



**HAL**  
open science

# Exploring the Nuances of Reduction in Conversational Speech: Lexicalized and Non Lexicalized Reductions

Kübra Bodur, Corinne Fredouille, Stéphane Rauzy, Christine Meunier

► **To cite this version:**

Kübra Bodur, Corinne Fredouille, Stéphane Rauzy, Christine Meunier. Exploring the Nuances of Reduction in Conversational Speech: Lexicalized and Non Lexicalized Reductions. 2023. hal-04414140

**HAL Id: hal-04414140**

**<https://hal.science/hal-04414140v1>**

Preprint submitted on 24 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Exploring the Nuances of Reduction in Conversational Speech: Lexicalized and Non-Lexicalized Reductions

*Kübra Bodur<sup>1</sup>, Corinne Fredouille<sup>2</sup>, Stéphane Rauzy<sup>1</sup>, Christine Meunier<sup>1</sup>*

<sup>1</sup> Aix-Marseille Université, LPL, CNRS, France

<sup>2</sup> Avignon Université, LIA, France

kubra.bodur@univ-amu.fr

## Abstract

In spoken language, a significant proportion of words are produced with missing or underspecified phonetic forms, a process known as phonetic reduction. We have established two different types of reduction for studying reductions in spontaneous speech: lexicalized and non-lexicalized reductions. Lexicalized reductions are regularly encountered in daily language, produced by nearly all speakers, and are related to the lexicon while the non-lexicalized reductions are not driven only by the lexicon, occurring at specific points of speech without regular reduction patterns. Instead, this type of reduction is influenced by various linguistic and conversational factors. Previously existing methods are not sufficient in detecting all potential reductions produced in the speech flow.

In this study, we propose a novel bottom-up approach for identifying reduction prone areas in a corpus of conversations in French, in addition to a manual top-down detection based on the known reduced forms. Our novel approach involves the development of a script that automatically detects sequences containing at least 6 phonemes in a 230ms window -indicating temporal reduction- allowing the detection of both the lexicalized and non-lexicalized sequences.

Manually verified extracted sequences were analyzed for the distribution and frequency of reductions, speaker characteristics, as well as the words and phonemes constituting these reduced sequences. Overall, our findings revealed significant variability in terms of reduction frequency and speaker performance, depending on the reduction type, thereby validating the distinction we established between the two types of reductions. Furthermore, the characteristics of the words and the phonemes present in these sequences were found to influence their likelihood of being reduced.

**Keywords:** spontaneous speech, phonetic reduction, conversation, variation, speaker behavior

## INTRODUCTION

Spoken language is one of the essential ways in which communication is made possible between humans, enabling them to exchange information, ideas and build connections on a daily basis. The study of spoken language is of utmost importance for understanding human cognition and developing technological tools capable of interacting with humans. However, despite the growing body of literature on speech, the characteristics of speech in spontaneous conversations remain less well-characterized due to their inherent complexities. The spontaneous productions of speakers often exhibit several phenomena specific to casual speech (feedback, repetitions, unfinished utterances, disfluencies, etc. (Prévost, et al., 2022)). It is imperative to provide a more comprehensive explanation of the intricate organization and dynamics of these phenomena to enhance our understanding of speech production, as they tend to have a significant impact on the quality and characteristics of the produced sounds.

The variation in the speech signal is one of the intrinsic features of spoken language (Ernestus, 2000; Johnson, 2004; Meunier & Espesser, 2011), appearing as a result of the permanent adaptation of one's productions to the speaking situation (Lindblom, 1990). Speaking in interaction involves permanent adjustments supported by different stages of speech production such as convergence between speakers (e.g., Pardo, 2013; Guardiola & Bertrand, 2013). The analysis of large corpora of spontaneous speech, coupled with the use of automated tools for their exploration, has enabled the identification of variations in the speech signal. This includes phenomena like reduction, which provides valuable insights into language production processes and cognitive load. Reduction in conversations has proven to be challenging to study, as its manifestation is influenced by interactions between various linguistic domains. Reduction can be defined as the phenomenon in which linguistic units (e.g., phonemes, syllables, words) are realized with relatively less acoustic and phonetic substance (Clopper, et al., 2018) often resulting in shorter durations and less precise articulations. Indeed, the productions of speakers in spontaneous conversations are typically less clearly articulated compared to formal or scripted speech (Ernestus & Warner, 2011). Phonetic segments may be weakened or even completely absent in some cases (Schuppler, Ernestus, Scharenborg, & Boves, 2011; Niebuhr & Kohler, 2011; Ernestus, 2014) all the while allowing the interaction to flow smoothly between speakers and listeners. In addition to absent segments, coalescence phenomena are frequently encountered, making it challenging to determine the nature of the phonemes realized in the reduced sequence, suggesting that words in conversation are not always produced as distinct phonemes. These reduced realizations can exhibit characteristics vastly different from those in scripted speech (Prévot et al., 2022). The reduction process and its distribution in spontaneous conversations remain poorly characterized, posing a challenge to comprehending how interactions are facilitated when phonetic information is lacking.

Previous research has indicated that speech reduction is influenced by a range of linguistic and conversational factors, including semantic and lexical predictability (Aylett & Turk, 2006; Clopper & Pierrehumbert, 2008; Turnbull, 2017), lexical frequency (suggesting that common words are more likely to be reduced than low-frequency words) (Pluymaekers, Ernestus, & Baayen, 2005; Bybee, File-Muriel, & de Souza, 2016), repetitive mention in discourse (Baker & Bradlow, 2009), speaking style (Gendrot, Adda-Decker, & Schmid, 2012), and the phonological properties of segments found in reduced sequences (Pharao, 2010; Meunier & Bigi, 2016). Many of these factors are closely related to the lexical properties of languages.

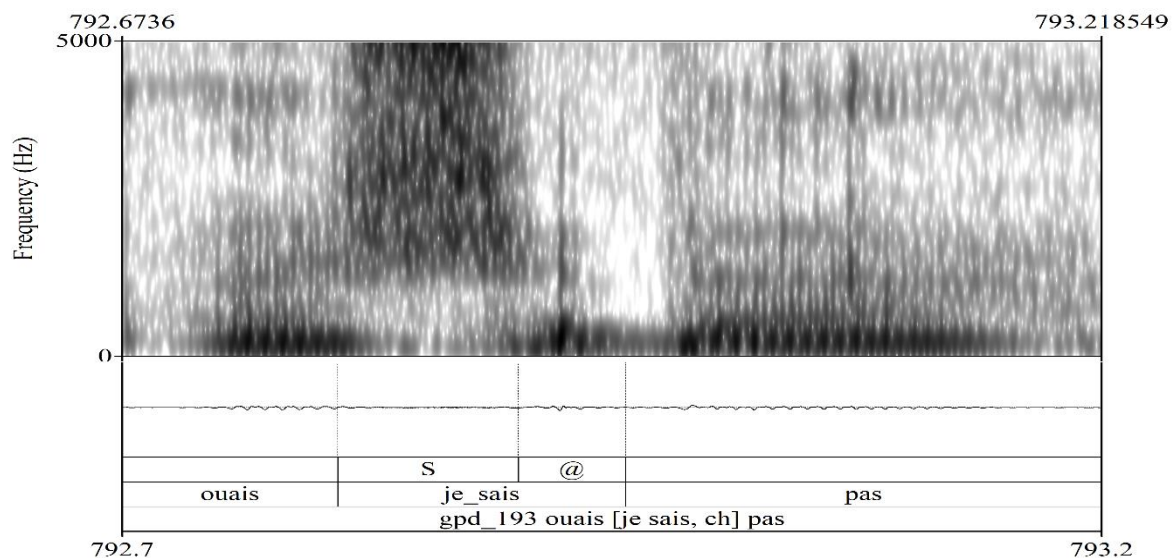


Figure 1. An example of a lexicalized reduction “je sais pas (/ʃɛpa/, “I don’t know”)” (top), involved tokens (middle) and the IPU (bottom).

Reduction is a complex phenomenon that has not been studied in all of its forms. For this reason, we have established a distinction between two types of reductions occurring in casual speech which we will call lexicalized and non-lexicalized reductions from now on. Lexicalized forms can be described as different reduced examples of a same lexicalized representation, relating to words or sequences that occur frequently in everyday language (Bodur, Fredouille, & Meunier, 2022). These forms are more commonly studied in literature and are often recognizable by listeners and transcribers who have perceptible representations of several examples of these sequences (see Figure 1). Moreover, these forms are included in Automatic Speech Recognition (ASR) dictionaries with their reduced counterparts, enabling systems to identify and transcribe the reduced versions of a given sequence. Examples of lexicalized reductions can be found in many languages, often occurring multiple times within a conversation, and produced by various speakers. Examples from English and French are given below based on Ernestus & Warner 2011:

- English : “a little” produced /ɔlə/ ; “yesterday” produced /jɛʃei/
- French : “je sais pas” (I don’t know) produced /ʃɛpa/ ; “c’était” (It was) produced /stɛ/

On the other hand, we have also observed a substantial number of reductions that do not conform to specific sequences and are hardly perceptible to listeners. We refer to these reductions as “non-lexicalized” reductions because they lack regularity, reproducibility, and direct association with specific lexical units. These reductions seem to be influenced by, or interact with various conversational and linguistic levels, including phonology, prosody, and syntax. They might be found scattered across multiple words (as seen in Figure 2), appear only once, and affect any sequence or word in the speech flow. All these factors contribute to the difficulty of their identification in the speech signal and collection for further analyses. In Figure 2, we can observe a temporally reduced sequence, “qui avait amené” (who had brought) which does not consist of the most frequent words, and appears to involve phonemic coalescence (with merged phonemes). Non-lexicalized reductions have not been as extensively studied as lexicalized ones since the

processes involved and the rules governing them are not well understood. This presents two significant challenges: devising appropriate methods for collecting reduced sequences (as they are imperceptible and challenging to visually detect in the speech signal) and determining whether the collected sequences exhibit reduction or not by analyzing their acoustic and phonetic characteristics. Detecting non-lexicalized reductions necessitates innovative methods that could be complementary to the existing approaches, such as the top-down approach, which primarily focuses on the search of the reduced occurrences of words that are known to be reduced frequently.

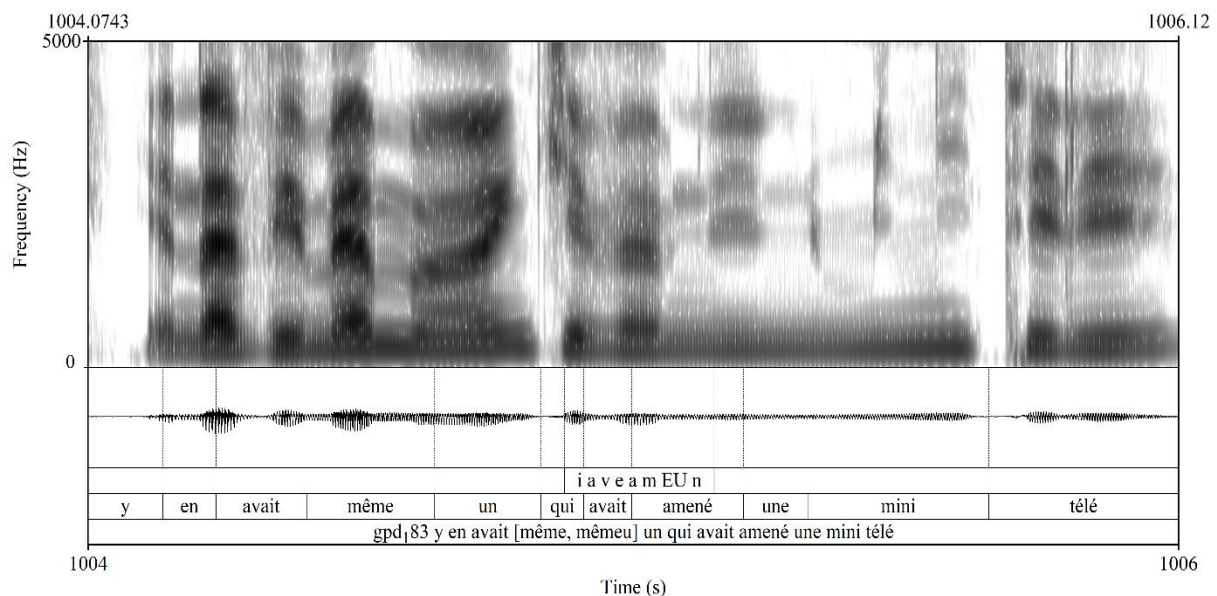


Figure 2. A non-lexicalized reduction: phonemes comprising the reduced sequence “qui avait amené” (“who had brought”, top), involved tokens (middle) and the IPU containing the reduced sequence (bottom).

Previous research on speech reduction has mainly relied on manual detection by human experts or automatic speech recognition systems, such as automatic forced alignment (e.g., Adda-Decker & Snoeren, 2011; Wu & Adda-Decker, 2020). However, manual detection of reductions by listening to the productions and deciding whether they are reduced or not might be difficult as it is time consuming and more importantly, human speech perception tends to be biased by existing linguistic knowledge (Elman & McClelland, 1988). Furthermore, some of the missing phoneme’s acoustic features may even be present in surrounding phonemes, and be perceptible to the listeners’ ears, making the annotation process even more effortful.

In contrast, advances in analyzing large corpora of spontaneous speech and automatic tools that have led to their exploitation, have improved our ability to study variation and phonetic reduction efficiently. Automatic forced alignment involves aligning transcriptions with audio signals to segment the acoustic signal into words and phonemes. This allows for the detection of sequences with extra short segments, indicating potential temporally reduced sequences. Additionally, these reduced segments often lead to a considerable number of automatic transcription errors, making it easier to identify mismatches between orthographic transcriptions and the speaker’s production. Despite these methods, comprehensive studies exploring all forms of reduction in conversation are lacking. Our research aims to bridge this knowledge gap and contribute to a deeper understanding of reduction phenomena.

## The current study

The current study aims at filling the gaps identified in previous research. Many of the previous studies have used top-down approaches, analyzing data based on previously existing knowledge, focusing mostly on well-documented lexicalized sequences with multiple lexical representations. For detecting the reductions that do not have lexicalized exemplars, we propose a bottom-up method which departs from the data in order to highlight less known reduction types. This method would result in capturing various reduced zones in the flow of speech, including both lexicalized and non-lexicalized sequences, facilitating a comprehensive preliminary analysis. This method allows us to collect more reduction examples than is possible with manual detections.

In addition, the corpus we chose for this study features informal and spontaneous speech between colleagues, which may provide more reductions than what can be found in other types of non-prepared speech (e.g., news broadcasts). This emphasis on uncovering lesser-known reduction forms in a natural context is expected to expand our understanding of reduction processes in casual conversations.

By employing two complementary approaches (top-down and bottom-up), we aim to thoroughly investigate various aspects such as the distributions of reductions, contextual factors influencing reduction, components of reduced sequences, and individual variations in reduction patterns. Our study would shed light on one of the most common phenomena in casual speech, contributing to a field that has not been extensively explored.

## METHOD

In this section, we present the corpus used in this study and two complementary approaches for detecting various reduction prone areas in a corpus of conversational speech in French.

### 1. The Corpus

We selected the Corpus of Interactional Data (CID, Bertrand, 2008) in order to conduct analyses on reduction in spontaneous French conversations. The corpus comprises 8 audio-visual recordings of dialogues between pairs of colleagues (10 females and 6 males, *average age*= 34 years) with relatively familiar relationships. In each pair, the speakers shared the same gender. All participants are native French speakers, with half originating from the South-East of France, and the other half from different regions.

The choice of the corpus seems ideal in that it involves spontaneous conversations between familiar colleagues, a context known to exhibit greater variation and speech reduction (Gendrot et al., 2012). The corpus was available with phonetic annotations generated through automatic phonetization and alignment tools capable of processing large volumes of speech data and providing consistent outputs for phonetic studies. The segmentation relies on Enriched Orthographic Transcriptions (Bertrand et al., 2006). An orthographic transcription is converted into phonemes using a grapheme to phoneme converter (Di Cristo & Di Cristo, 2001) which is based on a dictionary of inflected forms and a set of modifiable rules. The output of the converter is a sequence of tokens phonetized in SAMPA (Wells, 1997). Then, based on these phoneme

sequences and corresponding audio signals, the LORIA aligner (Brun et al., 2004) provides temporal location for each phoneme in the signal. The output of the aligner is then evaluated for error rates and corrected manually by human experts. Two versions of the phonetic annotation, v2 and v3, were available with the corpus. The difference is that the latter contains the same orthographic tokens and phonemes while the alignment was performed using SPPAS (Bigi & Hirst, 2012) software with extra manual corrections on the output of alignment, phonetization, tokenization and liaison. Both versions were tested for this study and will be further explained in the bottom-up approach section.

## 2. Detecting Reductions

### 2.1. Top-Down Approach

We initially adopted a top-down approach to investigate reductions in spontaneous speech. For doing so, we began by selecting some frequently encountered lexicalized sequences in French which have been previously documented in the literature (Adda-Decker & Snoeren, 2011; Wu & Adda-Decker, 2020). To expand our inventory, we conducted an informal questionnaire with 5 female students, asking them to provide examples of reductions they commonly produced and perceived in their daily conversations. It is important to note that this inventory is thus based on speakers' representations and is not comprehensive. In the end, we compiled a total of 13 items from different morphosyntactic categories which are presented in Table 1.

<b>Adverbs</b>	<b>Pronoun + Verb</b>	<b>Conjunctions</b>
alors ( <i>then</i> )	je peux ( <i>I can</i> )	alors que ( <i>whereas</i> )
enfin ( <i>finally</i> )	je sais ( <i>I know</i> )	parce que ( <i>because</i> )
puis ( <i>then</i> )	je suis ( <i>I am</i> )	puisque ( <i>since, as</i> )
voilà ( <i>here, there</i> )	tu sais ( <i>you know</i> )	
de toute façon ( <i>anyway</i> )		
quand même ( <i>all the same</i> )		

Table 1. The 13 selected items (and their translations) for the study of lexicalized reductions.

The extraction of the selected lexicalized reduced forms was performed semi-automatically. The 13 selected items were searched in the token (word) tier, which was already available in the CID corpus, using Praat (Boersma & Weenink, 2022) to detect all instances of these items. Then, two steps allowed us to identify an example as reduced:

- i. We utilized the Enriched Orthographic Transcription (*TOE*, Bertrand et al., 2008) available for the CID corpus. In this set of transcriptions, brackets and square brackets indicate the presence of particular reductions or realizations perceived by the annotators.
- ii. Each exemplary sequence found in the word tier has been identified and checked manually on Praat by one of the authors also with the help of the *TOE*. In the case of a detection of reduction, the reduced form was extracted with the three preceding and following tokens (to capture context of occurrence) and was annotated in a new TextGrid with the phonetic transcription corresponding to each form (Figure 1).

Based on the annotations provided by transcribers (in the TOE) and the first author (through manual verifications on Praat), we conducted a top-down analysis of reduced sequences to examine their distribution within the speech flow.

## 2.2. Bottom-Up Approach

To expand our inventory of reductions, we also adopted a bottom-up approach allowing us to discover various reduction prone areas in the corpus. Complementary to the previously described method, this approach focuses not only on the lexicalized reductions (in contrast to the top-down approach) but also on the ones involving multiple words that are likely to be influenced by various linguistic and conversational factors (e.g., phonology, prosody, pragmatics).

### 2.2.1. The Script

To identify reduction-prone areas, it is essential to define two key parameters: (i) an analysis window which is large enough to observe non-lexicalized reductions, and (ii) a threshold in terms of the number of successive phonemes observed within this window, so that the sequence could be considered as a potential reduction when the threshold is exceeded.

We tested a number of window sizes to determine which one would yield more sequences deemed as reductions. Ultimately, we defined a window size of 230ms and set the threshold for the number of phonemes within this window to 6. Other window sizes that we tested were 180ms and 200ms (while keeping the number of phonemes constant) respectively, which pointed at more highly reduced areas in the corpus. However, they yielded a smaller number of detected reductions (see Table 2). The choice of 230ms is thus critical as a larger window accommodates more phonemes and words (especially sequences made up of multiple words), providing a sufficient number of reductions for a more reliable analysis. However, we must point out that this choice of window size limits the scope of reductions we can detect in the corpus. The reductions detected here are constrained by the size of this window. Our aim is therefore not to compile an exhaustive list of reductions, but rather to establish a suitable window size allowing us to observe less-explored reductions.

With the selected window size, the system takes individual alignment files containing phoneme sequences and their temporal boundaries as input. For each phoneme, the script counts from its start time how many phonemes (including the starting phoneme itself) belong into the predefined window (230ms). If the threshold of the number of phonemes (6) is reached, the sequence is saved. As the 230ms window slides, the script continues adding subsequent phonemes to the sequence as long as there are 6 phonemes in each window. The output is a Textgrid file containing the identified reduced sequences along with their start and end times.

Window size	180ms/6 phonemes		200ms/6 phonemes		<b>230ms/6 phonemes</b>	
	v2	v3	v2	v3	v2	v3
Detected reductions	456	95	933	429	2144	1212
Error Rate					36%	21%



*Table 2. The numbers of reductions detected by the script using different window sizes and different phonetic annotations as well as the error rates calculated for the sequences detected using the v2 and v3 for the 230ms window size.*

### 2.2.2. Manual evaluation and errors

Assuming that the precision and the quality of the input is of utmost importance for detecting reductions, we evaluated the script performance using both v2 and v3 versions of phonetic annotations. After obtaining a number of potentially reduced sequences, we manually assessed the accuracy of automatic detection by looking at the reduced sequences given as the output of the script and the corresponding audio-visual signal on Praat.

From the 1212 sequences detected based on v3 annotations, we identified 256 errors, resulting in an error rate of 21% (see Table 2), while this rate was higher for the v2 (36%), even though it detected a greater number of sequences.

Our manual evaluation indicated that v3 provided more accurate results, meaning more accurately detected sequences which should be due to the improved alignment and eliminated unnecessary liaisons. The v2 of the corpus involved more brute annotations and no corrections. As a result, the results presented in the following sections are based on the v3 annotations.

## RESULTS

In this section, we present the results i) based on the annotations of the reductions made by the transcribers and the first author and ii) extracted from the set of reductions considered as correct after performing manual evaluations on the outputs of the script. The analyses were carried out for all 16 speakers of the CID corpus. The results will be presented in three parts: general observations, speaker-related factors, and linguistic factors.

### 1. General Observations

#### 1.1. The Total Numbers of Detected Reductions

We observed how many reductions were detected for both approaches. Adopting the top-down approach, we detected 2566 reduced targets out of a total of 3634 occurrences of the 13 selected target items in the corpus. This indicates that a substantial 71% of these items were identified as reduced in the corpus, representing a notably high proportion.

With the bottom-up approach, we obtained a total of 956 correctly detected reductions that lasted a total of 255.9 seconds within 25009.2 seconds which is the total duration of speech (excluding pauses and laughs) in the whole corpus. This shows that the durations of the reduced sequences correspond to 1% of the whole speech duration.

#### 1.2. The Number of Phonemes Affected by Reduction

As the sequences obtained through our top-down approach were pre-defined, we analyzed only the phonemes found in the sequences detected via the bottom-up approach. Using a custom-made

script, we extracted all the phonemes encountered in reduced areas and counted how many times they occurred in reduced sequences. 6351 phonemes were found in the reduced sequences, making up 2,1% of the total number of phonemes involved in the corpus (297703 phonemes in total).

### 1.3. The Number of Words Affected by Reduction

For the bottom-up approach, we also created a custom R script in order to extract and analyze all the words making up the reduced sequences. The script takes the Textgrid containing the reduced sequences and the one containing orthographic tokens as input. Looking at the start and end times of the detected reduced sequences, it extracts all the words coinciding with these intervals from the token Textgrid. We obtained a list of 716 different words from various morpho-syntactic categories, occurring a total of 3058 times in the detected reduced sequences. This number represents 2.39% of the whole corpus (containing a total of 127.713 words/tokens).

### 1.4. Distribution of the 13 lexicalized items within the bottom-up detections

We also wanted to see whether the words obtained by the bottom-up approach contained the selected 13 lexicalized items. Their proportions were overall much lower in the sequences detected automatically; they were detected a total of 126 times with the bottom-up approach while the number of detections using the top-down approach was 2566, suggesting only 4.9% of the reduced lexicalized items were caught by the script. The bottom-up approach requires the presence of at least six reduced phonemes, whereas lexicalized reductions (LR from now on) typically affect fewer than six phonemes. As a result, for LR to be detected using the bottom-up approach, they must often be part of more extensive reduction sequences, which appears to be infrequent.

For instance, we had detected (with top-down approach) 579 instances where '*parce que*' was reduced. However, only 26 of these were encountered among the results of the bottom-up detection. While '*enfin*', was found to be the most frequently reduced lexicalized item in the corpus, it was detected only 3 times via the bottom-up approach. The exact number of appearances can be seen in Table 3.

Lexicalized Item	Detection		Lexicalized Item	Detection	
	Bottom-Up	Top-Down		Bottom-Up	Top-Down
<i>alors</i>	32	132	<i>tu sais</i>	4	176
<i>alors que</i>	0	7	<i>je sais</i>	3	328
<i>parce que</i>	26	579	<i>enfin</i>	3	586
<i>je suis</i>	17	128	<i>je peux</i>	3	28
<i>voilà</i>	14	108	<i>de toute façon</i>	2	53
<i>puis</i>	14	266	<i>puisque</i>	2	37
<i>quand même</i>	6	138	<b>TOTAL</b>	126	2566

Table 3. The number of appearances for the 13 selected lexicalized items in the sequences detected using top-down and bottom-up approaches.

## 2. Speaker Related Factors

In this part, we will report the reduction rates for each speaker of the corpus. Based on the manual and automatic annotations, we calculated the total numbers of reductions obtained through both the top-down method (lexicalized reductions, (LR)) and the bottom-up method (not only LR but also non-lexicalized reductions (NLR)). The distributions of the reductions by method of extraction and by speaker are reported in Figure 3.

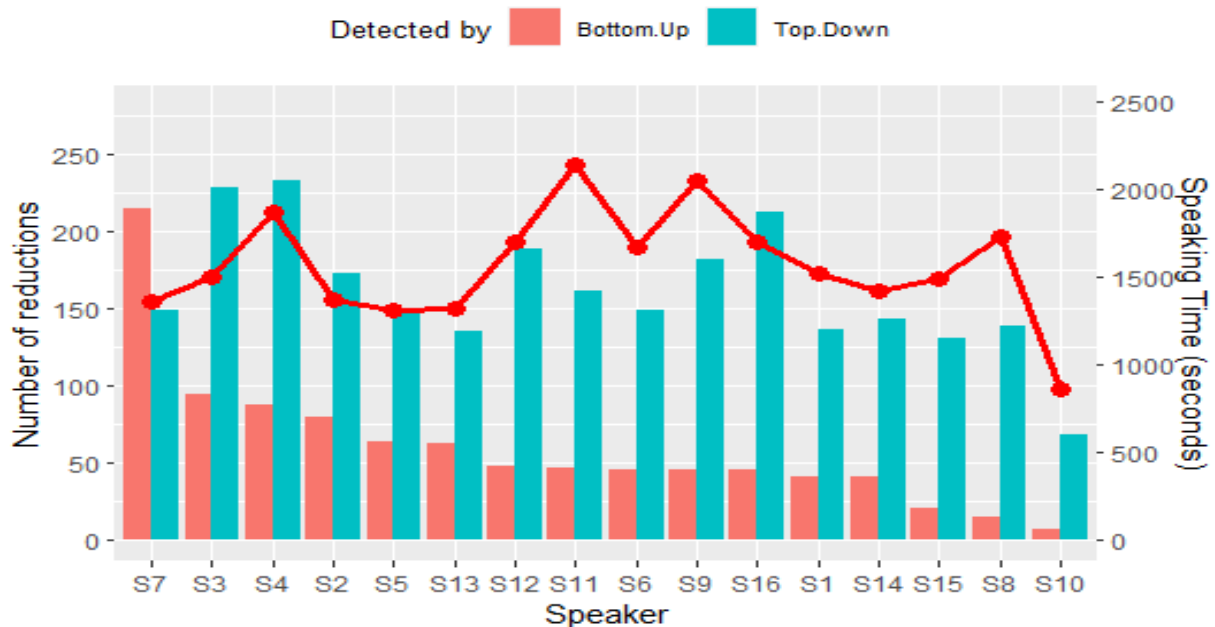


Figure 3. The numbers of reductions per speaker (in descending order based on the number of NLR), obtained using two complementary methods. LR are considered as the result of a top-down method while the NLR are the result of a bottom-up approach. The line represents the speaking time (in seconds) for each speaker.

As can be seen in Figure 3, the distribution of the reductions is highly heterogeneous among speakers. For example, the speaker with the most reductions (S7) produced 364 reductions (149 LR and 215 NLR), while the speaker with the lowest total number of reductions (S10) produced 75 sequences (68 LR and 7 NLR). The other speakers fall between these two extremes. However, the number of NLR of S7 is more than twice that of the following speaker (S3).

We observe that the minimum reduction rate for lexicalized items is 57% (S8 and S15), which suggests that speakers who reduce little do so on at least half of the items. It is important to note that the reduction ratio (number of reduced target items divided by the total occurrences of the same target item) for lexicalized items might provide more insights into a speaker's reduction patterns than the number of LR he/she produced (see Table 4). For instance, S7 achieved a reduction ratio of 81.4%, having reduced 149 out of a total of 183 target items, whereas the speaker with the lowest LR rate (S8) reduced 139 reduced out of 243 targets, resulting in a 57.2% reduction rate. Additionally, as speaking time varies among speakers in the corpus, a speaker can have a substantial reduction rate despite producing few target items (e.g., S10 with a 73.9% reduction rate). Conversely, S4 produced the most reduced items (233), but his reduction rate (74.4%) was not the highest, indicating that the frequency of use of an item does not necessarily correlate with the reduction rate.

	Male Speakers						Female Speakers									
	S3	S4	S7	S9	S15	S16	S1	S2	S5	S6	S8	S10	S11	S12	S13	S14
LRR (%)	80%	74%	81%	66%	57%	74%	80%	80%	73%	75%	57%	74%	64%	66%	62%	77%

Table 4. The ratio of LR for each speaker. The ratios are computed by dividing the number of reduced occurrences of the target items to their total number of occurrences in the corpus.

Considering that the number of phonemes produced by each speaker can be linked to their speaking times, investigating the proportion of reduced phonemes to the total number of phonemes produced by a speaker would also provide us with useful insight to better explore the various patterns among speakers. The proportion of the phonemes found in NLR sequences for each speaker are reported in Table 5.

	Male Speakers						Female Speakers									
Speaker	S3	S4	S7	S9	S15	S16	S1	S2	S5	S6	S8	S10	S11	S12	S13	S14
Phon. Red.	711	633	1000	293	151	331	302	578	427	312	100	50	359	359	454	291
Phon. Tot.	18.742	23.289	18.035	22.503	15.584	19.808	17.569	17.074	15.732	19.562	19.460	8237	25.626	21.754	16.972	17.756
%	3,7%	2,7%	5,5%	1,3%	0,9%	1,6%	1,7%	3,3%	2,7%	1,6%	0,5%	0,6%	1,4%	1,6%	2,6%	1,6%

Table 5. The number of the phonemes and their proportions for each speaker. Phon.Red. is the number of phonemes found in NLR. Phon.Tot. is the total number of phonemes produced by the speaker. % is the proportion of the reduced phonemes within the total.

Given this heterogeneity in reduction performances (from 0.5% to 5.5%), we explored four potential contributing factors: speaking time, articulation rate, gender, and the correlation between the numbers of the two types of reductions.

## 2.1. Speaking times

In Figure 3, we present the speaking times for each speaker, excluding pauses and laughs, to examine the distribution of speech duration among speakers and its potential relationship with the number of reductions they produce. Like the numbers of reduction, the speaking times were also highly variable from one speaker to another. Notably, the speaker with the highest number of NL reductions and the highest ratio of L reductions (S7) did not have the longest speech duration. Furthermore, the speaker with the shortest speaking time had the lowest number of both the bottom-up (NL) and the top-down (L) detections. The speaker with the longest speaking time (S11) did not seem to be performing very differently than the others.

To explore whether longer speaking times lead to more reductions, we conducted a Pearson's correlation test. The results revealed a statistically significant positive correlation ( $cor= 0.593211$ ,  $p-value= 0.01543$ ) between the speech duration and the number of the reductions detected with the top-down approach, indicating that longer speaking times tend to result in more LR. In contrast, the correlation coefficient between speech duration and the number of reductions detected with the bottom-up approach was  $-0.066503$ , suggesting a weak, statistically

insignificant negative correlation ( $p\text{-value}= 0.8067$ ). This implies that NLR are less influenced by the speaking times of the speakers. Additionally, we calculated the proportion of reduction durations within the total durations of speech for each speaker. The NLR ratios for each speaker are reported in Table 6.

	Male Speakers						Female Speakers									
Speaker	S3	S4	S7	S9	S15	S16	S1	S2	S5	S6	S8	S10	S11	S12	S13	S14
NLR ratio (in speaking time)	2%	1,2%	4,6%	0,6%	0,3%	0,7%	0,7%	1,4%	1,3%	0,7%	0,2%	0,2%	0,6%	0,8%	1,3%	0,7%

Table 6. The distribution of the speakers of the corpus according to their gender. NLR ratio is the proportion of the reduced sequence durations within the total speaking durations for each speaker.

## 2.2. Articulation rate

It was previously proposed in the literature (Fosler-Lussier & Morgan, 1999) that a higher articulation rate might induce more variation in speech given the established connection between phonemic duration and reduction. Articulation rate (number of phonemes per second) values in the CID Corpus of the speakers were extracted during a recent study (Chardenon, Meunier, & Fougeron, 2022). Interestingly, the speaker with the highest articulation rate (S7; 14.4 phonemes per second) also exhibited the highest reduction rate for both LR and NLR (149 LR and 215 NLR) while the speaker with the slowest rate (S10; 11.08 phonemes per second) produced the fewest reductions (68 LR and 7 NLR).

To quantify the relationship between the number of reductions produced by speakers and their articulation rates, a Pearson correlation coefficient was computed. The analysis revealed a positive correlation ( $cor=0.5252152$ ;  $p=0.4203$ ) between articulation rate and the proportion of reductions for lexicalized sequences, although it did not reach statistical significance. However, a statistically significant positive correlation was observed for non-lexicalized sequences ( $cor=0.786749$ ;  $p=0.0002996$ ), indicating a robust relationship between these two variables in this context.

## 2.3. Gender of the speaker

Previous research suggested a potential relationship between the gender of a speaker and their reduction rate. Specifically, it was suggested that male speakers tend to exhibit higher levels of reduction (Byrd, 1994; Bell, Jurafsky, Fosler-Lussier, Girand, & Gildea, 1999; Keune, Ernestus, van Hout, & Baayen, 2005). Nevertheless, it is noteworthy that some studies failed to find a statistically significant difference in reduction rates between male and female speech (e.g., Binnenpoorte & et al., 2005).

It should be acknowledged that the distribution of gender within the corpus is imbalanced, with 10 female and 6 male speakers. Additionally, the conversations exclusively involved either two female or two male speakers. More detailed information about the speakers and their reduction rates can be found in Tables 4, 5 and 6.

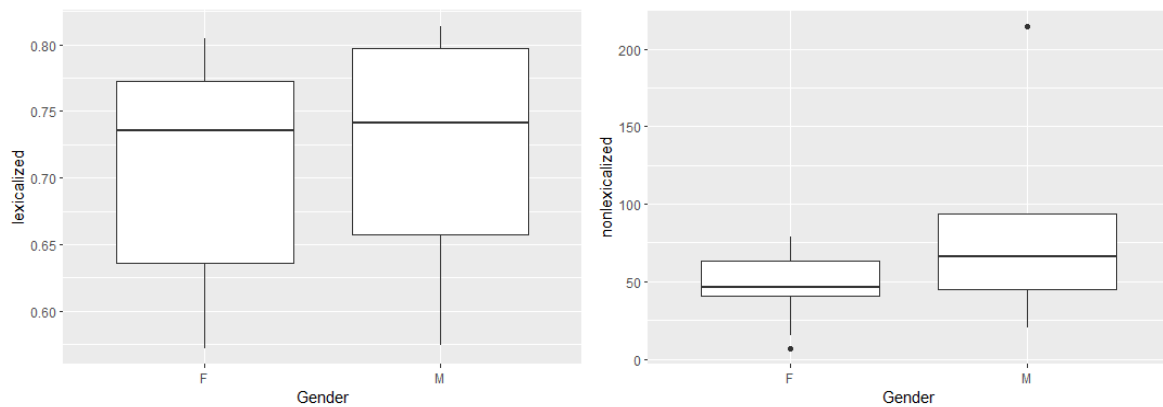


Figure 4. Mean values of the female and male speakers for the LR ratios (left) and the number of NLR they produced (right).

To see if the gender might be a contributing factor in the variation in reduction rates, we performed a t-test for both reduction types. The comparison of LR ratios between females and males revealed no statistically significant difference ( $t = -0.40423$ ,  $df = 21.555$ ,  $p\text{-value} = 0.69$ ). There was also no statistically significant difference in NLR scores between the two groups ( $t = -1.9814$ ,  $df = 12.339$ ,  $p = 0.07028$ ). However, it can be seen in Figure 4 that the difference between the means of the two groups for the NLR is more easily visible.

#### 2.4. The correlation between the number of lexicalized and non-lexicalized reductions

The distribution of lexicalized and non-lexicalized items varied among individual speakers in the corpus. We wanted to see whether speakers who had higher numbers of LR would produce more NLR and vice-versa. A Pearson's correlation test was performed to examine the correlation between the numbers of LR and NLR. The correlation coefficient between the lexicalized and non-lexicalized variables is 0.3069723, but this correlation is not statistically significant ( $t = 1.2069$ ,  $df = 14$ ,  $p = 0.2475$ ), suggesting there is insufficient evidence to conclude that there is a true correlation between the two variables.

Seeing that one of the speakers (S7) performed very differently than the others, we performed a correlation test excluding him. In this case, a strong positive correlation emerged ( $r = 0.74$ ,  $p = 0.001$ ), suggesting that a higher number of LR might be correlated with more NLR and vice-versa.

### 3. Linguistic Factors

In this section we will examine two linguistic factors, namely phonetic and lexical elements, and explore how their interplay influences the process of reduction in speech, shedding light on their respective contributions.

#### 3.1. Phonetic Factors

##### 3.1.1. Reductions detected with the bottom-up approach

For the reductions detected using the bottom-up approach, we first extracted all the phonemes encountered in the detected sequences for further analysis. As mentioned in the general observations, 6351 phonemes were identified in these reduced sequences.

First, we calculated how many phonemes on average are found in each of the sequences. When all the detected sequences in the corpus are considered, the average length of a detected sequence is 7.55 ( $SD = 2.27$ ) phonemes. The longest detected sequence encompassed 22 phonemes (/i p a r e g z a ~ p l E U d i r e k t E U r d E U r /: **si par exemple le directeur de recherche** (*if, for example, the research director*)). The speaker with the highest average length of reductions (S7) produced 8.1 phonemes (on average) per sequence, while the speaker with the least number of phonemes in a sequence averaged 6.5 phonemes. The majority of speakers (13 out of 16) maintained an average of more than 7 phonemes in their detected sequences.

Then, we looked at the distribution of the phonemes in these sequences. At a first glance, the most frequently appearing phonemes are /a/ (747 times), /e/ (701 times), /r/ (538 times) and /l/ (515 times) respectively. However, since these are also some of the most frequently appearing phonemes in French language, we calculated the ratio of reduced phonemes to their total occurrences in the corpus. The results show that these phonemes are not associated with higher reduction rates, despite their overall significant representation in reduced forms. Figure 5 shows that /H/ has the highest reduction rate, found reduced 6,7%, /l/ 2,6% while /e/ seems to be the most frequently appearing phoneme in the corpus but with a reduction rate of only 1.48%. The phonemes with the highest reduction ratios were mostly approximants: /H/, /j/, /l/, /w/ and /r/.

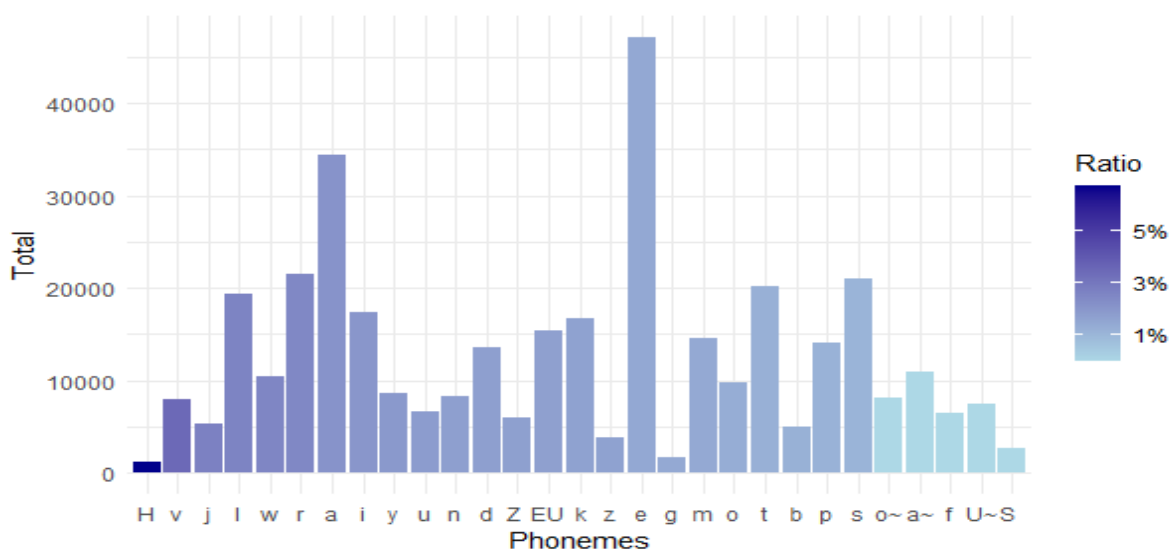


Figure 5. The total number in the corpus (y axis) and the ratios (color intensity) for the reduced phonemes found in the non-lexicalized sequences calculated based on their total occurrence in the corpus.

Additionally, even though the fricatives do not have high reduction ratios in general, /v/ was found to have a high ratio of reduction. These results suggest that the chances of a phoneme being reduced depend mostly on its articulatory characteristics and not on their frequency of occurrence.

### 3.1.2. Reductions detected with the top-down approach

Regarding phonetic factors, comparing the two types of reductions was not feasible since lexicalized items and the phonemes comprising these sequences were predefined. The 13 selected items contain various phonemes from different categories (refer to Table 3), and a single

lexicalized item could be reduced in multiple ways. For example, “*parce que*”, typically consisting of 6 phonemes, could appear as /sk/ with only two phonemes or /psk/ with three phonemes, while “*de toute façon*” which normally contains 10 phonemes, was often encountered as /tfsɔ̃/. The specific phonemes present in these reduced variants hinged on the phonetic and phonological properties of the item.

For instance, despite “*enfin*” being frequently reduced, the potential reduction forms were more constrained compared to others, primarily due to the presence of the fricative /f/, which was constantly retained in reduction. In addition, Figure 6 illustrates that the number of phonemes and words in the target item show no discernible relationship with the rate of reduction. This implies that the frequency of reduction for an item does not depend on the length of the sequence but rather on the inherent characteristics of the item itself.



Figure 6. Reduced (orange) and total (green) occurrences for each of the 13 selected items. The red line represents the ratio of reduction for each target item.

### 3.2. Lexical Factors

In this section, we will explore some of the lexical properties of the detected reduced sequences to investigate whether any relationship exists between lexical factors and the likelihood of reduction.

#### 3.2.1. Reductions detected with the bottom-up approach

For the bottom-up analyses, we created a custom R script to extract and analyze the words comprising the reduced sequences. This script utilizes two Textgrid files as input: one containing the reduced sequences and the other containing orthographic tokens. By aligning the start and end times of the identified reduced sequences, the script extracts all words that fall within these



intervals from the token Textgrid. From this process, we obtained a list of 716 tokens spanning different morphosyntactic categories. These tokens occurred a total of 3058 times within the detected reduced sequences. Importantly, this number accounts for 2.39% of the entire corpus, which encompasses a total of 127.713 words/tokens. It is worth noting that our method identifies a limited number of reductions based on the parameters we configured (requiring a minimum of 6 phonemes and 230ms).

The words identified within the reduced sequences show great variability. In a general context, pronouns and verbs emerged as some of the most frequent candidates. However, when we calculated the proportions of these reduced words in relation to all tokens utilized in the corpus, they exhibited lower reduction rates. For instance, "je" (5.7%), "tu" (3.4%), "de" (3.7%), "il" (4.1%), "la" (4.9%), and "que" (5.1%) all showed comparatively modest reduction rates. Again, we must specify that these values, extracted according to the constraints of our method, only represent a part of the reductions present in the corpus.

In contrast, a total of 68 words from different categories (including 30 nouns, 25 verbs, 4 adjectives and 9 disfluencies) exhibited a 100% reduction rate, even though some of them appeared only once in the corpus. Examples of such words include "fourrure (*fur*)", "orthophoniste (*speech therapist*)", "signalisations (*signalings*)", and "SNCF (*French Railways*)", among others. It is worth noting that certain lexicalized items detected as reduced did not appear within sequences identified using the bottom-up method. This divergence is attributed to the presence of missing phonemes in these items, albeit the remaining phonemes did not fall below the predetermined duration threshold.

Number of words	Occurrence	Frequency
10 words	1	0.1%
8 words	4	0.4%
7 words	12	1.2%
6 words	30	3.1%
5 words	79	8.2%
4 words	205	21.4%
3 words	359	37.6%
2 words	196	20.5%
1 word	62	6.4%

Table 7. The number of words in reduced sequences (left), the number (middle) and the frequency of occurrences in all of the detected reduced sequences (right).

We also conducted an analysis to determine the number of words present in these detected reduced sequences. Our findings revealed that a reduced sequence could consist of anywhere from 1 to 10 words (see Table 7). Typically, these sequences comprise 2 to 4 words. However, it was less common to encounter sequences containing more than 6 words, accounting for less than 2% of the detected sequences.

#### *Morphosyntactic categories of the words found in the detected sequences*

We looked at the morphosyntactic categories of words identified within sequences detected using the bottom-up approach. Reduction processes affected words across all morphosyntactic

categories, although certain categories were more prevalent in reduced sequences than others. Their distributions were as follows: 27% verbs, 25.7% pronouns, 11% nouns, 9% determiners, 7% prepositions, 6% adverbs, 5% discourse markers, 4% conjunctions, 3% adjectives, 1% disfluencies and 0.3% interjections. This breakdown illustrates the varying degrees to which different morphosyntactic categories were subject to the reduction phenomenon.

Subsequently, we computed the proportions of reduced word categories in relation to their overall occurrences, i.e., the number of reduced adjectives divided by the total number of adjectives in the corpus, and so on. The results are detailed in Table 8. Once again, verbs (4.3% of the total number of verbs in the corpus) and pronouns (3.2% of the total number of pronouns in the corpus) emerged as categories most frequently subject to reduction. However, this calculation revealed that adverbs (3.9% of their total) and conjunctions (3.4%) were also prevalent morphosyntactic categories in comparison to the others, exhibiting high reduction rates relative to their total occurrences.

Category	Reduced Number	Total in the Corpus	Ratio
Verb	834	19058	4.3%
Pronoun	786	24141	3.2%
Noun	323	12260	2.6%
Determiner	279	10334	2.7%
Preposition	229	7405	3%
Adverb	186	4806	3.9%
Discourse Marker	161	22168	0.7%
Conjunctions	135	4005	3.4%
Adjectives	93	3557	2.6%
Disfluency	31	10583	0.3%
Interjection	1	103	0.97

Table 8. The morphosyntactic categories of the words found in reduced sequences (bottom-up) and their proportions in the total of the corpus.

### 3.2.2. Reductions detected with the top-down approach

The items selected for the top-down collection are already well-documented in the literature as reduced. However, our findings reveal that while these items are indeed often subject to reduction (71% of the target sequences), there are notable variations in their reduction frequencies, indicating that the likelihood of reduction is influenced by the specific sequence itself. Figure 6 illustrates the distribution of different items and their reduction ratios (the number of reductions divided by the total number of items).

Within the CID corpus, the most commonly reduced lexicalized form is “*parce que*”, reduced in 98% of cases. Interestingly, “*enfin*” is the most frequently occurring among the 13 selected items in the corpus, yet it exhibits a lower reduction ratio compared to less frequently occurring words. This suggests that an item can be rare in the corpus yet display a high proportion of reduction. For example, “*puisque*” appears only 45 times in total within the corpus but is reduced 82% of the time. Conversely, “*puis*” is relatively frequent but quite rarely reduced. This variation in reduction ratios appears to be linked to the inherent characteristics of the items themselves.

### *The effect of morphosyntactic categories on the reduction of lexicalized items*

Regarding the lexicalized sequences, we wanted to see whether they are more susceptible to reduction when occupying specific morphosyntactic roles. For that purpose, we selected a subset of 3 lexicalized items (*alors*, *enfin*, *tu sais*) that can serve various functions within the syntactic structure. For example, “*enfin*” and “*alors*” can function both as adverbs and discourse markers while “*tu sais*” is typically a pronoun+verb composite but can also be used as a discourse marker in French. Building on the general assumption that less informative parts of speech are more frequently reduced, we hypothesized that these items would be more frequently reduced when functioning as discourse markers rather than as adverbs or pronoun+verbs.

A logistic regression analysis was conducted to investigate the relationship between the likelihood of reduction (reduced vs not reduced) and the morphosyntactic category for each item. In the case of “*alors*”, which included 248 data points, none of the categories showed a statistically significant effect on reduction. The coefficients for all categories (adverb, discourse marker, feedback, and disfluency) were not statistically significant, nor were the intercepts. However, looking at “*alors que*”, which had 48 data points, we observed that it was rarely reduced (15%) and predominantly served as a conjunction (it occurred as a disfluency only 2 times). The disfluency variable did significantly impact reduction (*coefficient: -15.848, p = 0.995*), while the intercept did (*coefficient: -1.717, p < 0.001*).

Based on 731 data points, “*enfin*” revealed no significant associations between any of the morphosyntactic categories and reduction. However, the intercept was found to have a significant effect on reduction (*coefficient: -1.557, p = 0.991*). Lastly, for “*tu sais*”, which had 247 data points, the category “pronoun+verb” was linked to a significantly lower likelihood of reduction compared to the discourse marker (*coefficient: -1.1722, p < 0.001*), and the intercept also exhibited a significant effect (*coefficient: 1.1722, p < 0.001*).

## DISCUSSION

In the current study, we conducted a preliminary analysis on a corpus of French conversations with the aim of gaining a deeper understanding of the phenomenon of reduction in spontaneous and casual speech. While some reductions are easy to identify and thus well-documented, others are not explicitly perceptible to the human ear, thus complicating the detection and analysis. Having established a distinction between these two types of reductions, we classified them as lexicalized (LR) and non-lexicalized (NLR) reductions and used two complementary methods to extract reductions. LR were collected using a top-down approach, focusing on 13 frequently occurring forms in everyday French. For NLR, we proposed a novel bottom-up approach allowing us to discover other types of reductions -dependent on the window size we defined-, focusing not only on the lexicalized ones but also on the ones involving multiple words that are likely to be influenced by various linguistic and conversational factors (e.g., phonology, prosody, pragmatics). The detection of these multi-word reductions is essential for understanding which factors are involved in the NLR process. The number of detected reduced sequences varied based on the chosen window size, with a deliberate selection of a 230ms window to capture as many multi-word sequences as possible. In exchange, we were able to detect only a limited number of reductions. Using a smaller window size and fewer phonemes (e.g., 30ms for 2 phonemes) would have yielded a larger number of reductions. However, our aim here was to identify large areas of

reduction affecting several words. Manual evaluations of the script's output highlighted the critical role of input quality (alignment) in accurate reduction detection. A wrong alignment could lead the script to identify a sequence as reduced, even if it contained fewer phonemes. This is why we observed a considerable improvement in output quality when using the v3 of phonetic annotations, despite resulting in a smaller sample size. One advantage of this bottom-up method is its ability to uncover various reduced forms that may not be apparent through manual listening and annotation. Manual detections and annotations are highly subjective, and different annotators may provide different judgments. Therefore, an objective and automatic method greatly simplifies the task. To lay a robust foundation for our decision to distinguish between two types of reduction, we examined the distribution, frequency, and characteristics of reduced sequences of both types, along with their correlation with potential factors of variation.

The analysis of the collected data has allowed us to observe a very heterogeneous distribution of reductions in conversations. While the top-down method and subsequent analysis revealed a substantial 71% of the selected lexicalized items exhibited reduction by speakers, suggesting a massive phenomenon of reduction, our exploration highlighted a nuanced and often overlooked speaker-related variability in reduction. Previous research has occasionally touched upon this aspect, mainly within the context of sociological factors (Labov, 1986; 2001). However, we observed that it contributes significantly to the heterogeneity in reduction rates (e.g., Figure 3 and Tables 4 and 5); speakers behave differently in terms of the number of reduced productions both for the lexicalized and non-lexicalized sequences. Notably, one speaker (S7) emerged as a standout, producing 215 NLR and had a reduction ratio of 81,4% for the lexicalized sequences whereas the others were below this production rate. Conversely, the speaker with the lowest LR rate (S8) produced 139 reductions out of 243 targets (57.2%) and also had the second lowest number of detected NLR, with only 15 sequences. Adding complexity to the results, the speaking time among speakers in the corpus was highly variable. A speaker could produce relatively few target items (e.g., S10) while still maintaining a considerable LR rate (73.9%), even with the lowest number (7) of NLR. These subtleties highlight how individual speaker traits, language preferences, and phonetic factors are interconnected, encouraging us to further investigate the reasons behind the differences.

It is crucial to remember that these two collection methods and reduction types can be considered as complementary to one another. The top-down method allows only a limited number of - previously known- sequences to be collected while the bottom-up method is able to go further and detect larger and more various sequences. However, this method is duration based (230ms/6 phonemes), meaning that even if the lexicalized items were sometimes found in the detected reduced sequences, it did not detect all of the manually identified lexicalized sequences. Only 4.9% of the sequences detected with the top-down approach were present in the bottom-up detections. This can be explained by the fact that a LR such as /*ʒepa*/ (which is one of the reduced representations of *je ne sais pas*) might be missing some phonemes but it is not necessarily pronounced with a shorter duration all the time. We could consider these reduced forms as exemplars of the same lexical form. The fact that they are considered reduced stems mostly from the missing phonemes or syllables and not from the phonemic durations. Bybee (2002) suggested that the phonetic form of each token of experiences is recorded in the exemplar representation, providing a range of variation for each word or phrase. High frequency words and phrases are more susceptible to variation, thus more reduction, due to the greater automatization of highly practiced sequences and due to the greater predictability of these items (Jurafsky, Bell, Gregory, & Raymond, 2001). As the selected items are frequent in everyday language, it is possible that

their various reduced forms are stored in the exemplar as well. Indeed, the correlation between the number of LR and the articulation rates was found to be weak ( $cor=0.5252152$ ;  $p=0.4203$ ). This finding is compatible with the observations of Van Son & Pol (1990; 1992), who noted that high speech rates do not necessarily lead to speech reduction: some speakers are able to control their articulators in such a way as to produce non-reduced pronunciations at high rates (e.g., S14). Thus, it can be suggested that these types of reductions are not highly dependent on the speaker's articulation rate. On the other hand, the correlation between the number of NLR and the articulation rates of the speakers was shown to be statistically more significant ( $cor=0.786749$ ;  $p=0.0002996$ ), confirming Fosler-Lussier & Morgan's (1999) finding that, at least in the case of NLR, a higher articulation rate might induce more variation and reduction; contributing to the heterogeneity in the distribution of reductions among speakers. This difference in correlation could be explained by the fact that LR are produced as exemplars of some stereotyped lexical items, and they are not necessarily as influenced by the characteristics of the speaker and the phonologic/prosodic context of speech.

The reductions detected with two different methods seem to be correlated with different factors, confirming our assumption that they should be considered as two different types of reductions. In addition to the fact that the NLR are correlated with higher articulation rates, we looked at the speech durations of the speakers, hypothesizing that longer speech durations would result in more LR but not necessarily more NL ones. The analysis showed that the LR had a significant positive correlation with the speaking time ( $cor= 0.593211$ ,  $p-value= 0.01543$ ) while it was not the case for the NLR. This can be explained by the fact that extended speech durations would provide speakers with more chances to produce some lexical items, thus reductions, while the NLR are not influenced by the same factors.

Another speaker dependent factor we investigated to see whether it contributes to the heterogeneity in the distribution of reductions was the gender of the speaker. When we took the gender of the speakers into account, we saw that the previous hypothesis that males tend to make more reductions was not confirmed. There was no significant difference between the two genders for the production of LR and NLR. However, the males had a higher average of NLR. Thinking that this effect might be stemming from the S7, who had an exceptional reduction performance compared to the other speakers, we decided to perform another t-test for the number of reductions and gender by excluding him. The results revealed that the difference between the two groups is still not statistically significant without this speaker ( $t = -1.9814$ ,  $df = 12.339$ ,  $p-value = 0.07028$  for NLR and  $t = -0.40423$ ,  $df = 21.555$ ,  $p-value = 0.69$  for LR). These analyses suggest that gender is not a crucial factor for the varying reduction rates. Other factors should also be taken into consideration to account for the difference in the rates of reduction. In a recent paper, Clopper et al. (2023) discussed the influence of the dialect of the speakers on the reduction in English (American Midwest) and concluded that phonetic reduction processes interact with dialect variation only weakly. In our corpus, the majority of the speakers come from South-Eastern France, and the number of speakers may not be significant for the exploration of the influence of dialect on reduction. However, it should be tested to see whether we observe similar effects in other dialects of French as well.

We also looked at the correlation between the two types of reductions as another potential predictor of the production rate. The fact that the speaker with the highest reduction rate for lexicalized sequences (S7) also produced the higher number of non-lexicalized sequences and that the one with the lowest rate of LR (S8) produced very few NLR, might suggest a correlation

between the two types of reductions. Overall, the correlation was not statistically significant, and the rate of LR cannot be considered as a predictor of the rate of NLR. It should be noted that these results depend on the specific context and population of speakers we studied. Further research with larger sample sizes may be warranted to explore this relationship more comprehensively.

Additionally, we observed that the two reduction types have different mechanisms, and they are not influenced by the same factors. For example, as the name suggests, the LR are highly dependent on the lexicon. Their distributions in conversations are quite heterogeneous, and the frequency with which an item is reduced depends on the item itself and not on the frequency of the item. Some are reduced more often in the corpus (e.g., *parce que*, 98%; *je suis*, 88%), while others are reduced less often (e.g., *alors*, 33%; *voilà*, 39%). These rates are compatible with those of Wu & Adda-Decker (2020), who suggested that these reduced sequences often fulfill a role as discourse markers that speakers use in casual speech. Furthermore, reduction ratios for the lexicalized sequences are not dependent on their frequency in the corpus. For example, the sequence "*je peux*" (I can) has a reduction ratio of 84.8%, even though it appeared only 33 times in total in the corpus, whereas "*alors*" (then) appeared 400 times in total but was reduced only 132 times (33%).

Interestingly, when we look at the selected lexicalized items, we see that the items with the highest reduction ratios (*parce que* (98.4%, conjunction), *je suis* (88%, pronoun+verb), *je peux* (84.8%, pronoun+noun)) might not always be considered as highly informative as they are easily predictable by the context (Bell, Brenier, Gregory, & Jurafsky, 2009). The "pronoun+verb" combinations such as "*je sais*", "*je suis*" and "*je peux*" have a higher reduction ratio compared to adverbs such as "*alors*", "*voilà*", etc., and it is often the pronoun that is reduced in these sequences, which is highly predictable. To see whether there is a relationship between the informativeness and the morphosyntactic category of a lexicalized item and its possibility of reduction, we analyzed a subset of 3 sequences (*alors*, *enfin*, *tu sais*) which can function as either discourse markers or adverb/pronoun+verb. Our findings suggested that, within this subset, there is no evidence to support the hypothesis that specific morphosyntactic categories are associated with a higher likelihood of reduction and any of the specific morphosyntactic categories these forms would occupy are associated with a higher likelihood of reduction except for "*tu sais*". The category of pronoun+verb is associated with a significantly lower likelihood of reduction compared to the discourse marker category, showing that reduction rates of some items might indeed vary due to their informativeness and category, but they are more influenced by the lexicon. This finding is in line with the proposition that the more pragmatic and the less lexical or compositional an item is, the more it will be reduced (Schubotz et al., 2015) and that the presence of reduction would be helpful for the listener to grasp the intended meaning. A comparison between the durations of the reduced and non-reduced variants of these items would help us draw more conclusions to complement these findings in the future.

On the other hand, NLR are not as dependent on the lexicon, and they may contain various words from various categories. Many of the words with the highest reduction rates were the ones appearing rarely in the corpus (e.g., *orthophoniste*, *bouquinais*) while the most frequently encountered words in reduced sequences had lower proportions due to their overall higher presence in the corpus. In this case, previous proposition (Pluymaekers, Ernestus, & Baayen, 2005) & (Jurafsky, Bell, Gregory, & Raymond, 2001) that frequent words get reduced more frequently is only due to the fact that these words appear more frequently in the corpus, but it does not necessarily mean that they have a higher tendency to become reduced and the words

involved in reduction show great variability. Additionally, Jurafksy et al.'s (2001) proposition that the less informative parts of the speech are reduced more frequently seems to be contradicting our findings as the majority of the words with a 100% reduction rate were content words which are informative in their context of occurrence. However, it should be acknowledged that many of these words appeared only once in the corpus, limiting our ability to draw more reliable interpretations.

Even though verbs (27% of all the detected words) and pronouns (25.7%) had higher numbers of representation (see Table 8), the main reduction ratios (times reduced/total number of that same category in the corpus) of adverbs (3.9% of all the adverbs in the corpus) and conjunctions (3.4%) were as important. Only 0.3% of the disfluent words appeared in the reduced sequences, which is an interesting finding in that these disfluent words can be considered as less informative and as mostly repetitions, which is a condition thought to be contributing to the frequency of reduction (Fowler, 1988). However, one might still wonder if reductions appear more in sentences containing disfluencies. In a previous study (Bodur et al., 2023) on the relationship between reduction and the prosodic position, we found that around 37% of the reductions occur within disfluent units. These results suggest that further analysis is needed to understand whether there is a relationship between the disfluencies and whether there are more reduction prone categories in general or not. Additionally, the length of the detected reduced sequences varied in terms of the numbers of words involved. More commonly, a detected reduced sequence was made up of 3 to 4 words. This effect is mostly due to the detection method and parameters we selected. As we wanted to observe larger areas of reductions, we used a 230ms window with 6 phonemes at least, assuming a duration of around 38.33ms for each phoneme. A smaller window size would yield more reductions involving –possibly- a smaller number of words. It was suggested in some of the previous studies (e.g., Moon & Lindblom, 1994; Baltazani, 2007), that longer sequences have a higher tendency to get reduced both temporally and spatially. The fact that extra short sequences (e.g., consisting of only one word) have a much lower reduction rate in our corpus might be considered as an effect of this factor. However, extra long sequences (e.g., 8 to 10 words) were also quite rarely influenced by the reduction process. A possible explanation for this could be as following. For the case of shorter sequences, it is plausible that speakers, perceiving the brevity of such utterances, opt to maintain a higher degree of linguistic integrity, as further reduction may risk compromising the clarity of their intended message. Conversely, in the context of longer sequences, the observed rarity of reduction can be attributed, in part, to the inherent complexity and informational richness associated with such utterances, which may prompt speakers to exercise caution in employing reduction strategies to ensure that essential details are not lost in the process.

Additionally, as the articulatory characteristics of phonemes also influence the reduction rate (Meunier & Bigi, 2021), some phonemes might appear more often in reduced sequences than the others. Previous studies (e.g., Pluymaekers, Ernestus, & Baayen, 2005; Voigt & Schüppert, 2013) suggested that a higher frequency would cause a higher rate of reduction. Calculating the proportions of reduced phonemes in the corpus, we saw that the phonemes with the highest reduction proportions are /H/, /j/, /l/, /w/ and /v/ respectively, meaning that the category of approximants (liquids and glides) appears more frequently in the detected reduced sequences in our corpus while fricatives, plosives and nasals are rarely present in these sequences. It is also crucial to note that in terms of duration, these are also some of the shortest phonemes in French. This suggests that the properties and characteristics of phonemes are potentially influential factors for reduction (Meunier & Bigi, 2021). In addition, reduced sequences detected with the

bottom-up approach were found to consist of up to 22 phonemes while the shortest sequences consisted only of 6 phonemes, which is the minimum number of phonemes we set for the detection.

For the selected lexicalized sequences, the phonetic compositions are also heterogeneous. While the number of phonemes in the target items, and the number of words in it, show no relationship with the rate of reduction, the forms an item can take when reduced depended on the nature and of the phonemes it contains. For example, when “*alors*” is reduced, it is often the /l/ that is deleted, and the vowels /a/ and /o/ were mixed together, turning the sequence into a long vocalization. “*Enfin*” is very often pronounced as “*fin*”, and omission of the fricative consonant has never been observed. On the other hand, for “*parce que*”, both the /r/ and the schwa are systematically omitted, even in the least reduced forms. These observations suggest that the proportion of reduced material may be linked, in part, to the nature of the phonemes making up the target item. We might add, although we haven't carried out any specific analysis on this subject, that the item “*parce que*” is the most massively reduced item in the corpus, even though it contains two schwas. As the majority of speakers come from south-eastern France, we could have assumed that this item would be less reduced (as is the tendency in the southern parts of France to preserve schwas), but this was not the case.

In conclusion, our examination of preliminary results obtained through two distinct data collection methods has revealed a remarkable variability in the distribution of reductions across speakers, phonemes, and words, encompassing both lexicalized and non-lexicalized sequences. We believe that this variability underscores the methodological soundness of establishing two reduction types within our study. While reduction has traditionally been regarded as a primarily physiologically driven, emerging as a result of an effort to minimize articulatory effort and often occurring in the absence of any constraints to hyper-articulate (Lindblom, 1990), our findings suggest a more nuanced perspective. It can be proposed that certain reduced productions are not just ephemeral but rather stored as exemplars of the initial forms within the lexicon. Notably, the frequency of LR production remains relatively unaffected by articulation rates, persists across all speakers, and does not exhibit a strong physiological underpinning, indicating that these items may be retrieved and pronounced at varying rates without significant dependence on contextual factors. On the other hand, the NLR seems to be more context sensitive, manifesting in non-predictable words, spanning multiple words, and contingent upon articulation rate and morphosyntactic category. Additionally, our new bottom-up detection method holds promise for the study of reductions in spontaneous speech, despite potential biases stemming from predefined parameters. Future research may explore ways to further refine this approach, such as individualizing scripts for each speaker by incorporating their articulation rates as a parameter. This advancement could potentially enhance the precision of reduction detection and reduce the need for extensive manual evaluations, thus expanding possibilities for more comprehensive investigations into the dynamics of reductions in natural language production. In light of these findings, it becomes evident that our study not only contributes to the evolving understanding of reductions but also highlights the potential for transformative advancements in the field of spontaneous speech analysis.



## References

- BIBLIOGRAPHY Adda-Decker, M., & Snoeren, N. (2011). Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics*, vol. 39, 261-270.
- Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 3048-3058.
- Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and speech*, 391-413.
- Baltazani, M. (2007). Prosodic rhythm and the status of vowel reduction in Greek. *Selected papers on theoretical and applied linguistics*, 17(1), 31-43.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gildea, D. (1999). Forms of English function words - Effects of disfluencies, turn position, age and sex, and predictability. *Proceedings of ICPHS*, (pp. 395-398).
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., & Rauzy, S. (2006). Le CID- Corpus of Interactional Data: protocoles, conventions, annotations. *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA)*, 25, 25-55.
- Bertrand, R. (2008). Le CID- Corpus of Interactional Data- annotation et exploitation multimodale de parole conversationnelle. *TAL*, 105.
- Bigi, B., & Hirst, D. (2012). SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody. In *Speech Prosody* (pp. 19-22).
- Binnenpoorte, D., & et al. (2005). Gender in everyday speech and language: a corpus-based study. *Proceedings of Interspeech*, (pp. 2213-2216).
- Bodur, K., Fredouille, C., & Meunier, C. (2022). Formes réduites en conversation: caractéristiques des sequences et des locuteurs. *JEP - 34èmes Journées d'étude sur la Parole*. Noirmoutier.
- Bodur, K., Bertrand, R., German, J., Rauzy, S., Fredouille, C., & Meunier, C. (2023, August). Speech reduction: position within French prosodic structure. In *Interspeech 2023* (pp. 117-121), Dublin.
- Boersma, P., & Weenink, D. (2022). Praat: A system for doing phonetics by computer, 2000. *Software available at www.praat.org*, 4(2).
- Brun, A., Cerisara, C., Fohr, D., Illina, I., Langlois, D., Mella, O., & Smaïli, K. (2004). Ants: le système de transcription automatique du Loria. In *JEP 2004-25èmes Journées d'Etude sur la Parole* (p. 4).
- Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in second language acquisition*, 24(2), 215-221.
- Bybee, J., File-Muriel, R. J., & de Souza, R. N. (2016). Special reduction: A usage-based approach. *Language and Cognition*, 421-446.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 39-54.
- Chardenon, E., Meunier, C., & Fougeron, C. (2022). Variations de débit articulatoire sur un corpus conversationnel avec différents types d'interactions. *JEP -34ème Journées d'Etude sur la Parole*, (p. 4). Noirmoutier.
- Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America*, 1682-1688.

- Clopper, C., Turnbull, R., Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., & Zellers, M. (2018). Exploring variation in phonetic reduction: Linguistic, social and cognitive factors. *Rethinking Reduction*, 25-72.
- Di Cristo, A., & Di Cristo, P. (2001). Syntax, une approche métrique-autosegmentale de la prosodie. *Traitement Automatique des Langues*, 42(1), 69-111.
- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, vol. 27, 143-165.
- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, 27-41.
- Ernestus, M. T. (2000). *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*. Utrecht: LOT.
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 253-260.
- Fosler-Lussier, E., & Morgan, N. (1999). Effects of speaking rate and word frequency on conversational pronunciations. *Speech Communication*, 137-158.
- Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and speech*, 31(4), 307-319.
- Gendrot, C., Adda-Decker, M., & Schmid, C. (2012). Comparaison de parole journalistique et de parole spontanée: analyses de séquences entre pauses. *Proceedings of the Joint Conference JEP-TALN-RECITAL*, (pp. 649-656).
- Guardiola, M., & Bertrand, R. (2013). Interactional convergence in conversational storytelling: when reported speech is a cue of alignment and/or affiliation. *Frontiers in psychology*, 4.
- Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, (pp. 29-54).
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language*, 45, 229-254.
- Keune, M., Ernestus, M., Van Hout, R., & Baayen, R. H. (2005). Social, geographical, and register variation in Dutch: From written mogelijk to spoken mok. *Corpus Linguistics and Linguistic Theory*, 183-223.
- Labov, W. (1986). The social origins of sound change. In *Dialect and language variation* (pp. 524-541). Academic Press.
- Labov, W. (2001). *Principles of linguistic change: social factors*. Blackwell Publishers, Oxford.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech Production and Speech Modelling*, 403-439.
- Meunier, C., & Bigi, B. (2016). Répartition des phonèmes réduits en parole conversationnelle: Approche quantitative par extraction automatique. *Proceedings of the Joint Conference JEP-TALN-RECITAL*, (pp. 615-623).
- Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, 271-278.
- Meunier, C., & Bigi, B. (2021). Variations temporelles des phonèmes en parole conversationnelle: propriétés phonétiques et facteurs lexicaux. *Studii de lingvistică*, 11, 11-38.
- Moon, S. J., & 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.

- Niebuhr, O., & Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, 319-329.
- Pardo, J. (2013). Measuring phonetic convergence in speech production. *Frontiers in psychology*, 4.
- Pharao, N. (2010). *Consonant reduction in copenhagen danish*. Copenhagen: Unpublished Ph.D. dissertation, University of Copenhagen.
- Prévoit, L., Bertrand, R., Blache, P., Meunier, C., Nguyen, N., & Pallaud, B. (2022). Étudier la conversation pour mieux comprendre le langage. TIPA. Travaux interdisciplinaires sur la parole et le langage, (38).
- Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken dutch. *The Journal of the Acoustical Society of America*, 2561-2569.
- Schubotz, L. M. R., Oostdijk, N. H. J., & Ernestus, M. T. C. (2015). Y'know vs. you know: What phonetic reduction can tell us about pragmatic function.
- Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmented transcriptions. *Journal of Phonetics*, 96-109.
- Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies*. (Doctoral dissertation , University of California, Berkeley).
- Turnbull, R. (2017). The role of predictability in intonational variability. *Language and speech*, 123-153.
- Van Son, R.J.J.H. & Pols, L.C.W., (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 88, 1683-1693.
- Van Son, R.J.J.H., Pols, L.C.W., (1992). Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, 92, 121-127.
- Voigt, S., & Schüppert, A. (2013). Articulation rate and syllable reduction in Spanish and Portuguese. *Phonetics in Europe: Perception and production*, 317-332.
- Wells, J.C. (1997). 'SAMPA computer readable phonetic alphabet'. In Gibbon, D., Moore, R. and Winski, R. (eds.), 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B.
- Wu, Y., Adda-Decker, M., & Lamel, L. (2020). Schwa deletion in word-initial syllables of polysyllabic words: Investigations using large French speech corpora. *Journal of Monolingual & Bilingual Speech (JMBS)*, 2(2).