexical interactions in non-native speech comprehension: Evidence from electroencephalography, eye-tracking, and functional magnetic resonance imaging. Ian FitzPatrick
60. Processing casual speech in native and non-native language. Annelie Tuinman
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. Stuart Robinson
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. Sonja Gipper
63. The influence of information structure on language comprehension: A neurocognitive perspective. Lin Wang
64. The meaning and use of ideophones in Siwu. Mark Dingemans
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. Marco van de Ven
66. Speech reduction in spontaneous French and Spanish. Francisco Torreira
67. The relevance of early word recognition: Insights from the infant brain. Caroline Junge
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. Matthias J. Sjerps