amU

Aix　　Marseille　　Université

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 29 octobre 2025 par

Kübra BODUR

Reduction phonétique en conversation: compétences phonétique et discursive

Dicipline
Sciences du langage

École doctorale
ED 356 - Ecole Doctorale Cognition, Langage et Education

Laboratoire/Partenaires de recherche
LPL - Laboratoire Parole et Langage

Composition du jury

Richard OGDEN	Rapporteur
Professeur des universités, University of York	
Christelle DODANE	Rapporteure
Professeur des universités, Université Sorbonne Nouvelle	
Barbara SCHUPPLER	Examinatrice
Directrice de recherche, Technical University of Graz	
Cédric GENDROT	Examinateur
Professeur des universités, Université Sorbonne Nouvelle	
Noël NGUYEN	Président
Professeure des universités, Aix-Marseille Université	
Christine MEUNIER	Directrice de thèse
Directrice de recherche, Aix Marseille Université	
Corinne FREDOUILLE	Co-directrice de thèse
Professeur des universités, Avignon Université	

Affidavit

I, undersigned, Kübra Bodur, hereby declare that the work presented in this manuscript is my own work, carried out under the scientific supervision of Christine Meunier and Corinne Fredouille, in accordance with the principles of honesty, integrity and responsibility inherent to the research mission. The research work and the writing of this manuscript have been carried out in compliance with both the French national charter for Research Integrity and the Aix-Marseille University charter on the fight against plagiarism.

This work has not been submitted previously either in this country or in another country in the same or in a similar version to any other examination body.

Marseille, 10 juillet 2025

List of publications and participations in conferences

1) List of publications produced as part of the doctoral project:

   1. Bodur, K., Meunier, C., & Fredouille, C. (2022). Formes réduites en conversation : Caractéristiques des séquences et des locuteurs. *Proceedings of the 34th Journées d'Etudes sur la Parole* (JEP 2022), Noirmoutier, France, June 13-17.

   2. Bodur, K., Bertrand, R., German, J., Rauzy, S., Fredouille, C., & Meunier, C. (2023). Speech reduction: Position within French prosodic structure. *Proceedings of Interspeech 2023*, Dublin, Ireland, August 20-24.

   3. Bodur, K., Fredouille, C., Rauzy, S., & Meunier, C. (2025). Exploring the nuances of reduction in conversational speech: Lexicalized and non-lexicalized reductions. *Speech Communication*.https://doi.org/10.1016/j.specom.2025.103268

   4. Bodur, K., Fredouille, C., & Meunier, C. (2025). Speech reduction in French: The relationship between vowel space and articulation dynamics. *Proceedings of Interspeech 2025*, Rotterdam, Netherlands, August 17-21.

   5. Bodur, K., Fourtassi, A., Fredouille, C., & Meunier, C. (2025). Speech reduction in middle childhood: The case of child–caregiver interactions. *First Language*, *0*(0). https://doi.org/10.1177/01427237251363936.

2) List of other publications produced during the doctoral project:

1. Bodur, K., Branje, S., Peirolo, M., Tiscareno, I., & German, J. S. (2021). Domain-initial strengthening in Turkish: Acoustic cues to prosodic hierarchy in stop consonants. *Proceedings of Interspeech 2021*, Brno, Czechia, August 30-September 3, 1459-1463.

2. Bodur, K., Nikolaus, M., Kassim, F., Prévot, L., & Fourtassi, A. (2021). Chico: A multimodal corpus for the study of child conversation. *Companion Proceedings of the 2021 International Conference on Multimodal Interaction*, Montreal, Canada, 158-163.

3. Bodur, K., Nikolaus, M., Fourtassi, A., & Prévot, L. (2022). Backchannel behavior in child-caregiver Zoom-mediated conversations. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Toronto, Canada, July 27-30, 3072-3079.

4. Liu, J., Nikolaus, M., Bodur, K., & Fourtassi, A. (2022). Predicting backchannel signaling in child-caregiver multimodal conversations. *Companion Proceedings of the 2022 International Conference on Multimodal Interaction*, 196-200.

5. Bodur, K., Nikolaus, M., Prévot, L., & Fourtassi, A. (2023). Using video calls to study children's conversational development: The case of backchannel signaling. *Frontiers in Computer Science*, 5, Article 1088752.

6. Agrawal, A., Liu, J., Bodur, K., Favre, B., & Fourtassi, A. (2023). Development of multimodal turn coordination in conversations: Evidence for adult-like behavior in middle childhood. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Sydney, Australia (hybrid), July 26-29, 3275-3282.

7. Mazzocconi, C., O'Brien, B., El Haddad, K., Bodur, K., & Fourtassi, A. (2023). Differences between mimicking and non-mimicking laughter in child-caregiver conversation: A distributional and acoustic analysis. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Sydney, Australia, July 26-29, 253-261.

8. Mazzocconi, C., El Haddad, K., O'Brien, B., Bodur, K. & Fourtassi, A. (2023). Laughter mimicry in parent-child and parent-adult interaction. *Proceedings of the 1st International Multimodal Communication Symposium* (MMSYM 2023), Barcelona, Spain, April 26-28.

9. Goumri, D. E., Agrawal, A., Nikolaus, M., Vu, H. D. T., Bodur, K., Emmar, E., & Fourtassi, A. (2024). Chica: A developmental corpus of child-caregiver's face-to-face vs. video call conversations in middle childhood. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING 2024), Turin, Italy, May 20-25, 3153-3164.

10. Goumri, D. E., Agrawal, A., Nikolaus, M., Vu, H. D. T., Bodur, K., Semmar, E., & Fourtassi, A. (2024). ChiCA: un corpus de conversations face-à-face vs. Zoom entre enfants et parents. *Proceedings of the 35th Journées d'Études sur la Parole (JEP 2024) / 31st Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2024) / 26th Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues* (RECITAL 2024) (Vol. 2). Toulouse, France, July 8-12, 4-5. ATALA & AFPC.

11. Mazzocconi, C., O'Brien, B., Bodur, K., & Fourtassi, A. (2024). Do children laugh like their parents? Conversational laughter mimicry occurrence and acoustic alignment in middle-childhood. *Journal of Nonverbal Behavior*, 49, 53-83. https://doi.org/10.1007/s10919-025-00478-z

3) Participation in conferences[1] and summer schools during the doctoral project:

1. Journée des 50 ans du LPL, Laboratoire Parole et Langage, Aix-en-Provence. 13 May 2022 (Poster)

2. Journées d'études sur la Parole (JEP), Université de Nantes, France. 13-17 June 2022 (Poster)

3. Chorin Winter (summer) School, Chorin, Germany. 4-8 July 2022 (Poster)

4. Cognitive Science Conference, Toronto, Canada. 27-30 July 2022 (Poster)

5. Laboratoire Parole et Langage Open Doors, Aix-en-Provence, France. 15 October 2022 (Oral presentation)

6. Multimodality in Social Interaction Workshop by ILCB, Marseille, France. 3 April 2023 (Panelist)

7. ILCB Retreat, Porquerolles, France. 11-13 April 2023. (Poster)

8. Interspeech 2023, Dublin, Ireland. 20-24 August 2023 (Oral presentation)

9. Society for the Neurobiology of Language Conference, Marseille, France. 24 October 2023 (volunteering in the organization)

10. International Seminar on Speech Production (ISSP), Autrans, France. 13-17 May 2024 (Poster)

11. International Max Planck Research School (IMPRS), Nijmegen, Netherlands. 5-7 June 2024 (Poster)

12. Interspeech 2025, Rotterdam, Netherlands. 17-21 August 2025 (Oral presentation)

---

[1] The term "conference" is generic. It refers to "conference," "congress," "workshop," "symposium," "national and/or international meetings," etc. Indicate whether you gave an oral or poster presentation.

Résumé et mots-clés

Cette thèse porte sur la réduction de la parole, une caractéristique fréquente du langage parlé, qui peut se manifester par l'omission, la compression ou l'affaiblissement de segments dans la parole en continu. Longtemps considérée comme un phénomène phonétique ou articulatoire de bas niveau, motivé par une économie d'effort, la réduction est aujourd'hui reconnue, à la lumière de travaux récents, comme un phénomène plus structuré et multidimensionnel. L'objectif de cette thèse est de contribuer à une reconceptualisation de la réduction de la parole en tant que processus émergeant à l'intersection de la réalisation phonétique, de l'organisation prosodique, de la représentation lexicale et de la planification de la parole.

Ce travail propose une distinction entre deux types de réduction. Les réductions lexicalisées sont supposées être fréquentes, plus systématiques, et probablement représentées comme des formes holistiques dans le lexique mental. À l'inverse, les réductions non-lexicalisées sont envisagées comme des zones temporellement compressées de réduction extrême, sensibles au contexte et apparaissant de manière dynamique au cours de la production de la parole. Quatre études empiriques constituent le cœur de cette thèse et explorent ces deux types de réduction en s'appuyant à la fois sur des méthodes de détection descendantes (basées sur les représentations) et ascendantes (basées sur le signal).

Les résultats des quatre études montrent que les réductions lexicalisées sont plus stables entre locuteurs et à travers différents débits articulatoires, ce qui reflète des schémas d'usage bien ancrés. En revanche, les réductions non-lexicalisées varient selon le débit articulatoire, les frontières prosodiques et les catégories morphosyntaxiques, et présentent une plus grande variabilité interindividuelle et contextuelle. Contrairement à nos attentes, les indices acoustiques traditionnels comme la distinctivité vocalique ou la réduction de l'espace vocalique ne permettent pas de prédire de manière fiable les réductions non-

lexicalisées, soulignant le rôle central du débit articulatoire comme principal modulateur de la compression temporelle. Les données développementales recueillies auprès d'enfants âgés de 7 à 11 ans soutiennent également la distinction entre les deux types de réduction. Les enfants produisent des taux de réduction lexicalisée similaires à ceux des adultes. En revanche, les réductions non-lexicalisées sont largement absentes dans leur parole, ce qui peut s'expliquer par un débit de parole plus lente, suggérant qu'elles nécessitent des compétences linguistiques et motrices plus avancées.

Cette thèse apporte des contributions conceptuelles et méthodologiques. Elle remet en question les modèles qui considèrent la réduction comme un simple phénomène articulatoire, en la reconsidérant comme un processus stratégique lié à la planification, façonné par des facteurs linguistiques et interactionnels. Sur le plan méthodologique, elle propose une approche de détection duale, adaptée à l'analyse de corpus spontanés, permettant d'extraire un maximum d'occurrences de réduction selon les paramètres choisis et le phénomène ciblé.

Dans l'ensemble, les résultats soutiennent l'idée que la réduction constitue un continuum de comportements linguistiques, liés à la structure du langage et à son développement. En intégrant des perspectives issues de la phonétique, de la prosodie et de la recherche développementale, cette thèse propose un modèle unifié de la production de la parole, dans lequel la réduction offre un levier essentiel pour comprendre comment le langage est planifié, réalisé et acquis.

Mots clés : réduction phonétique, parole spontanée, variation, conversation, développement linguistique

Abstract and keywords

This thesis examines speech reduction, a common feature of spoken language, characterized by the omission, compression, or weakening of segments in connected speech. While previously regarded as a low-level phonetic or articulatory artifact driven by economy of effort, recent research suggests that reduction is a more structured and multi-dimensional phenomenon. The goal of the thesis is to support the reconceptualization of speech reduction as a process that emerges at the intersection of phonetic realization, prosodic organization, lexical representation, and speech planning. This work proposes a distinction between two types of reduction. Lexicalized reductions are assumed to be frequent, more consistent, and likely represented as holistic forms in the mental lexicon. Non-lexicalized reductions, by contrast, are supposed to be temporally compressed zones of extreme reduction that are context-sensitive and arise dynamically during speech production. Four empirical studies that make up the core of this thesis examine these two types of reductions using both top-down (representation-driven) and bottom-up (signal-driven) detection methods.

Findings from the four studies show that lexicalized reductions were more stable across various speakers and articulation rates, reflecting more entrenched usage patterns. Non-lexicalized reductions, on the other hand, varied with articulation rate, prosodic boundaries and morphosyntactic categories. They also showed greater variability across individuals and contexts. Contrary to our expectations, traditional acoustic cues such as vowel distinctiveness and smaller vowel space areas did not consistently predict non-lexicalized reductions, emphasizing the role of articulation rate as the strongest modulator of temporal compression. Developmental data from children aged 7 to 11 further support the contrast between the two types of reduction we proposed. Children produced lexicalized reduction ratios comparable to caregivers. Non-lexicalized reductions were largely absent in child speech, which can be explained by their slower articulation rates, suggesting they depend on more advanced linguistic and motor skills.

This dissertation offers both conceptual and methodological contributions. Challenging the models that treat reduction as purely articulatory, it frames reduction as a strategic, planning-based process shaped by linguistic and conversational factors. Methodologically, it introduces a dual detection approach suitable for naturalistic corpora to extract as many examples of reduction as possible depending on the phenomenon interested in and the selected parameters. Overall, the findings show that reduction is a spectrum of linguistic behaviors linked to linguistic structure and development. By integrating insights from phonetics, prosody, and developmental research, this work proposes a more unified model of speech production; one in which reduction provides key insight into how language is planned, realized, and learned.

Keywords: phonetic reduction, spontaneous speech, variation, conversation, linguistic development

To my family...

Acknowledgments

Writing this section feels, in many ways, more challenging than composing the dissertation itself. How can one possibly find the right words to express heartfelt gratitude to all those who have supported, encouraged, and inspired them along the way? This journey has been a true emotional rollercoaster, not only filled with intense learning about phonetic reduction but also profound personal growth. I am deeply grateful for these past four years, which have shaped me both as a researcher and as a human being.

It is with immense appreciation that I extend my thanks to everyone who has been part of this journey: my supervisors, the jury members, colleagues, friends, family, and all those -seen and unseen- who have directly or indirectly contributed to this work.

Even though words may fall short in conveying the full depth of my gratitude, I will try my best to honor the people who have made a difference during this important chapter of my life. If I do not mention everyone by name, it is not for lack of appreciation but only due to the limitations of space. Please know that I carry your support with me.

First and foremost, my deepest and sincere gratitude goes to my supervisors, Christine Meunier and Corinne Fredouille. Christine and Corinne, je vous suis profondément reconnaissant·e pour votre accompagnement constant, votre rigueur scientifique, et votre bienveillance humaine. Your unwavering support, thoughtful feedback, and constant encouragement have been the foundation of my growth as a researcher. Your warmth, patience, and empathy made all the difference, turning challenges into opportunities and uncertainty into confidence. You guided me without ever imposing, supported me while giving me space to explore, and offered reassurance when I doubted myself. Working under your guidance has been an honor and a privilege. I truly cannot thank you enough.

I am also very grateful to the members of my PhD jury: Prof. Richard Ogden, Prof. Christelle Dodane, and Dr. Barbara Schuppler, for kindly accepting to evaluate my work.

My sincere thanks go as well to the members of my Thesis Follow-Up Committee, Prof. Cédric Gendrot and Prof. Noël Nguyen, for their time, guidance, and expertise throughout the years. I am especially thankful for your support and for joining my defense committee, your input has helped shape this dissertation during the three committees.

Three individuals deserve a special mention for being early sources of inspiration and encouragement in my academic path.

Dr. Marie-Hélène Sauner, my teacher at Galatasaray University, whose influence on my academic journey has been truly decisive. In many ways, she is the reason I am here today, writing this dissertation. Marie-Hélène, you were the first to see this path for me and to encourage me to pursue it. You pushed me to apply for the Master's program and for the bourse d'ambassade, opening the door to studying in France and to everything that followed. Throughout the years, your constant encouragement, guidance, and trust never wavered. You believed in me when I was still learning to believe in myself. Words cannot fully express my gratitude, but I want you to know that it is thanks to you that I found this place, this program, and so many opportunities beyond it. Thank you, Marie-Hélène, from the bottom of my heart.

Dr. James German, thank you for your generosity, your mentorship, and your trust. You gave me my first real research opportunity during my second year of Master's studies; an experience that led me to my first conference presentation, as a nervous but determined student. That moment was transformative. Your unwavering support, even before I embarked on this PhD journey, and your reassuring smile throughout it all, meant more than I can say.

Dr. Abdellah Fourtassi, you have been a guiding light in my journey. From our first project together during my Master's internship, you have shown me what it means to be both a brilliant researcher and a supportive mentor. Your constant encouragement, your thoughtful "*feedback*" and the incredible research community you created in CoCoDev have been instrumental in my development. Thank you for believing in me and for giving me so many opportunities to grow. I am immensely grateful.

To all my CoCoDev colleagues, thank you for your collaboration and friendship. A special shoutout to Floor Meewis and Chiara Mazzocconi, our shared moments, both in and outside the lab, made this experience all the more memorable. Who else could convince me to cycle under the blazing Provençal sun through lavender fields?

I also wish to thank the researchers at LPL who supported my work in countless ways, especially Stéphane for your invaluable help with R scripts, and Roxanne and Philippe for your insightful feedback and thoughtful questions throughout my thesis.

To my lab mates, past and present: Airelle, Auriane, Axel, Clara, Dasha, Estelle, Giulia, Giusy, Lena, Lydia, Leela, Marine, Martha, Outhmane, Salvatore, and Yen Lin, thank you for

creating a supportive, collaborative, and inspiring work environment. Your friendship, advice, and solidarity made our lab a special place to grow as a researcher.

I must express a special thank you to Dasha, not just as a lab colleague, but as a dear friend and writing companion. From co-working sessions to moments of shared frustration, from casual chats to supplement discoveries to survive the stress, your presence made this process more bearable and even enjoyable. I'm truly grateful to count you among my friends.

Thank you also to Abhishek for being part of the co-working team and for your friendship.

To my FAVA singing group, thank you for bringing joy and balance into my life. Singing with you was a precious escape, a time to breathe, to laugh, and to reconnect with myself.

And now, to the pillars of my personal life:

To my mother, my anchor and my unwavering source of strength, Teşekkür ederim anne, iyi ki varsın. Your belief in me, even across the distance, gave me courage. Your love, quiet but fierce, gave me everything I needed to keep going.

To my best friend, Neslihan, thank you for being my mirror, my cheerleader, and my partner-in-complaints about academia. Your genuine joy for my successes, even more than my own, is a rare gift, and your loyalty and understanding mean the world to me. I am so lucky to have you by my side.

Finally, to Özer, my partner in every sense of the word, thank you for your patience, your emotional and practical support, and your unshakable faith in me, even when I had none in myself. You helped me stay grounded, helped me focus when my mind wandered, and celebrated even the smallest milestones with me. I am endlessly grateful to share this life with you, and excited for whatever comes next, together.

This dissertation is not just my achievement. It is the result of the collective support, love, and wisdom of everyone mentioned here, and many more. I carry your contributions with me, always.

Thank you, from the bottom of my heart.

Table of contents

List of figures

List of tables

General Introduction

Human spoken communication is a dynamic process which thrives on variability, with no two utterances from the same speaker ever being exactly the same. Listeners must constantly navigate through a speech signal that is inherently variable across speakers, contexts, and even within an individual's own speech (Foulkes & Docherty, 2006; Johnson, 2004; Lindblom, 1990). Any given word in a continuous stretch of speech can have countless acoustic realizations, adding to the complexity of spoken communication. Despite the continuous flow of varied speech signal, the primary goal of communication remains the successful transfer of a message while ensuring intelligibility for the listener.

Lindblom (1990) suggests that speech production relies on a listener's ability to discriminate among stored lexical items based on the distinctiveness of the acoustic signal. Crucially, the amount of explicit signal information required for successful lexical access is not fixed, it varies depending on contextual and communicative factors. This adaptability in speech processing highlights the idea that speakers do not always produce fully articulated forms but rather adjust their articulation based on communicative demands. Therefore, speech production occurs along a continuum, shifting between hyperspeech (output-oriented control, where clarity is prioritized) and hypospeech (system-oriented control, where effort is minimized). This balance allows them to maintain intelligibility while managing the physical and cognitive effort of speaking. Importantly, successful communication does not always require highly contrastive or fully specified acoustic signals. In many cases, listeners interpret reduced or ambiguous forms not only through bottom-up processing based on the discriminability of the acoustic signal, but also through top-down integration of linguistic, prosodic, and communicational context. This suggests that speech perception is not strictly dependent on the discriminability of the signal alone but rather involves a dynamic interplay between sensory input and higher-level predictive mechanisms (Kleinschmidt & Jaeger, 2015).

In much of traditional phonetic and phonological research, the concept of variation in speech is grounded in the assumption that there exists an underlying canonical or citation form, an idealized, theoretical representation of words often associated with dictionary pronunciations. In these forms, each phoneme is expected to be distinctly and fully articulated, making its acoustic and phonetic properties maximally salient (Jakobson & Halle, 2002). However, speech research has long grappled with a fundamental dichotomy (Ladd, 2006; Kohler, 2009; Ohala, 1991):

1) The theoretical model of speech, which assumes discrete, segmental units such as phonemes
2) The actual physical reality of speech production, which is fluid and continuous

This discrepancy poses a major challenge in phonetic and linguistic research. While traditional linguistic models conceptualize speech as being composed of discrete, clearly defined segments, natural speech exhibits far greater variability. In spontaneous communication, identifiable speech events do not always align neatly with theoretical phonemic categories, making it difficult to draw rigid distinctions between different phonetic forms. Expanded, canonical, and reduced forms frequently co-occur within the same conversation, phrase, or even within a single word, creating a dynamic interplay between phonetic characteristics. Within a given utterance, some segments retain a clear and distinct pronunciation, while others undergo systematic processes such as lenition or elision, resulting in a continuum of phonetic realizations.

Variation of speech sounds: early perspectives

This issue of phonetic variability and speech reduction has deep historical roots in linguistic theory. Earlier linguistic frameworks did not fully account for the variability in speech production. Classical perspectives, such as those proposed by Jakobson & Halle (1956), framed speech sounds in terms of invariant, distinctive features, emphasizing clarity and systematic contrast in perception. Phonemes would serve as stable mental representations despite variations in their physical realization. In this tradition, reductions were considered as "lazy" or "slurred" speech, implying a failure to articulate words properly (compared to their idealized code) rather than recognizing reductions as an inherent feature of fluent speech. As Jakobson and Halle put it:

"The slurred fashion of pronunciation is but an abbreviated derivative from the explicit clear speech form which carries the highest amount of information…. When analyzing the patterns of phonemes and distinctive features composing them, one must resort to the fullest, optimal code at the command of the given speakers." Jakobson & Halle (1956: 6)

Such accounts prioritized the analysis of canonical forms, implicitly treating variability as noise to be normalized or corrected. However, this perspective began to shift in the latter half of the 20th century, with increasing recognition of the regularity and communicative function of variation. While the traditional view assumed a relatively invariant phonemic structure, contemporary research increasingly suggests that phonemes are better understood as dynamic categories, continuously shaped by contextual and communicative constraints. Research in phonology, phonetics, and sociolinguistics (e.g., Browman & Goldstein, 1990a; Labov, 1986; Lindblom, 1990) highlighted that variation is not random or peripheral, but central to how speech is produced and understood in real-world contexts.

Within these emerging frameworks, reduction came to be viewed not as imprecision, but as a reflection of efficient, listener-oriented communication. Furthermore, reduction phenomena might challenge the classical concept of a fixed phonemic inventory, as they demonstrate that the boundaries between phonemes can be fluid and highly context dependent. Models such as Lindblom's Hyper and Hypo-speech Theory (1990) proposed that speakers dynamically adjust articulation based on communicative goals, balancing clarity and effort. Similarly, articulatory phonology (Browman & Goldstein, 1990) emphasized gestural coordination over fixed segmental targets, accounting for fluid transitions and overlap in natural speech. This reconceptualization of speech sounds as context-sensitive, gradient, and functionally adaptive, laid the groundwork for the contemporary perspectives on reduction.

Today, speech reduction is more widely understood as a pervasive and systematic phenomenon in which segments, syllables, or even entire words may be compressed, coarticulated, or deleted in fluent speech. Far from being anomalous, such reductions are intrinsic to naturalistic communication across languages and styles. Yet, despite its ubiquity, reduction was long sidelined in theoretical models, in part due to methodological challenges and historical bias toward carefully articulated, citation-form speech. Crucially, speech reduction is not merely the result of articulatory constraints or

communicative pressure; it is shaped by the structural properties of the language system itself. Languages differ in how reductions are realized, guided by typological features such as prosodic structure, morphological complexity, and lexical frequency patterns (Bybee, 2002b). This suggests that reduction is not a deviation from linguistic norms but a language-specific, rule-governed aspect of phonological organization, one that reflects and reinforces systematic variation rather than opposing it.

Canonical vs Reduced productions

A growing body of research suggests that "canonical" and "reduced" forms do not constitute strictly categorical oppositions, but instead exist along a continuum of pronunciation variants, ranging from fully articulated to highly reduced forms (Nolan, 1992). The recognition of speech as existing on a continuum challenges the traditional view of reduction as a mere deviation from an idealized, canonical form. Rather than being arbitrary or irregular, reductions emerge as a systematic and adaptive aspect of spoken communication, shaped by efficiency and contextual demands.

Articulatory phonology (Browman & Goldstein, 1990) further supports this gradient view, explaining reduction in terms of gestural overlap and magnitude. According to this model, pronunciation variants arise due to changes in articulation rate, leading to increased gestural overlap and decreased gestural magnitude and not necessarily due to the addition, deletion, or modification of gestures. This inherently implies that reduction is not a binary process but rather a continuous phenomenon. From this perspective, speech reduction is not merely an artifact of casual speech but a natural extension of variation in spoken language, resulting in a spectrum of reduced forms, and demonstrating how efficiency is maintained in real-time communication.

The privileging of canonical productions in traditional linguistic research likely stems from their association with faster and more accurate recognition in laboratory-based word recognition tasks (Ernestus & Baayen, 2007; Pitt et al., 2011; Ranbom & Connine, 2007). Canonical forms are often considered ideal targets in models of lexical access, due to their clearer segmental structure and assumed acoustic stability. Studies such as Bradlow and Pisoni (1999), and Bradlow and Bent (2002) have reinforced this perspective by demonstrating a "clear speech" benefit in intelligibility for native listeners,

particularly in noise or for individuals with hearing impairments. However, the generalizability of this advantage has been challenged. For instance, Bradlow and Bent (2002) found that non-native listeners do not necessarily benefit from clear speech to the same extent, suggesting that the processing advantage of canonical forms may be language-specific or listener-dependent.

Moreover, the assumed canonical advantage is often based on experimental designs involving words in isolation, where dictionary-like pronunciations are expected to be more frequent (Tucker, 2011). Yet, in spontaneous conversational speech, highly frequent pronunciation variants, such as flaps in American English or schwa deletion in French, can be processed as efficiently as, or even more efficiently than their canonical counterparts, especially when supported by contextual cues (Bürki et al., 2010; Connine, 2004). In fact, Dilley and Pitt (2010) show that perception of reduced forms is highly sensitive to temporal context and speaking rate, further problematizing the notion of a fixed, "ideal" form.

Nonetheless, research on larger linguistic contexts, including full sentences (Ranbom et al., 2009) has also found evidence for a canonical advantage, even in environments that favor reduced forms. While reduced forms may align more naturally with conversational speech patterns, canonical forms still provide perceptual advantages. However, these studies primarily involve carefully produced speech rather than spontaneous productions, making it unclear whether the speaking style really favored canonical variants. In contrast, studies on casual speech (Sumner, 2013; Tucker, 2011) suggest that reduced forms can be processed as efficiently as canonical forms when the speech is naturally produced. That said, recognition of severely reduced forms in spontaneous speech remains challenging for listeners, even though contextual cues improve comprehension (Ernestus et al., 2002).

Thus, while canonical forms have traditionally served as the baseline in phonetic and psycholinguistic models, their privileged status may reflect methodological bias more than empirical necessity. One perspective in this debate is that the canonical form alone does not define a word's "phonetic essence" (Niebuhr & Kohler, 2011). In light of research emphasizing the flexibility and adaptiveness of speech production, it becomes increasingly clear that canonical forms represent only one end of a continuum of

variation, rather than the standard against which all speech should be measured. Ernestus and Smith (Ernestus & Smith, 2018) propose that in studying the perception of reduced forms; the focus should shift away from what is present in the full, canonical form and toward what remains when certain phonetic elements are reduced.

Reconceptualizing reduction in speech

This evolving perspective has led to a reassessment of the role of reduction in everyday speech, offering new insights into the mechanisms underlying spoken language production, perception, and variation. As the recently growing body of research indicates that reduction is not merely a random or imprecise deviation from an idealized form, but rather a highly systematic and contextually governed process shaped by linguistic, cognitive, and communicative factors (Mitterer & Ernestus, 2008; Warner & Tucker, 2011).

Much of the prior research on speech reduction has primarily focused on highly perceptible and lexically driven reductions. Two examples from French include (Wu & Adda-Decker, 2020):

- je suis /ʒə sɥi/ (*I am*) produced as → [*ʃɥi*]
- je ne sais pas /ʒə nə sɛ pa/ (*I don't know*) produced as → [*ʃɛpa*]

These forms follow regular and predictable phonetic patterns and are relatively easy to identify because they involve frequent function words or commonly reduced syllables. Yet, speech reduction is not confined to these surface-level phonological transformations. It can encompass a wide range of subtler articulatory modifications that do not always conform to pre-established lexicalized patterns.

Some of these reductions are far more elusive, often escaping perceptual detection by the human ear due to their gradual, gradient nature. For instance, reductions that involve weakened consonantal gestures, vowel centralization, or coarticulatory effects may not be immediately recognized as distinct phonetic alterations yet still play a crucial role in shaping spoken language dynamics (Ernestus, 2014). Such subtler forms of reduction challenge traditional methodologies of speech analysis, as they do not always result in

clearly segmentable phonological transformations but rather emerge through continuous variation along articulatory and acoustic dimensions (Niebuhr & Kohler, 2011). Consequently, their identification requires more refined acoustic analysis tools, advanced corpus-based methodologies, and computational modeling to accurately capture and quantify the extent and nature of these processes (Schuppler et al., 2011; Tucker, 2011).

Given the systematic and structured nature of reduction, it is clear that this phenomenon deserves more extensive empirical investigation. Understanding reduction not only provides crucial insights into the efficiency and adaptability of spoken language but also has implications for speech perception models, automatic speech recognition systems, and language acquisition research (Aylett & Turk, 2004). Recognizing reduction as an integral aspect of natural speech, rather than a mere deviation from an idealized norm, shifts the focus toward a more comprehensive view of spoken language as a dynamic and adaptive system, warranting further interdisciplinary research.

Reduction as a multi-faceted phenomenon

Despite significant advancements in phonetic research, methodological challenges continue to impede a comprehensive and nuanced understanding of speech reduction in all its forms. One major difficulty is the high degree of variability in reduced forms and the inherent limitations of human perception (Johnson, 2004). Some reductions occur so subtly that they are nearly imperceptible, requiring advanced acoustic and computational analysis techniques for their detection (Ernestus & Warner, 2011). Consequently, many studies have concentrated on highly perceptible and lexically driven reductions, as these are more easily transcribed and analyzed using conventional methods. This focus on salient reductions has resulted in an incomplete picture of the phenomenon, potentially overlooking less obvious, yet equally systematic, reduction processes.

Additionally, a reliance on transcription-based methods introduces further limitations, as these approaches are inherently constrained by the listener's ability to detect and categorize reductions. This issue is exacerbated by the fact that different listeners may perceive reductions differently based on linguistic background, experience, and contextual expectations. As a result, speech reduction research has historically been skewed toward reductions that fit pre-established phonological frameworks, limiting our

understanding of its full variability. Addressing these challenges requires a more inclusive approach that integrates perceptual, acoustic, and computational methods to systematically analyze reductions. By refining methodologies and expanding the scope of investigation, we can develop a more accurate and comprehensive model of how speech reduction operates within natural communication.

Given this shift in understanding, it becomes essential to examine the mechanisms and patterns of speech reduction more closely, exploring the conditions under which reductions occur and their implications for models of speech production and perception. Extending the scope of reduction to include less explored forms will contribute to a deeper understanding of spoken language variability, highlighting the delicate balance between intelligibility and efficiency in human communication.

This thesis aims to investigate a range of speech reductions in casual and familiar speaking contexts with the goal of better understanding how and why these reductions occur by analyzing sufficiently large, natural speech datasets. The main hypothesis is that reduction is a broad phenomenon that requires a combination of theoretical perspectives and experimental methods, in order to discover various types and to better analyze them. This thesis will explore the influence of factors such as speaking rate, grammatical structure, articulation, prosody and how these reductions emerge in children's developing speech.

The next section provides a brief overview of the current state of research on speech reduction, highlighting key findings and ongoing debates. This is followed by a detailed description of the methodology used to detect and analyze reductions in spontaneous speech. The four main chapters that follow each present a core article, exploring different dimensions of speech reduction and together offering a deeper, more nuanced understanding of how and why reduction occurs in everyday conversations.

Chapter 1- State of the art

1.1. Understanding reduction: definition and prevalence

The work in this dissertation focuses on speech reduction in spontaneous French conversations, a phenomenon that plays a crucial role in shaping the fluidity and efficiency of spoken communication. Speech reduction is a pervasive and fundamental process in which linguistic units —such as segments, syllables, and/or entire words— are produced with weakened or altered articulatory and acoustic properties (Ernestus & Warner, 2011). This process exists on a continuum of acoustic-phonetic alterations, ranging from subtle articulatory changes to significant temporal compressions.

Phonetic reduction occurs when segments, syllables, or entire words are produced with less articulatory effort and acoustic-phonetic substance compared to their full, canonical forms (Clopper, et al., 2018). Within this broad framework, Johnson (2004) introduced the term "massive reduction" to describe two key phenomena: (i) syllable deletion and (ii) segmental changes. These two processes, though somewhat overlapping, can be systematically identified and quantified within phonetically transcribed speech corpora. Reduction can occur at various levels and impact not only spectral properties and temporal aspects but can also lead to the absence of segments. A unit can be considered reduced if:

(i)   Its acoustic representation involves fewer segments than its canonical pronunciation would suggest, such as the absence of segments (e.g., [ste] instead of [sete] for "c'était" (it was)) (Johnson, 2004).

(ii)   A segment of the canonical pronunciation is realized with the acoustic properties of another segment associated with less articulatory effort (e.g., lenition) (Browman & Goldstein, 1990b; Kirchner, 2013).

(iii)   The segments in the canonical form undergo phonological processes that create shortened or assimilated forms (Ohala, 1993; Kohler, 1991).

Additionally, reduction can affect different linguistic levels, from individual segments to entire phrases. Segment level reduction, such as vowel and consonant reductions has been widely investigated in phonetics and speech processing.

In French, the findings of Adda-Decker and Snoeren (2011) underscored the prevalence of speech reduction across various contexts, particularly in spontaneous French conversations. Their analysis of segment duration distributions revealed that casual speech exhibited significantly shorter segment durations compared to careful speech, with more than 30% of segments in conversational speech lasting only 30–40 ms. Cross-linguistic comparisons indicated that this pattern of reduction was not exclusive to French but also occurred in English, suggesting that casual speech might universally lead to a flattened duration distribution. Furthermore, they found that reductions were not limited to specific phonemes but affected both vowels and consonants. However, certain segments were particularly prone to reduction, with schwa [ə] being the most frequently deleted vowel, especially in high-frequency function words. Among consonants, liquids (/l/) and glides (/w/) showed high reduction rates, particularly in words like *je suis* (I am), where schwa deletion and consonantal assimilation contributed to substantial phonetic alterations. Voiceless fricatives tended to maintain longer durations, while voiced fricatives, particularly /v/, exhibit notable reduction in casual speech.

In his analysis of a corpus of American English conversations, Johnson (2004) found that 6% of content words and 4.5% of the function words exhibited reduction. Furthermore, reduction was highly pervasive at the segmental level: more than 60% of words in the corpus deviated from their citation form on at least one phone, while 28% of words deviated on two or more segments, which is a notable deviation rate, given that 68% of word tokens contained three or fewer phones. Segment deletion was particularly widespread. A syllable deletion analysis revealed that 5-6% of words in the corpus had at least one syllable omitted. However, segmental deletion was even more frequent; over 20% of words contained at least one deleted segment, and 5% of words exhibited two or more segment deletions, representing a strikingly high rate of deletion, underscoring the extent of phonetic reduction in conversational speech.

Similarly, quantitative research on speech reduction in Dutch highlighted its high prevalence in casual conversations. Schuppler and colleagues (2011) analyzed 153,200-word tokens from 10 informal Dutch conversations, reporting a 19% syllable deletion rate and a 40% rate of segment alteration or deletion. Additionally, Pluymaekers and colleagues (2005) found that reductions occur frequently in both function and content words, with lexical frequency playing a crucial role; high-frequency words were more likely to undergo reductions, such as vowel centralization or deletion. Taken together, these findings emphasize the systematic and widespread nature of phonetic reduction in spontaneous speech, affecting a range of speech segments and occurring across different languages and styles. In the following section, we will present an overview of the specifics of reduction, beginning with reduction of segments, such as vowels and consonants before expanding our focus to larger linguistic units.

1.1.1. Vowel Reduction

Vowel reduction is a widespread phenomenon observed across many languages, particularly in the context of unstressed syllables or informal speech (Van Bergem, 1995). The phenomenon of vowel reduction is particularly significant because of the continuous and overlapping articulatory nature of vowel transitions, unlike the more discrete articulatory events observed for consonants. This continuity makes vowel segments especially susceptible to variation and reduction in natural speech, as articulatory inertia and coarticulatory pressures accumulate more easily across adjacent vowels.

Vowel reduction typically involves centralization, shortening, or even complete deletion of vowels, resulting in a simplified vowel system in specific prosodic or contextual environments. Lindblom (1963) characterizes vowel reduction as a gradual and systematic process that results in a compressed and centralized vowel space due to a decrease in spectral and durational distinctiveness. Reduced vowels tend to assimilate to surrounding segments and become more centralized in the vowel space. Vowel reduction has found to be linked with many conditions and prosodic contexts such as in unstressed syllables, at lower prosodic boundaries and at increased speech rates (Lindblom, 1963), as well as in function words compared to lexical words (Van Bergem,

1993) and in durationally shortened areas in connected speech (Gendrot & Adda-Decker, 2005).

As the reduction in the spectral properties of vowels was often accompanied by a shortening of the vowel, Lindblom (1963) proposed that vowel reduction was a result of the formant undershoot caused by durational constraints. Undershoot implies that due to temporal constraints the articulators do not reach vowel-specific articulatory targets, resulting in a shrunken vowel space and less distinct formant patterns. However, it is also possible for vowels to be spectrally reduced without shortening, due to force-dependent undershoot, where reduced articulatory effort leads to slower and less extreme articulatory movements (Moon & Lindblom, 1994). Vowel reduction manifests itself in several ways acoustic, phonetically, and durationally.

Acoustic manifestations of vowel reduction include:
(i)   Centralization: Reduced vowels tend to move towards the center of the vowel space, often approaching schwa-like productions (Bates, 1995).
(ii)   Formant changes: F1 values tend to decrease, particularly in unstressed vowels; and the acoustic vowel space becomes reduced (Strycharczuk, Ćavar, & Coretat, 2021).

Phonetic manifestations of vowel reduction include:
(i)   Decreased articulatory effort: Reduced vowels show less extreme articulatory positions and movements (Bates, 1995).
(ii)   Increased coarticulation: Slower articulatory movements lead to more overlap between adjacent sounds.
(iii)   Underspecification: Reduced vowels, especially schwa, may be largely unspecified for tongue and jaw position, allowing for greater contextual variability (Bates, 1995).

In addition to these, vowel reduction can manifest itself in the shortening of durations; reduced vowels are typically shorter than their full counterparts (Fourakis, 1991). However, the occurrence and extent of vowel reduction can vary significantly between languages. For instance, English exhibits extensive vowel reduction (Lindblom, 1963), (Johnson, 2004) with many unstressed vowels reducing to schwa, while languages like Spanish show less dramatic reduction patterns (Broś, 2017). Factors influencing vowel

reduction include stress patterns, speech rate, word frequency, and the phonological structure of the language.

1.1.2.　　Consonant Reduction

Consonant reduction, on the other hand, has received comparatively less attention, except in studies focusing on consonant cluster simplification in children and second language acquisition (e.g., McCarthy, 2008; McLeod et al., 2001). Van Son and Pols (1999) state that there is a lack of data that would allow a comparison between the acoustic consequences of vowel and consonant reduction in a unified framework. This limits the utility of the knowledge obtained from vowel reduction research in understanding the consonant reduction patterns. Unlike vowels, which are produced through continuous movements within the vocal tract, consonants are typically articulated through more discrete positions and gestures. This fundamental difference means that reduction does not operate in the same way across the two categories: while vowel reduction often involves centralization or temporal compression, consonant reduction can take many diverse forms (van Son & Pols, 1999) and is generally harder to generalize.

In informal speech styles, consonants often exhibit articulatory weakening, reduced precision, or even deletion (Van Son & Pols, 1999). Acoustic evidence suggests that reduced consonants may lack clear closure, formant transitions, or bursts (Byrd & Tan, 1996; Duez, 1995; Son & Pols, 1995). For instance, Crystal and House (1982) found that almost half of stop consonants in their corpus of read connected speech lacked either a closure or a burst. Similarly, Warner and Tucker (2011) observed that in American English, highly reduced productions of stops and flaps occur frequently, with approximant-like tokens even in careful speech. They also reported that approximant-like realizations of expected stops and flaps in some conditions constitute the majority of tokens. In extreme cases, entire consonants may be deleted, as demonstrated in Greenberg (1999), where consonantal omissions were found to be a common feature of conversational speech. Similar to vowel reduction, consonant reduction manifests itself acoustically, phonetically, and durationally.

Acoustic manifestations of consonant reduction include:

(i) Spectral changes: Reduced consonants often show less distinct spectral features compared to their full forms (Parrell & Narayanan, 2018).

(ii) Formant transitions: The formant transitions into and out of consonants may be less pronounced in reduced forms (Van Son & Pols, 1999).

(iii) Center of Gravity (COG) shifts: Non-plosive consonants typically show a decrease in COG in spontaneous speech, indicating a reduction in articulatory effort (Van Son & Pols, 1999).

Phonetic manifestations of consonant reduction include:

(i) Consonant cluster simplification: One or more consonants in a cluster may be omitted or simplified (McCarthy, 2008).

(ii) Assimilation: Consonants may partially or fully adopt features of adjacent sounds (Duez, 2002).

(iii) Lenition: Stop consonants may be realized as approximants in certain contexts (Warner & Tucker, 2011).

(iv) Deletion: In extreme cases, consonants may be completely omitted (Greenberg, 1999).

In addition to these, consonant reduction can manifest itself through shortened segment durations, a process similar to vowels. Van Son and Pols' (1999) report that the decrease in duration of consonants is such that the relative duration, as a fraction of total VCV segment duration, remains unchanged, suggesting that the change in duration seems to be a result of a "global" increase of speaking rate in spontaneous speech. However, it should be stated that not all consonant categories are affected by the reduction in the same way, it is possible to observe category specific manifestations of reductions.

Consonant reduction is influenced by several key factors. Speaking style plays a crucial role, with spontaneous and casual speech exhibiting more reduction than careful or read speech (Greenberg, 1999). Phonetic environment also shapes reduction patterns, as surrounding sounds can affect the degree and type of modification (Duez, 2002). Additionally, faster speech rates tend to increase reduction, while high-frequency words are more prone to undergoing these changes (Parrell & Narayanan, 2018). Prosodic

position further influences reduction, with consonants in weak prosodic positions being more susceptible (Parrell & Narayanan, 2018).

1.1.3.   Challenges of a segment-based approach to reduction

Studying reduction using a segment-based approach presents methodological and theoretical challenges. In spontaneous speech, the assumption that speech unfolds as a sequence of discrete, identifiable segments often fails to reflect the complex, continuous, and overlapping nature of articulatory events. Reduced words are often difficult to annotate and segment; articulatory gestures blend across segment boundaries, and phonetic realizations often diverge from canonical forms to the point where individual segments may become acoustically untraceable.

Even in cases where a segment appears to be absent in the acoustic signal, it may still exert an influence on surrounding sounds through coarticulatory or prosodic effects, leaving behind what Niebuhr and Kohler (2011) describe as phonetic traces. These effects challenge the notion of the phoneme as a stable, recoverable unit. As Farnetani and Recasens (Farnetani & Recasens, 1997) point out, coarticulation results in a dynamic vocal tract configuration that reflects the influence of multiple articulatory targets at once, rather than clean, segment-by-segment articulation.

This raises important questions about the ontological status of the phoneme in connected speech. Is the phoneme a psychologically real unit, or an abstraction imposed by linguistic theory? In frameworks like Articulatory Phonology (Browman & Goldstein, 1992), the traditional segment is replaced by articulatory gestures that may overlap and coordinate in time, offering a more realistic model of how speech is produced and perceived. From this perspective, reduction is not an anomaly but a natural outcome of gestural dynamics and efficiency in communicative contexts.

Given these complexities, a purely segmental approach may obscure rather than illuminate the nature of reduction. Instead, there is increasing recognition of the need to

analyze reduction at larger linguistic units, such as syllables, words, or multi-word chunks, where patterns of articulatory coordination, frequency, and predictability better account for observed variability in spontaneous speech. This shift in perspective aligns with usage-based and exemplar models of speech processing (Bybee et al., 2016), which emphasize gradient variation and context-sensitive production over idealized, segmental templates.

1.1.4. Reduction at the word and multi-word levels

While research on speech reduction has spanned multiple levels of linguistic structure, a large focus has been placed on segmental reduction. This emphasis stems from the relative ease with which segmental processes can be analyzed within existing phonological frameworks (Kirchner, 2013). Phonetic models traditionally treat the segment as the fundamental unit of speech analysis, making it more straightforward to describe reduction in terms of phonological rules, constraints, or articulatory processes. Consequently, large-scale corpora and computational models have been developed to analyze and predict segmental reduction, facilitating a more systematic exploration of consonant weakening, vowel centralization, and segment deletion (Ernestus & Warner, 2011).

However, reduction extends far beyond individual segments. Entire words and multi-word sequences undergo systematic phonetic changes that present unique analytical challenges. Unlike segmental reduction, which can often be described in terms of articulatory undershoot or coarticulation, word- and phrase-level reductions involve interactions between prosody, frequency effects, and syntactic structure (Ernestus & Smith, 2018). The challenge of identifying and measuring these reductions lies in their subtle and context-dependent nature, they often involve gradient phonetic modifications rather than categorical segment loss. Moreover, prosodic restructuring, stress shifts, and rhythmic reorganization contribute to the variability of reduced forms, making them harder to codify in formal phonological models (Bell et al., 2009; Bybee, 2002a). In particular, Adda-Decker and colleagues (2005) demonstrate that in spontaneous French,

phenomena such as re-syllabification where segmental material is re-assigned across syllable boundaries, play a central role in shaping the surface form of utterances. These dynamic reconfigurations often obscure canonical word and syllable boundaries, complicating the task of aligning surface realizations with their underlying phonological representations and further highlighting the gradient, context-sensitive nature of reduction.

Another reason why word and multi-word reductions remain less studied is the lack of annotated large-scale corpora. Segmental reduction has benefited from phonemically transcribed datasets that enable systematic analysis, whereas comparable resources for multi-word reductions (annotated for fine-grained phonetic, prosodic, and usage-based factors) are still scarce. The statistical modeling and machine learning techniques that have advanced segmental reduction studies are still adapting to the complexities of phrase-level reduction, which requires capturing higher-level linguistic structures and their interactions.

Despite these challenges, research on word-level reduction has yielded significant insights. Studies have shown that high-frequency words and function words are particularly susceptible to reduction, often undergoing phonetic weakening, segment deletion, and vowel centralization (Ernestus, 2000; Johnson, 2004). These reductions make common words difficult to distinguish in fluent speech, leading to a reliance on contextual cues for comprehension. One well-documented example is schwa deletion in French, where words such as *petite* [pətit] may be realized as [ptit] (Adda-Decker & Snoeren, 2011). Similarly, in English, phrases like *going to* frequently reduce to [gʊnə] or even [g̃ə] in conversational speech.

Beyond isolated words, multi-word reduction represents another major area of study. Frequently co-occurring word sequences tend to form highly fused units, blurring word boundaries and creating reduction patterns that challenge traditional lexical segmentation (Bybee, 2002a). In French, for instance, the phrase *je ne sais pas* (I don't know) is often realized as [ʃepa] or [ʃpa], while in Dutch, function words are frequently shortened or even disappear entirely in casual speech (Binnenpoorte et al., 2005). These reductions are not arbitrary, they emerge from usage-based linguistic tendencies, where

high-frequency phrases become entrenched as fixed, reduced chunks, leading to stress loss, phonetic erosion, and eventual grammaticalization (Bybee et al., 2016).

Importantly, multi-word reduction is shaped by both linguistic and cognitive factors. Reduction patterns reflect not only historical sound changes but also real-time speech processing constraints, where speakers optimize articulatory effort and listeners rely on predictability to reconstruct meaning. This interaction between production ease and perceptual efficiency highlights the importance of studying reduction beyond the segmental level, as it provides insights into both language evolution and speech processing mechanisms (Torreira & Ernestus, 2012). Therefore, extending the study of reduction beyond the segments is crucial to gain a more comprehensive understanding of how spoken language varies dynamically across contexts, registers, and communicative needs.

Given the complexity and pervasiveness of speech reduction, it is necessary to examine it across multiple linguistic levels—from individual segments to entire multi-word sequences. This dissertation will specifically focus on multi-word sequences and propose a novel method for their detection (Chapter 2). However, before introducing this method, the next section will provide an overview of the key factors influencing reduction in casual speech**.**

1.2. Influencing factors in speech reduction

Reduction involves a complex interplay of linguistic, conversational and cognitive factors. Understanding these factors is crucial for developing comprehensive models of speech production, perception, and language processing (Tucker & Ernestus, 2016). The growing interest in spontaneous speech along with the development of large conversational corpora and advanced tools for speech analysis such as phonetic annotations and forced alignment (Adda-Decker & Snoeren, 2011) allowed a more in-depth analysis of reduction and the underlying mechanism. The following section focuses on linguistic and conversational influences that are known to shape reduction patterns, while acknowledging these are not the only factors that are in play in the production of reductions.

1.2.1. Linguistic factors

This section focuses on the phonological, prosodic, and lexical factors that influence speech reduction. While other aspects, such as syntactic or more complex morphological structures, may also be relevant, they are beyond the scope of the present discussion.

1.2.1.1. Phonological and prosodic factors

Phonological processes have been found to influence the phonetic realization of words in connected speech (Fruehwald, 2013). Most phonological alterations arise from modifications in the sequencing of articulatory gestures, which minimize muscular activity by either reducing the degree of displacement from a neutral position or shortening the temporal dimension through the overlap of consecutive gestures (Browman & Goldstein, 1992). Bybee and colleagues (2016) argue that reduction is determined by phonetic and phonological trends already present in the language which are accelerated in high-frequency phrases, and that this reduction is phonetically gradual.

One key factor is prosodic structure, which governs how reduction operates at different levels of phrasing and prominence (Aylett & Turk, 2004; Wightman et al., 1992). Research has shown that unstressed syllables and weak prosodic positions are particularly prone to reduction, as seen in vowel centralization and deletion across languages (Bybee, 2002a). For instance, vowels in unstressed syllables often undergo reduction, making them more similar to schwa, and word final /t/ may be deleted in certain contexts. These alterations are not random; they are governed by phonological rules and constraints that are sensitive to the surrounding phonetic environment (Kilbourn-Ceron et al., 2020). Phonotactic constraints determine whether reduced forms remain phonologically acceptable within a given language. For example, in English, word-final /t/ and /d/ deletion is more likely to occur when the resulting consonant cluster conforms to native phonotactic patterns (Guy, 1991b, 1991a). Similarly, schwa deletion in French is influenced by sonority principles, favoring reductions that preserve syllabic structure (Bürki et al., 2011).

Reduction is further modulated by lexical storage and retrieval processes. Words with multiple stored pronunciation variants, as suggested by exemplar-based models of speech production (Goldinger, 1998), may be produced in more reduced forms depending on contextual and speaker-related factors. This is evident in studies showing that speakers retrieve and produce more frequent pronunciation variants with greater ease (Bürki et al., 2010). Furthermore, variation in reduction processes across languages and dialects suggests that phonological rules governing segmental deletion can be both gradient and categorical, depending on language-specific constraints and lexical organization (Torreira & Ernestus, 2011).

Phonological neighborhood density, defined as the number of words that can be formed by a single phoneme substitution, addition, or deletion from the target word (Gahl & Strand, 2016), has significant implications for speech reduction by influencing both word recognition and production processes. In read speech, words with fewer phonological neighbors are recognized more easily and rapidly, but they require greater articulatory effort in production. Conversely, in conversational speech, high-density words —those with many phonological neighbors— tend to undergo more temporal and spectral vowel reduction than low density words (Gahl et al., 2012). However, this effect may be language-dependent; for example, Yao and Meunier (2014) report an opposite trend in French, suggesting that the relationship between neighborhood density and reduction may vary across linguistic systems. These findings align with broader research showing that neighborhood density inhibits word recognition due to lexical competition (Luce & Pisoni, 1998), but facilitates production by enhancing activation and retrieval efficiency (Stemberger, 2004). The tendency for high-density words to be reduced in spontaneous speech supports the idea that frequent lexical access allows for articulatory streamlining, contributing to phenomena such as vowel centralization, elision, and assimilation. This interplay between lexical structure and phonetic reduction highlights the adaptive nature of speech production, where communicative efficiency influences the extent of reduction depending on speech context.

1.2.1.2. Lexical factors: Frequency, contextual predictability and morphology

Numerous studies have shown that listeners process reduced word pronunciations more efficiently when these words are highly frequent (Brand & Ernestus, 2018; Ranbom & Connine, 2007). From a production perspective, high-frequency words are also more susceptible to reduction than low-frequency words. This phenomenon is attributed to the fact that frequently used words are more often exposed to production biases that lead to phonetic reduction (Bybee et al., 2016). The high frequency of usage facilitates the automation of articulatory routines, thereby increasing the likelihood of lenition, centralization or even elision in casual speech.

The relationship between frequency and reduction is further influenced by contextual ease. Clopper and Turnbull (Clopper & Turnbull, 2018) propose that phonetic reduction occurs more frequently in "easy" contexts —contexts that impose lower processing demands on both the speaker and the listener. They identify several characteristics that define such contexts:

(1) High-frequency words: More commonly used words tend to be reduced more often (Jurafsky et al., 2008).

(2) Low phonological neighborhood density: Words with fewer phonological neighbors are less likely to be confused with other words, making reduction less risky (Gahl et al., 2012).

(3) High semantic predictability: Words that are easily inferred from their linguistic context are more prone to reduction (Bell et al., 2003a).

(4) Second mention in discourse: Words that have already been introduced in conversation tend to be pronounced less clearly upon repetition (Fowler & Housum, 1987).

(5) Casual speaking styles: Reduction is more prevalent in informal speech than in careful or formal speech (Gendrot et al., 2012).

The fact that high-frequency words are produced more quickly suggests that lexical frequency influences both lexical access and motor planning during speech production (Balota & Chumbley, 1985), highlighting the decreased processing demands for both the speaker and listener. More frequent words have stronger memory representations and are retrieved with greater ease, thereby reducing the cognitive load required for articulation (Dell, 1990). This ease of retrieval, in turn, allows speakers to economize articulatory effort, leading to reduced pronunciations.

Furthermore, contextual predictability also plays a crucial role in reduction. Predictability refers to the degree to which a linguistic element (e.g., word, syllable) can be anticipated by a listener based on the surrounding linguistic and discourse context. When a word is highly predictable, either because it has been mentioned previously or because it fits well within the syntactic or semantic structure, speakers often reduce its articulation. Early experimental evidence for the influence of predictability on speech reduction comes from Lieberman (1963), who demonstrated that words with low contextual predictability in conversation are more intelligible when removed from their context and presented in isolation. In contrast, highly predictable words tend to be more reduced, suggesting that unpredictability prompts clearer pronunciation to aid listener comprehension. Similarly, Fowler and Housum (1987) found that repeated words in a narrative -typically more predictable on second mention- are less intelligible than their first occurrences. This finding aligns with the principle that new or unpredictable information is articulated more clearly than old or predictable information, as listeners rely less on contextual inference for word recognition when encountering novel lexical items.

Words that are highly predictable within a given linguistic context tend to undergo greater reduction, a phenomenon known as probabilistic reduction (Bell et al., 2003b; Jurafsky et al., 2000). This effect has been observed not only for lexical words but also for grammatical morphemes, indicating that predictability influences phonetic realization across multiple linguistic levels. Aylett and Turk (2004) argue that such reduction reflects a communicative trade-off: when context provides sufficient cues, speakers economize effort by reducing articulation, relying on the listener's ability to infer the intended word. Conversely, words that are less predictable tend to be pronounced more clearly to

enhance their perceptual salience, ensuring that listeners can successfully identify them (Priva, 2015).

Speech reductions tend to first affect the least informative elements of an utterance (Jurafsky, Bell, Gregory, & Raymond, 2001). These include function words that are predictable from the context, idioms, morphological markers (especially inflectional endings), discourse markers, and common temporal expressions such as dates (Johnson, 2004). Because these elements contribute little new information to a conversation, they are more prone to reduction without significantly impairing comprehension.

Additionally, morphological structure can also influence reduction (Hanique & Ernestus, 2012), as morphemes are not only carriers of meaning but also subject to articulatory and perceptual constraints. One key factor influencing reduction is morphological redundancy—when a suffix or an inflectional marker conveys information that is already retrievable from context, it becomes more prone to weakening or deletion (Zimmermann, 2016). For instance, in highly inflected languages, grammatical morphemes such as case markers, agreement suffixes, or plural markers may be reduced in spontaneous speech when they are predictable from syntax and discourse (Schuppler et al., 2012). In English, morphological reduction is commonly observed in function words and contracted forms, such as *"he is"* → *"he's"* or *"going to"* → *"gonna"* (Bybee, 2010). Similarly, in Dutch, segment deletion is frequently found in morphologically complex words. This suggests that certain morphological components are more vulnerable to reduction than others, especially when they carry less semantic weight or when contextual cues allow for their omission without impeding comprehension. Notably, the higher degree of reduction observed in Dutch function words may also stem from the fact that many of these words (e.g., *eigenlijk (actually)* from (Ernestus & Smith, 2018)) are at least bisyllabic, offering more material for phonetic reduction.

The position of a morpheme or segment within a word also influences its susceptibility to reduction. Word-final segments, particularly in unstressed syllables, are more prone to weakening or deletion due to articulatory lenition (Beckman, 1998). This may be particularly evident in languages like French, where final consonants in morphologically complex words are frequently elided unless followed by a vowel in liaison contexts (Durand et al., 2001). Similarly, in English, word-final /t/ and /d/ sounds are often subject

to deletion, especially in clusters (*west coast → wes' coast*) (Labov, 1969). Additionally, inflectional suffixes in agglutinative languages, such as Turkish, may be reduced when predictability is high.

1.2.2.   Conversational factors

This section aims to provide an overview of conversational factors such as speaking style and speech rate, as well as speaker and listener related constraints.

1.2.2.1.   Speaking style

A growing body of research has explored how conversational setting and speaking style influence the degree of reduction in speech. Rather than being fixed properties of a language, reduction patterns appear to vary significantly depending on the communicative context. Speaking style encompasses various aspects of articulation, including speech rate, precision, and formality, all of which shape how speakers produce linguistic units. In casual and informal settings, where communicative efficiency, processing ease, and speaker comfort are prioritized, speech is often more reduced compared to formal contexts (van de Ven et al., 2012). Reduction in casual speech is driven by a combination of articulatory economy, predictability, and interactive dynamics, which allow speakers to rely more on contextual cues rather than precise articulation.

In contrast to casual speech, formal contexts often demand clearer articulation, leading to reduced instances of coarticulation and deletion. Speakers aim for greater intelligibility, often adhering more closely to pronunciation norms and employing hyperarticulated speech (Lindblom, 1990). Clear and careful speech, which is typically associated with formal speaking styles or situations requiring enhanced intelligibility, is characterized by:

(1) Larger vowel space and increased vowel duration (Bradlow, 2002), reflecting enhanced articulatory effort.

(2) Increased intensity and slower speech rate (Picheny et al., 1986), allowing for more precise production of phonetic segments.

(3) Enhanced segmental contrasts (Smiljanić & Bradlow, 2005), which reduce the likelihood of confusion between similar sounds.

These phonetic adjustments serve to optimize comprehension, particularly in communicative situations where speech clarity is essential, such as when speaking to non-native listeners or in noisy environments.

On the other hand, casual speech typically involves less precise articulations than formal speech. The more relaxed articulation observed in informal contexts leads to shorter segment durations, greater gestural overlap, and a higher frequency of reduced word forms. Ernestus and colleagues (Ernestus et al., 2015) showed that the duration of schwa prefixes and word-final /t/ in Dutch past participles is affected by phonetic variables, indicating a gradient process of reduction in informal speech. Similarly, Johnson (2004) found that casual speech leads to phoneme deletion, particularly in high-frequency function words and morphologically predictable structures.

Further supporting the distinction between speech styles, studies comparing journalistic speech and spontaneous conversations highlight the greater reduction observed in the latter. Unlike read speech, journalistic speech constitutes a distinct speaking style that retains a relatively high level of articulation, marked by a moderate speech rate and fewer hesitations or fragmented words (Vaissière, 1997). While some degree of reduction is observed even in journalistic speech, such as vowel shortening, it remains less extensive than in fully spontaneous interactions. Importantly, research suggests that in spontaneous speech, reduction increases beyond what can be attributed solely to phoneme duration or speech rate. This indicates that additional linguistic and cognitive factors play a role in shaping reduction patterns in natural, unplanned discourse.

Beyond articulatory constraints, social and interactional factors play a crucial role in determining the degree of reduction. According to Bell's (1984) *audience design theory*, speakers actively adjust their speech to accommodate their listeners, modifying features such as formality, articulation, and phonetic detail based on the perceived needs and identities of their interlocutors. This social adaptation can lead to greater reduction in casual or informal settings, where mutual understanding is presumed and communicative efficiency is prioritized over articulatory precision. In such contexts, reduced forms may signal familiarity, solidarity, or shared knowledge, whereas in more formal or unfamiliar settings, clearer and less reduced speech may be used to ensure intelligibility and social appropriateness.

Tucker and Ernestus (2016) have emphasized the need for more extensive research on casual speech, arguing that linguistic insights based solely on careful speech may be incomplete. Without investigating spontaneous, informal speech, researchers may fail to capture essential aspects of real-world language processing and production.

1.2.2.2.   Speech rate

Speech rate, or tempo, is widely recognized as an important factor influencing phonetic variation in conversational speech, and it has been frequently examined in relation to speech reduction. Numerous studies have explored the idea that as speakers accelerate their speech rate, articulatory gestures tend to become compressed and simplified, potentially leading to more pronounced reduction effects (Moon & Lindblom, 1994). This acceleration might increase temporal constraints on speech production, often resulting in omission, weakening, or modification of sounds.

At faster speech rates, the articulatory gestures required to produce individual speech sounds are executed more rapidly, sometimes at the expense of clarity and precision. One common consequence of this is segmental elision, where certain phonetic elements are omitted entirely (Duez, 1995). Another consequence of increased speech rate is articulatory undershoot, where the articulators fail to reach their intended targets (Lindblom, 1963), leading to a centralization of vowel contrasts, reduced consonantal

gestures, and an overall blurring of phonetic boundaries. Additionally, at high tempos, speech segments frequently overlap, leading to increased coarticulation and assimilation effects, where sounds become more similar to adjacent segments or are entirely deleted due to gestural blending (Byrd & Tan, 1996).

Beyond phonetic reduction, speech rate also shapes listeners' perception of spoken language. Dilley and Pitt (Dilley & Pitt, 2010) demonstrated that contextual speech rate can influence lexical perception, showing that slowing the rate of surrounding speech can cause function words to be perceived as missing, while faster rates can create perceptual illusions of words that were never actually spoken. Similarly, Jacewicz and colleagues (Jacewicz et al., 2009) investigated variation in speech tempo among American English speakers and found that, although individuals tend to have characteristic speaking rates, some consistently speak in faster, more compressed phrases. This suggests that speech rate is deeply intertwined with both individual speaking style and regional dialect variation.

While many studies emphasize the connection between increased speech rate and greater phonetic reduction, Van Son and Pols (R. J. J. H. van Son & Pols, 1999) challenged this assumption, arguing that high speech rates do not necessarily lead to reduced articulation. Their acoustic study of Dutch speech demonstrated that speakers are capable of maintaining clear, fully articulated speech even at high tempos, suggesting that reduction is not an automatic consequence of faster speech but rather a controlled, language-specific adaptation that depends on communicative intent, linguistic structure, and contextual demands. This perspective aligns with models of hyper- and hypo-articulation (Lindblom, 1990), which propose that speakers dynamically adjust their articulation based on listener needs, environmental conditions, and cognitive load.

Further cross-linguistic research supports the idea that reduction patterns are not solely governed by tempo but also by language-specific phonetic constraints and stylistic norms. Torreira and Ernestus (Torreira & Ernestus, 2012), for example, examined the weakening of intervocalic /s/ in spontaneous Spanish speech and found that while higher speech rates correlated with increased reduction, the extent of weakening varied based on prosodic position and lexical frequency, highlighting the interplay between speech rate, phonological structure, and reduction processes.

These findings raise an important question: is it the act of accelerating speech that triggers reduction, or are some speakers inherently more prone to reducing due to their naturally faster speaking styles? In other words, is reduction a universal response to increased rate, or do individual speaking styles mediate this effect? This question points to the need to consider speaker-specific and listener-related constraints in the study of reduction, factors that go beyond general linguistic or conversational patterns and involve individual variability in both speech production and perception.

1.2.3.   Speaker and listener related constraints

While linguistic, conversational, and cognitive factors provide a broad framework for understanding speech reduction, it is essential to recognize that individual variability also plays a crucial role in shaping reduction patterns. Speakers differ significantly in their articulation due to a range of factors, including anatomical, physiological, and sociolinguistic influences. Similarly, listeners' perceptual abilities and expectations also impact how reduced speech is processed and understood.

One of the primary sources of variation in speech reduction is speaker identity, which encompasses characteristics such as age, gender, dialect, and speaking style. These speaker-specific characteristics, also called *indexical properties* (Clopper, 2004; Pisoni, 1993) might affect both baseline articulatory patterns and the degree of reduction. Research has shown that individuals exhibit distinct speech tempos and reduction tendencies based on these properties. Jacewicz, Fox, and Wei (Jacewicz et al., 2009) examined speech tempo in American English and found significant variation across different dialects and genders. They reported that women generally speak faster than men, a finding that has implications for speech reduction, as faster speech rates were found to correlate with increased phonetic reduction (Byrd & Tan, 1996).

Another relevant dimension in the study of speech reduction is anatomical differences between speakers. While physiological differences such as vocal tract length do not directly explain reduction mechanisms, they do shape the acoustic profile of speech sounds (Fuchs et al., 2008), influencing features like vowel space size and formant

patterns. These variations contribute to speaker-specific phonetic signatures that listeners learn to accommodate in perception. Rather than driving reduction processes themselves, anatomical differences highlight the need to account for individual variability when analyzing reduction patterns in spontaneous speech.

Beyond anatomical differences, paralinguistic factors, such as emotional state and cognitive load, also influence reduction. Polzin and Waibel (1998) demonstrated that emotions such as joy, anger, or sorrow modulate articulatory precision and speech clarity. For instance, angry speech often exhibits hyper-articulation and expanded vowel spaces, whereas casual or tired speech tends to be more reduced and less precise.

Environmental conditions can further shape speech production. The Lombard effect (Lombard, 1911; (Lau, 2008) describes how speakers instinctively modify their speech in response to background noise. Under noisy conditions, speakers increase their vocal intensity, slow their speech rate, and enhance articulatory clarity, thereby reducing phonetic reduction (Van Summers et al., 1988). This adaptive mechanism highlights how reduction is not merely an automatic consequence of speech tempo but is context-sensitive and listener-oriented (Smiljanić & Bradlow, 2009).

Just as speakers vary in their degree of reduction, listeners also differ in their ability to perceive and interpret reduced speech. Listeners develop speaker-specific expectations, adjusting their perception based on a speaker's typical phonetic patterns (Goldinger, 1996). Research suggests that experienced listeners are often able to restore reduced segments through top-down processing, using lexical and contextual cues to reconstruct missing phonetic material (Ernestus et al., 2002).

Listener adaptation is also shaped by linguistic background. Native and non-native listeners process reduced speech differently, with non-native speakers struggling more with heavily reduced forms (Broersma & Cutler, 2011). In highly reduced speech, native listeners rely on prosodic and syntactic cues to resolve ambiguity, while non-native listeners may experience greater difficulty due to differences in phonetic cue weighting (Mitterer & Ernestus, 2008).

A developmental perspective further reveals how reduction is shaped by the dynamic interaction between speaker abilities and listener needs. Children, in particular, represent a case where both production skills and perceptual systems are in flux, offering insight into how reduction patterns emerge and stabilize through communicative experience.

A particularly illustrative case of speaker-listener interaction in reduction behavior can be found in children. As both novice speakers and developing listeners, children exhibit distinct and evolving patterns of phonetic reduction shaped by motor control limitations, growing linguistic knowledge, and communicative awareness. Early child speech often simplifies complex articulatory gestures through processes such as cluster reduction and vowel centralization, but these reductions are developmentally constrained rather than communicatively optimized (Fikkert, 2006; Vihman et al., 1985). Kinematic and acoustic studies show that children over four years of age exhibit greater variability in producing unstressed syllables than stressed ones, suggesting that effective reduction of weak syllables demands fine motor control not yet fully developed at this stage (Ballard et al., 2012; Goffman, 1999). In fact, children tend to produce relatively longer and larger-amplitude articulatory movements, especially in unstresses positions, compared to adults (Green et al., 2000).

Importantly, reduction in child speech is not only motorically constrained but also shaped by listener awareness and prosodic structure. Function words, often reduced in adult speech, might remain relatively unreduced in child productions. Research has shown that children do not differentiate unstressed syllables based on morphosyntactic status (e.g., function vs. lexical syllables) to the same extent as adults, limiting their ability to form adult-like prosodic words (Goffman, 2004). Overall, reduction in speech offers a unique lens into the intersection of motor, linguistic, and listener-oriented development. The delayed and variable emergence of adult-like reduction patterns, especially in semantically light elements such as function words or determiners, suggests that children must not only acquire the articulatory skills necessary for temporal compression but also the pragmatic and semantic cues that inform which elements can be safely reduced without compromising communicative clarity. This dual dependency highlights reduction as a fundamentally communicative phenomenon, modulated by both production capacity and the need to be understood.

Given the wide range of factors that influence speech reduction —spanning individual speaker characteristics, environmental conditions, and listener expectations, it becomes crucial to develop effective methods for detecting and analyzing these reductions. The variability introduced by speaker-specific traits, cognitive and emotional states, and conversational context means that reductions can manifest in diverse ways, from subtle vowel centralization to complete phoneme deletion. Accurately identifying these reductions requires a combination of acoustic, articulatory, and perceptual analysis techniques, as well as computational approaches that account for speaker- and context-dependent variability.

The following section explores the methodologies used to detect speech reductions, examining both traditional phonetic analyses and modern machine learning techniques that enable a more precise and automated identification of reduced speech forms.

1.3. Detecting reductions

The study of speech reduction presents significant methodological challenges, particularly at the word and multi-word levels. Traditional approaches have relied on manual annotation, where trained phoneticians listen to recordings and transcribe reductions based on auditory perception. However, this process is extremely time-consuming and prone to biases; listeners may fail to detect certain reductions, especially when they involve subtle phonetic modifications rather than categorical segment deletions. Furthermore, perceptual salience does not always align with actual acoustic reduction, leading to inconsistencies in manual transcriptions (Ernestus & Warner, 2011).

To address these limitations, Automatic Speech Recognition (ASR) and forced alignment techniques have emerged as valuable tools for detecting speech reduction patterns at scale (e.g., Adda-Decker & Snoeren, 2011). Early ASR research struggled with the high variability in word pronunciations, which significantly impacted system accuracy, as canonical dictionaries often failed to account for frequent reductions in spontaneous speech (Strik & Cucchiarini, 1999). As a result, much of the early ASR work on reduction was concerned with variation at the segmental level. As ASR technology

has evolved, new strategies have been introduced to better model reduction, including the integration of reduced word variants into pronunciation dictionaries.

While early ASR models primarily accounted for segmental reduction, more recent systems have attempted to model word and multi-word reduction by expanding pronunciation dictionaries. One strategy is the inclusion of multiword units, which are frequently co-occurring sequences that tend to be pronounced as a single, reduced unit (Strik, Elffers, Bavcar, & Cucchiarini, 2006). For example, going back to the previous example in French, the phrase *je ne sais pas* (I don't know) and its reduced forms such as [ʃepa] or [ʃpa] can be integrated into pronunciation dictionaries to improve recognition accuracy. However, ASR systems still face difficulties when dealing with less frequent multi-word reductions or context-dependent reduction patterns, particularly when no specific acoustic models have been trained to recognize them.

A particularly well-documented case of such reduction involves the omission of schwa in casual speech. Adda-Decker and Snoeren (2011) reported that approximately 5% of ASR errors could be attributed to the unpredictable appearance or disappearance of schwa, leading to frequent mismatches between actual spoken forms and the predictions made by acoustic word models. Similarly, ASR misrecognition of reduced forms can occur when homophonic or near-homophonic sequences emerge due to reduction. For instance, the French phrase *quai de Seine* [kɛdəsɛn] ("bank of the Seine") can be misrecognized as *quête saine* [kɛtsɛn] ("healthy quest"), due to a combination of schwa deletion, regressive voice assimilation, and segmental merging —common processes in casual speech (Snoeren et al., 2006).

Beyond ASR transcription errors, forced alignment methods provide a more systematic, bottom-up approach to detecting speech reduction. Forced alignment tools automatically map phonetic transcriptions onto acoustic signals, revealing regions where speech deviates from canonical forms (Adda-Decker & Lamel, 1999). This method has proven particularly useful for quantifying vowel reduction and duration trends (Adda-Decker, Gendrot, & Nguyen, 2008). Adda-Decker and Snoeren (2011) then extended this technique to effectively quantify speech reduction in large spoken corpora across various speech styles, from broadcast news to telephone and face-to-face conversations.

Forced alignment techniques rely on canonical pronunciation dictionaries to compare expected and actual realizations of speech. When words or segments are heavily reduced or deleted, the forced alignment system struggles to fit the canonical form onto the acoustic signal. This results in anomalously short segment durations, which serve as a quantitative indicator of reduction. For example, in the phrase *student athletes*, if the unstressed syllable *-dent* is reduced, the alignment system will assign very short segment durations to that region, reflecting its weakened or omitted articulation (Adda-Decker & Lamel, 2017). This approach has been successfully used to analyze reduction in both English and French.

By leveraging large speech corpora, forced alignment enables researchers to systematically track reduction patterns across different speech styles, ranging from broadcast news to casual conversation (Adda-Decker & Snoeren, 2011). Studies have demonstrated that higher proportions of short segments correlate with greater degrees of temporal reduction, providing a quantitative measure of speech compression (Schuppler et al., 2008).

Despite these advancements, ASR systems still struggle with reduction in spontaneous speech, particularly in casual and conversational registers, where pronunciation variability is greatest. Previous studies have reported that ASR word error rates range from 10% in journalistic speech to over 15% in casual telephone speech, even when trained on large corpora (Prasad et al., 2005). Reduction-related errors account for a significant portion of ASR misrecognitions, particularly in cases of homophone substitutions, extreme vowel weakening, and segment deletion (Vasilescu et al., 2009).

To mitigate these issues, ASR systems have introduced frequent reduction forms such as *wanna*, *dunno*, and *gonna* as separate dictionary entries (Binnenpoorte et al., 2005; Strik & Cucchiarini, 1999). However, this approach is inherently limited—it requires pre-specifying reduced forms, which fails to account for highly variable and context-dependent reductions. In contrast, forced alignment provides a more data-driven approach, enabling researchers to detect new and unpredictable reduction patterns by analyzing deviation patterns in phonetic segment durations (Adda-Decker & Snoeren, 2011).

In sum, while manual transcription has played a foundational role in documenting speech reduction, its limitations in scalability and perceptual bias have led to increasing reliance on ASR-based methods. The integration of pronunciation dictionaries in ASR has helped model word and multi-word reduction, though challenges remain in handling spontaneous, context-sensitive reduction patterns. Forced alignment methods offered a promising alternative, allowing researchers to automatically detect reduction patterns by identifying mismatches between expected and actual segment durations although the need for novel and complementary methods is clear. Building on existing methods for identifying reductions, the following section will outline the aims and the scope of this dissertation, situating them within the context of the current state of the art before introducing a novel bottom-up approach for detecting reduced zones in the speech flow.

1.4. Aims and scope of the thesis

This dissertation aims to explore reductions in spontaneous dyadic conversations by establishing a distinction between two types of reductions. The primary objective is to gain a deeper understanding of the mechanisms underlying reduction in casual speech. To achieve this, it is essential to analyze a sufficiently large speech dataset that provides a natural and familiar context for speakers, allowing reductions to occur spontaneously and reflecting their everyday language use.

Much of the existing research, as discussed in the previous section, has primarily focused on either segmental level reductions or perceptible reduced variants. Consequently, frequently occurring reduction forms have been examined in greater depth, perhaps to the extent that they are overrepresented in reduction studies. However, these forms represent only a fraction of the phenomenon. In connected speech, numerous reductions go unnoticed because they do not correspond to stored reduced variants in speakers' mental lexicons. This dissertation seeks to broaden the scope of reduction research by introducing a complementary type of reduction—one that encompasses more subtle, temporally compressed speech segments. Establishing this distinction is crucial, as it necessitates methodological advancements that move beyond traditional manual detection methods, which rely heavily on listeners' perceptual biases. Instead, this research takes a bottom-up approach, working directly with speech data to capture and characterize multi-word reductions in a more objective manner. This approach will provide empirical justification for distinguishing between the two proposed types of reduction.

The dissertation is structured as follows:

Chapter 2: Methodology - How are various reduction phenomena captured?
This chapter outlines the methodological framework used to identify reductions within conversational corpora. The chapter begins with a detailed presentation of the two corpora selected for analysis, highlighting their relevance and structure. Following this, two complementary detection approaches are introduced. The first is a top-down manual

procedure, used to identify reductions based on a set of perceptible forms. The second is a bottom-up automated method, which targets temporally compressed zones in speech by applying predefined temporal and phonemic parameters, serving as indicators of potential reduction zones. Together, these approaches form a robust foundation for the extraction of reduced speech examples, which are further examined in the subsequent chapters of the dissertation.

Chapter 3: Can we distinguish different types of reductions in casual speech, and what factors influence their production?

This chapter is dedicated to establishing a conceptual distinction between different types of reduction in casual speech, addressing the limitations of existing methods, which primarily capture frequent, regularized forms. In addition to proposing a distinction between these and more subtle reductions, this chapter introduces a novel detection approach based on temporal parameters. By comparing manually identified lexicalized reductions with automatically detected temporally reduced zones, corresponding to non-lexicalized reductions, this chapter examines key factors influencing reduction production, including articulation rate, morpho-syntactic categories, and phoneme characteristics. This analysis provides a comparison of the two reduction types, offering a more nuanced understanding of their occurrence in spontaneous speech.

Chapter 4: To what extent do articulatory and phonetic factors such as vowel space size, vowel distinctiveness, and articulation rate account for the production of non-lexicalized reductions?

This chapter explores the relationship between vowel space characteristics, articulation rate and the production of non-lexicalized reductions. Building on Chapter 3's findings on speaker variability in reduction performance, this chapter investigates whether these differences stem from articulatory and phonetic factors; more specifically, vowel space size, vowel distinctiveness, and articulation rate. Vowel space reflects the range of tongue movements in speech production, while vowel distinctiveness measures how clearly vowels are acoustically separated. The central hypothesis is that speakers with a smaller vowel space and lower vowel distinctiveness, indicating a more compact and less differentiated articulation, would be more prone to producing reductions. By

linking reduction patterns to vowel space characteristics, this chapter provides insights into whether non-lexicalized reductions result from articulatory constraints or other linguistic and conversational factors, offering deeper insight into inter-speaker variability in spontaneous speech.

Chapter 5: How does prosodic structure, particularly the position in the prosodic hierarchy influence the occurrence and distribution of non-lexicalized reductions?

This chapter focuses on the role of prosody in reduction. Building on previous research, which has linked reductions to their position within the prosodic hierarchy, this chapter examines whether phonetic reductions occur more frequently in prosodically weak positions. Using the bottom-up approach mentioned in Chapters 2 and 3, this chapter aims to provide a more comprehensive perspective on the distribution of reductions within prosodic domains by taking into account the influence of prosodic phrase length.

Chapter 6: How do children produce lexicalized and non-lexicalized reductions, and what does this reveal about the acquisition of the linguistic and interactive skills involved in speech reduction?

This chapter extends the investigation into children's interactions, offering a developmental perspective on reduction mechanisms. By examining reductions in child speech, this chapter aims to demonstrate that reductions are complex linguistic and conversational skills requiring not only articulatory control but also advanced linguistic and interactive competencies. As reduction is considered a multifaceted skill dependent on various factors, this chapter analyzes the occurrences of both lexicalized and non-lexicalized reductions within a children's corpus (ages 7 to 11) to determine at what stage of development they occur more prevalently. The findings will shed light on how children progressively acquire the ability to navigate reductions, balancing efficiency with communicative clarity in their developing speech.

By establishing a more nuanced framework for understanding speech reductions, this dissertation aims to contribute to a more comprehensive view of phonetic variability in spontaneous speech, ultimately enriching our understanding of spoken language processing and production.

Chapter 2- General Methodology

This chapter presents a detailed account of the methodological framework developed to identify zones of reduced speech in spontaneous conversational data. The approach combines both top-down and bottom-up methods with the goal of assembling a sufficient and representative dataset of reduction instances. We begin by presenting the two conversational corpora that serve as the basis for this dissertation. Key details are provided regarding their structure, levels of linguistic annotation, and overall relevance for the detection of reductions. This section aims to establish the suitability of the data for investigating reduction in ecologically valid, interactional contexts.

The second part of the chapter is dedicated to the detection methods themselves. It is structured in two main sections. The first details the top-down approach, explaining its theoretical motivation, implementation steps and inherent limitations. The second section introduces the bottom-up approach, designed to capture reduction patterns that may fall outside the scope of a purely top-down strategy.

Together, these complementary approaches provide a robust methodological basis for identifying reduced forms in conversational speech. While the bottom-up detections are not exhaustive, they offer valuable access to reduction patterns that might otherwise go unnoticed. The data extracted through these methods serves as the foundation for the analyses presented in the following chapters.

2.1. Speech corpora

A well-established finding in phonetic and sociolinguistic research is that reductions are more frequent in informal, spontaneous speech than in read or scripted speech (Ernestus & Warner, 2011; Johnson, 2004; Nakamura et al., 2008). Casual speech contexts often involve faster speaking rates, less articulatory precision, and greater reliance on shared knowledge between interlocutors, all of which contribute to the likelihood and

extent of phonetic reduction. Consequently, the selection of a corpus composed of naturally occurring, informal conversations is essential for capturing reduction phenomena as they occur in ecologically valid communicative settings. Such data allows for the observation of reduction patterns in their most frequent and varied forms, thereby maximizing the number and diversity of analyzable instances.

In line with this, the present dissertation focuses primarily on adult speakers' production of reductions in spontaneous interactions, with the aim of quantifying and characterizing reductions in a naturalistic speech environment. In addition, a secondary objective is to examine how children produce reduced forms in comparable conversational contexts, which may shed light on the acquisition and development of reduction patterns in spoken language.

To address these goals, two corpora were selected, each designed to represent spontaneous, informal dialogue within one of the two speaker groups (i.e., adults and children). The next section will present the details of these two corpora.

2.1.1. Corpus of Interactional Data (CID)

The Corpus of Interactional Data (CID) (Bertrand et al., 2006, 2008; Blache et al., 2017) is a French audio-visual corpus designed to elicit a wide range of linguistics phenomena in spontaneous conversational settings. It consists of eight dyadic interactions between same-sex colleagues, totaling eight hours of dialogue. All interactions were recorded in a sound-proof booth using high-quality audio and video equipment to ensure clear signal capture. This setup is particularly important for analyzing subtle and often phenomena such as vocal feedback, turn-taking, and speech overlap.

The 16 participants (10 women and 6 men) are native speakers of French, primarily from southern France. Crucially, the interlocutors were chosen based on pre-existing familiarity, as they were colleagues with established social relationships. This familiarity was intended to support natural, fluid interactions and to allow speakers to engage

spontaneously and comfortably, minimizing the artificial constraints often introduced by experimental tasks.

Two thematic prompts were used to initiate conversation: one focused on professional conflicts, the other on unusual and unexpected personal experiences. These themes served only as starting points to help participants enter conversation. Importantly, speakers were explicitly told they could diverge from the prompts at any time. This design choice aims to preserve conversational authenticity and encourage the emergence of naturalistic linguistic features.

Consisting of a total of 127.713 tokens and 297.703 phonemes produced by the 16 speakers, the CID corpus also includes rich, multi-level and multi-modal annotations covering phonetic, prosodic, syntactic and gestural dimensions of speech among others. These detailed annotations enable comprehensive analyses of interactional dynamics and speech production and will be used for the studies mentioned in the following chapters. The corpus is publicly available through the Ortolang repository[1].

2.1.2. Child Interpersonal Communication Analysis Dataset (CHICA)

The Child Interpersonal Communication Analysis (CHICA) corpus (Goumri et al., 2024) was developed to address the underrepresentation of spoken language data from middle-childhood (6-12 years) in existing corpora. While most linguistic research has concentrated on early childhood or adult populations, middle-childhood remains relatively unexplored. CHICA aims to fill this gap by providing ecologically valid data for studying the development of linguistic and conversational skills in children aged 7 to 11.

The current version of the dataset includes recordings of 15 children interacting with one of their caregivers. These interactions were recorded in the children's homes to

---

[1] https://www.ortolang.fr/market/corpora/sldr000027/v2.1

preserve the natural dynamics of everyday conversation. Unlike laboratory settings, which can feel artificial to children and negatively impact the spontaneity of interactions, recording in familiar environments supports more fluid and authentic communication. This approach aligns with findings in interactional linguistics, emphasizing the context-sensitivity of social and linguistic behavior (Dideriksen et al., 2023).

Participants are grouped by age into three cohorts: 7, 9, and 11 years old. Each dyad participated in two types of sessions: (i) a remote conversation via Zoom video call and (ii) a face-to-face interaction recorded using eye-tracking equipment. During both sessions, dyads engaged in a loosely structured word guessing game. One participant would think of a word while the other tried to guess it, with both participants free to ask or offer hints. After a successful guess, roles were reversed. Parents were encouraged to end the session after 10–15 minutes, ensuring a roughly equal number of guessed words for each participant to maintain conversational balance.

For the purposes of this dissertation, we focus exclusively on Zoom video call sessions, as these recordings offer separate audio tracks for each speaker. This technical feature significantly enhances transcription accuracy, phoneme alignment, and subsequent analyses. Initial transcriptions were generated automatically using WhisperX (Bain et al., 2023), an enhanced version of the Whisper speech recognition model. WhisperX provides word-level timestamps through forced alignment. Additional multi-modal annotations were performed on the corpus, specifically for nonverbal behaviors such as gaze and head nods. To prepare the data for reduction analysis, additional manual corrections were applied to the WhisperX outputs. These included restoring disfluencies, repetitions, laughter, and short utterances such as feedback signals, all of which are crucial for accurate phonetic alignment and reduction analysis. Following these corrections, forced alignment was performed using SPPAS (Bigi, 2015) to ensure consistency with the phonetic annotations of the CID.

Finally, our study narrows its focus to two age groups, 7-year-olds and 11-year-olds, to explore how reduction mechanisms vary across developmental stages, specifically between younger and older children (and their caregivers) within the corpus. This age-grouping decision is grounded in developmentally sensitive theories of speech production, which emphasize that phonological and articulatory representations

undergo significant reorganization across childhood, with younger children relying more on holistic motor patterns and older children beginning to integrate perceptual and articulatory schemas in more adult-like ways (M. A. Redford, 2019a). By contrasting 7- and 11-year-olds, we aim to capture these developmental shifts and their impact on the production of reductions, which require flexible integration of motor and perceptual representations. This exploratory subset of the corpus consists of 16.768 tokens and 42,342 phonemes produced by 20 participants: 5 children from the 7-year-old group (mean age = 7;3), 5 children from the 11-year-old group (mean age = 11;3), and their respective caregivers.

Corpus	Speaker	Gender	Duration	Speaking Situation	Token Count	Phoneme Count
CID	16 adults	10F/6M	8 hours	spontaneous	127.713	297.703
CHICA	10 children 10 caregivers	4F/6M 5F/5M	2.5 hours	semi-spontaneous	16.768	42.342

Table 1. Summary of corpus characteristics for the two corpora used in analyzing reductions: CID and CHICA.

2.2. Detecting Reductions

2.2.1. Top-Down approach: detecting known reduction forms

The most documented forms of reduction concern words or sequences of words used very frequently in everyday language. Their reduced forms can appear several times within the same interaction and might appear as different examples of the same lexicalized representation due to their frequent occurrence. A frequent example of such reductions is *je sais* ("I know") in French, which is frequently produced as [ʃsə], [ʃə] or [ʃ] instead of [ʒə se] in various contexts (Figure 1 illustrates several examples of the sequence *je sais* produced by one speaker). The phonetic trace of the reduction is thus sometimes distant from this representation, and the examples form a heterogeneous set. These reductions are most often identifiable by listeners and transcribers who already have a perceptible representation of these forms due to their frequent occurrence in both perception and production.

A top-down approach is most effective when clear mental representations of reduction forms already exist in speakers' and annotators' mental lexicon. This approach focuses on frequently used sequences in daily French conversations, where both speakers and listeners can recognize these frequent forms. The top-down extraction method consists of two steps described in the following sections.
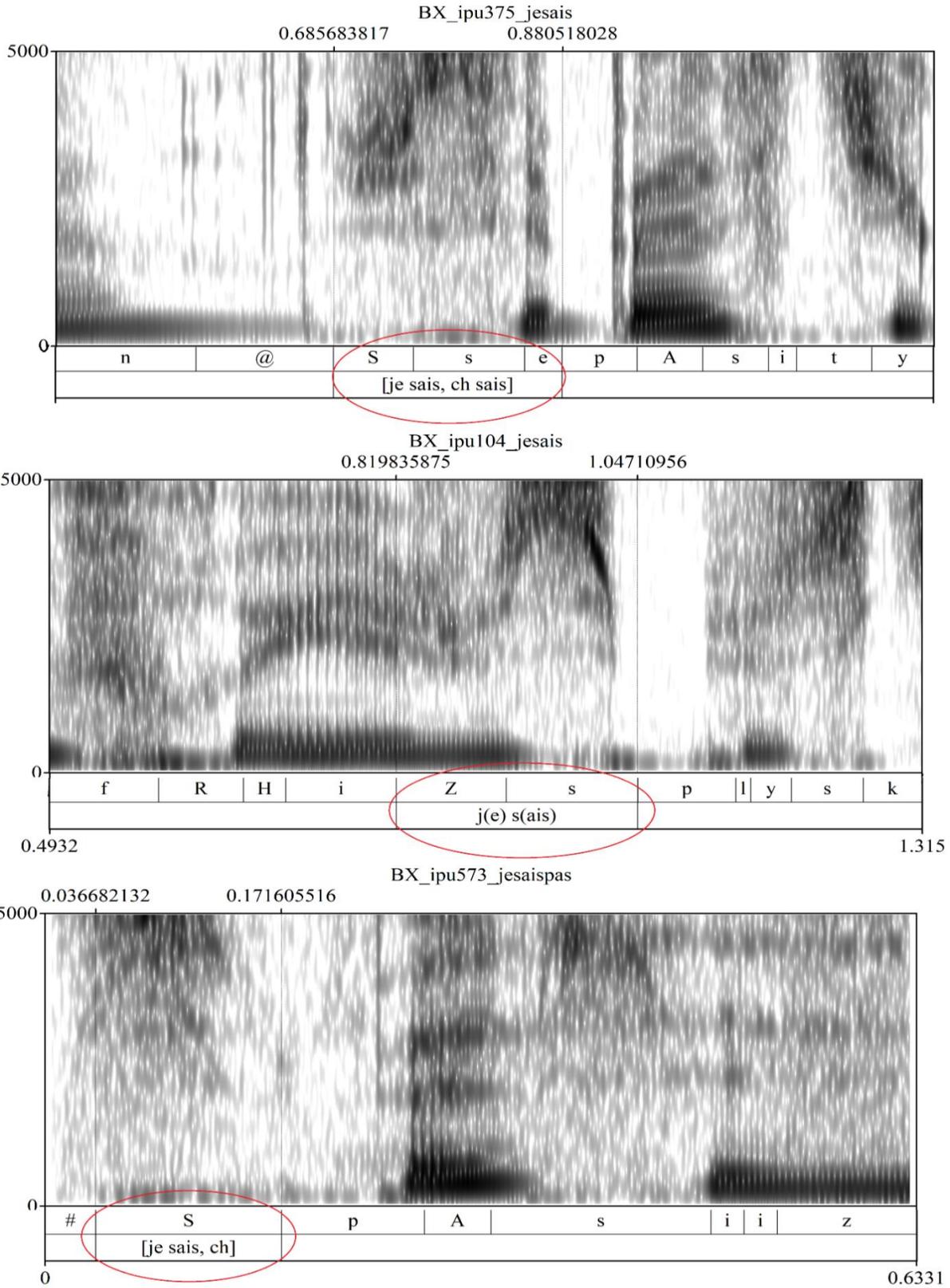
Figure 1. The sequence "je sais" (*I know*) reduced as [ʃsə], [ʒs], and [ʃ] by the speaker BX from the Corpus of Interactional Data (Bertrand et al., 2008).

2.2.1.1. Item selection

To gather reduced sequences using a top-down approach, we first compiled a list of commonly reduced sequences in French, as previously identified in the literature (Adda-Decker & Snoeren, 2011; Wu & Adda-Decker, 2020). To refine our focus, we conducted an informal questionnaire with five native speakers (mean age: 23 years; four had prior academic training in linguistics, while one worked as a brand advertiser), asking them to identify the reductions they most frequently produced and perceived in daily conversations. From the responses, we selected the items that overlapped with those mentioned in the literature, resulting in a speaker-informed inventory of 13 frequently occurring items from different morpho-syntactic categories. The target items for the top-down collection method are as follows:

- Adverbs : alors (then), enfin (finally), puis (then), voilà (there), de toute façon (anyway), quand-même (anyway)

- Pronoun + Verb combinations : je peux (I can), je sais (I know), je suis (I am), tu sais (you know)

- Conjunctions : alors que (whereas), parce que (because), puisque (because)

2.2.1.2. Detecting and collecting reduced items within a corpus

The collection of the aforementioned target items was conducted semi-automatically using Praat and the Corpus of Interactional Data (CID) (Bertrand et al., 2008), which comprises eight hours of dialogues between colleagues. A key advantage of this corpus is its detailed transcription and annotation across multiple linguistic and conversational levels, including phonetics, prosody, syntax, discourse, and gestures. Additionally, the corpus includes Enriched Orthographic Transcription (TOE) (Bertrand et al., 2008), which captures transcribers' perceptions of pronunciation variants rather than solely providing canonical transcriptions. For instance, if a speaker produces *tu sais* (you know) in a

reduced form, the annotation includes both the canonical form (tu sais) and the reduced variant [tse] using brackets (Figure 2).

Using the CID and its available annotations, the 13 target items were first identified within the token (words) tier to locate all occurrences. Each occurrence was then manually evaluated to determine whether it exhibited phonetic reduction. The TOE annotations provided useful insights into reduction patterns but were not entirely sufficient for identifying all reduced occurrences. In some cases, reductions were evident from the audio-visual signal but were not explicitly annotated in the TOE. Therefore, evaluations were conducted based on the audio-visual signal of speech, with reductions defined as instances where the produced sequence contained fewer segments compared to its canonical form.

For the purpose of this study, an item was considered reduced based on the following two criteria: (i) segment deletion, as in [*tse*] instead of [*tyse*], and (ii) segment assimilation, as in [*ʃe*] instead of [*ʒə se*]. The latter occurs when the /ʒ/ sound in *je* and the /s/ sound in *sais* merge, with the /ʒ/ influencing the /s/ to produce a /ʃ/ sound.

When a reduced item was identified, it was extracted along with its surrounding tokens to preserve the contextual information. Each instance was then annotated individually in a new TextGrid, with phonetic transcription corresponding to the observed pronunciation. Figure 2 illustrates an example of a target sequence '*tu sais*' showing its representation in the token tier, its corresponding enriched orthographic transcription in the IPU tier, and the phonetic transcription derived from forced alignment in the phon tier.
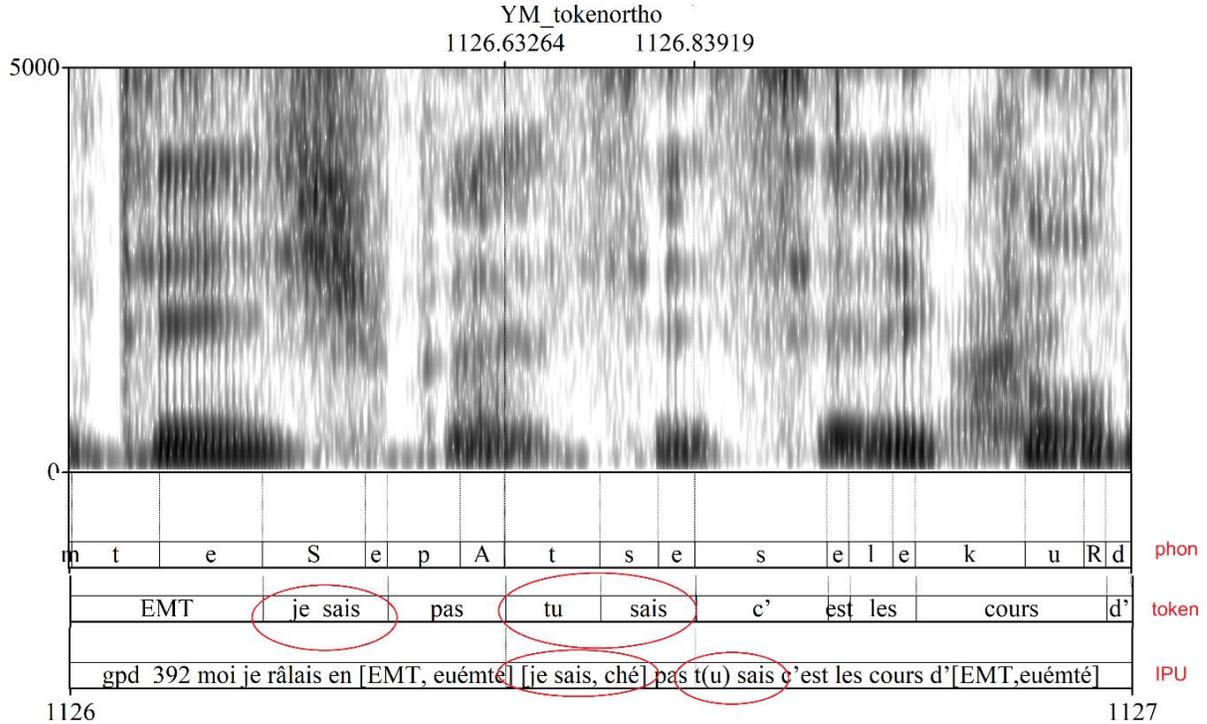
Figure 2. An example of a detected target sequence, '*tu sais*' and the annotations used for a manual reduction identification using audio-visual signal on Praat.

Using this top-down approach, a total of 2566 reduced items were identified out of 3634 target sequences produced by 16 speakers in the corpus, representing 71% of the total targets. These identified reductions will serve as the basis for the analyses presented in the subsequent chapters.

It is important to acknowledge the limitations of the top-down method in identifying reductions. This approach is restricted to a relatively small subset of reduction forms, as it relies on the existence of stable mental representations of these forms in both speakers' and listeners' mental lexicons. While effective in capturing exemplar-based reductions, the detection process remains largely manual and thus time-consuming. Moreover, accurately determining whether a sequence is reduced requires a certain level of phonetic expertise, as reductions are often subtle and may not be immediately perceptible in the audio signal alone. Furthermore, given that not all reductions have fixed or widely shared representations across speakers, additional methods are necessary to identify less frequent and more irregular reduction patterns that may not be systematically stored as exemplars of a target item.

67

2.2.2. Bottom-Up approach: detecting extreme reduction zones

As previously discussed, speech reduction is a complex and multi-faceted phenomenon that can manifest in diverse ways. The top-down approach, while effective in capturing frequently occurring and mentally represented reduction forms, is inherently limited in its scope. It is unable to systematically detect less conventional, highly compressed speech sequences that do not conform to established phonological patterns. To address this limitation, we introduce an automated bottom-up approach designed to detect extreme reduction zones within conversational speech.

Given that human perception of phonetic variation is influenced by prior linguistic knowledge (Elman & Mcclelland, 1984), subtle reductions may go unnoticed in manual annotation. The proposed bottom-up method leverages computational processing to systematically identify reduced sequences, with a particular emphasis on detecting extended multi-word reduction zones rather than isolated segmental reductions. While previous temporality-based approaches have been used to detect speech reductions (Chapter 1, Detecting reductions), our method refines this approach by introducing a dedicated set of parameters aimed at capturing longer, highly compressed speech sequences.

The primary objective of this approach is not to exhaustively catalog all possible reductions but rather to compile a sufficient dataset of extremely reduced sequences for further analysis. The obtained dataset will facilitate a deeper understanding of the mechanisms underlying speech reduction and the linguistic and conversational factors that influence the emergence of extreme reduction patterns.

The identification of extreme reduction zones was conducted using a custom Python script employing a sliding window approach with a fixed duration of 230ms. The detection process operates as follows:

a. **Input processing:** The script reads the phonetic alignment file (only a specific format (.lbl extension) is accepted as input) and initializes the sliding window at the beginning of the file.

b. **Initial detection:** The sliding window (230ms) begins at the first phone boundary of the alignment file.

c. **Phoneme count evaluation:** If at least 6 phonemes are detected within the window, the sequence is marked as a potential reduction zone.

d. **Incremental expansion:** The window advances by shifting its starting point to the next (second) segment. If the new window also contains at least 6 phonemes, only the newly included segments are added to the previously saved sequence to prevent redundant detections.

e. **Termination condition:** If a window contains fewer than six phonemes, the expansion process stops, and the detected sequence is finalized.

f. **Iteration:** The sliding window continues progressing through the alignment file until the final segment is reached.

g. **Output:** At the end of the process, the script provides a Textgrid file where the detected reduced sequences are indicated with their temporal information (start and end points, the duration etc.) in addition to a Text file where all of the detected sequences are found in a list.

A step-by-step application of the detection script is illustrated in Figure 3. This method mainly aims to ensure the detection of extended reduction zones. The identified reduction sequences will be analyzed to determine their phonetic, prosodic, and conversational characteristics, in order to provide further insight into the nature of phonetic reduction in conversational speech.

Following the development of the automated detection process, the accuracy of the output was assessed through manual evaluations based on the audio-visual speech signal. This step ensured that only correctly detected sequences were retained for further analysis. A sequence was classified as erroneous if the phonetic annotation did not align with the actual spoken segments. Most errors stemmed from misalignment issues such as inaccuracies in phoneme duration or positioning, or from inappropriate liaison insertions -where the aligner applied a liaison phoneme per French language conventions, even when the speaker omitted it during production.

1

885.646106   885.875818

| # | t | s | e | Z | e | t | Z | u | r | p | a | f | i | n | i | w | e |

First 230ms window ⟹ The window contains six phonemes. The sequence "tseZet" is saved.

2

885.67656   885.907143

| # | t | s | e | Z | e | t | Z | u | r | p | a | f | i | n | i | w | e |

Second 230ms window ⟹ There are six phonemes in the window. Only the "Z" is added to the previously saved sequence.

3

885.727027   885.95761

| # | t | s | e | Z | e | t | Z | u | r | p | a | f | i | n | i | w | e |

Third 230ms window ⟹ The window contains seven phonemes. "ur" are added to the previously saved sequence.

4

885.755741   885.986324

| # | t | s | e | Z | e | t | Z | u | r | p | a | f | i | n | i | w | e |

Fourth 230ms window ⟹ The window contains six phonemes. However, they were previously added to the sequence. Window stops adding to the sequence and the sequence is saved.

Figure 3. A step-by-step representation of the bottom-up detection of reduced zones in speech. Using the 230ms and 6 phonemes configuration. In the end the sequence "tseZetZur" (*tu sais j'ai toujours (you know I always had)*) was saved as a reduction zone. The script stopped adding phonemes to the segments in the fourth window as they had been previously added. The sequence ends and the window continues to slide.

During this evaluation, a crucial factor emerged: the quality of the phonetic alignment used as input significantly influenced the script's performance. To assess this impact, we evaluated the system using two versions of phonetic alignment provided in the CID dataset: v2 and v3. Both versions contain the same orthographic tokens and phonemes; however, v3 includes phonetic alignments performed using SPPAS software (Bigi, 2015). In this version, additional manual corrections were applied to further enhance alignment accuracy, particularly in terms of phonetization, tokenization, and liaison adjustments. Our system demonstrated more robust detection performance when using v3 as input, likely due to its improved alignment accuracy.

A total of 2144 sequences from 16 speakers were detected using the v2 phonetic annotations as input. Manual evaluation of these sequences revealed that 775 detections were erroneous, resulting in an error ratio of 36%. Table 2 details individual distributions of detection and error counts as well as the proportion of errors within individual detections.

SPEAKER (CID)	V2: 230MS/6 PHONEMES	ERRORS	ERROR RATIO %	V3: 230MS/6 PHONEMES	ERRORS	ERROR RATIO %
AB	84	26	31%	50	9	18%
AC	141	54	38%	93	14	15%
AG	185	39	21%	113	19	17%
AP	181	65	36%	94	7	7%
BX	141	64	45%	93	30	32%
CM	128	53	41%	67	21	31%
EB	348	92	26%	236	21	9%
IM	62	33	53%	28	13	46%
LJ	123	42	34%	63	18	29%
LL	32	16	50%	8	1	13%
MB	104	44	42%	71	24	34%
MG	141	48	34%	76	28	37%
ML	125	51	41%	83	19	23%
NH	104	13	13%	50	9	18%
SR	69	41	59%	27	7	26%
YM	176	94	53%	61	16	26%

Table 2. Using the v2 and v3 of phonetic alignment, the number of potential reduction zones flagged by the system and the number of erroneous detections based on manual evaluations. Error ratio is calculated as the proportion of erroneous detections within the total detections.

To further assess system performance, we tested the script using v3 of the phonetic annotations available for the CID dataset. With v3, the system identified a total of 1212 sequences as potential reduction zones. However, manual assessment indicated that 256 of these detections were inaccurate in terms of temporal alignment and segment correspondence, leading to a lower error ratio of 21% despite detecting fewer reductions (see Table 2 for individual distributions of detections and error rates for v2 and v3).

The improved accuracy with v3 suggests that more precise segmentation and better alignment lead to more reliable results. The higher number of detected reductions in v2, despite its higher error ratio, can be attributed to misaligned segment boundaries. When segment boundaries were not properly corrected, the system flagged extra short segments as reductions, even when they did not correspond to actual reduced productions in speech. In contrast, v3's improved alignment led to a significant reduction in false positives and overall better performance.

The selection of a 230ms window spanning 6 phonemes was a strategic compromise informed by both theoretical considerations and empirical testing. Our primary objective was to identify and analyze instances of extreme reductions in speech, where phonemes are realized with minimal temporal and articulatory cues. Given this, it was crucial to define a window that would be short enough to plausibly contain highly compressed sequences, yet not so short that it would drastically limit the number of analyzable cases.

The 230ms / 6 phonemes configuration corresponds to an average segment duration of approximately 38.3ms per phoneme, which is considerably below average segment durations reported for standard French pronunciation (e.g., typically 50-80ms per segment, depending on the prosodic context and phoneme class) (Bartkova & Sorin, 1987). Sequences falling within this window are therefore strong candidates for exhibiting phonetic reduction. However, we note that such averages do not imply uniform compression across all six phonemes. In some cases, the sequence may include only one or two extremely short segments, while the others remain closer to standard durations. Nonetheless, the overall compression serves as a useful indicator of potential zones of reduction.

To evaluate the trade-off between reduction sensitivity and data availability, we also tested two shorter window sizes: 180ms and 200ms, while keeping the phoneme count threshold constant at six. This choice was motivated by our interest in detecting longer reduction sequences that may span multiple words. These window sizes yield average phoneme durations of 30ms and 33.3ms, respectively, which indicate even more extreme cases of temporal compression. However, the number of sequences detected under these conditions was significantly lower: 95 sequences for 180ms and 429 sequences for 200ms, compared to 956 correctly detected sequences (out of a total of 1213 detections) for 230ms. Table 3 illustrates the number of detections for each set of configurations.

Additionally, erroneous detections in extremely short windows occurred due to challenges in annotation and alignment. In the 180ms and 200ms conditions, the alignment quality comes forward as an even more important factor. The error ratios for these configurations were not calculated as we decided to use the detections obtained through 230ms window. Taken together, these considerations led us to adopt the 230ms / 6-phoneme window as an effective balance, sensitive enough to identify meaningful cases of reduction while still ensuring data availability.

SPEAKER (CID)	180MS 6 PHONEMES	200 MS 6 PHONEMES	230MS 6 PHONEMES
AB	2	15	50
AC	9	29	93
AG	8	48	113
AP	11	33	94
BX	10	37	93
CM	1	23	67
EB	29	103	236
IM	0	8	28
LJ	0	18	63
LL	1	1	8
MB	7	25	71
MG	6	25	76
ML	7	26	83
NH	3	14	50
SR	0	8	27
YM	1	16	61

Table 3. The number of detected sequences using two alternative configurations for the detection.

Figure 4 illustrates an example of an output sequence detected using the bottom-up method. The script outputs the phonetic transcription and the temporal interval of the identified reduced sequence. To provide additional context, a second tier containing the token representation has been manually added. In this example, the detected sequence consists of six segments [*spɥiko*] within the 230ms window, corresponding to the larger phrase *je pense **puis** comme* ("I think since…"). The bolded letters indicate the segments found within the reduced sequence.

Because the algorithm continuously adds segments as long as they meet the reduction parameters (230ms and 6 phonemes), it is capable of detecting reduction zones containing anywhere from six to even twenty phonemes, capturing instances of extreme phonetic reduction.

Figure 4. An example from a sequence detected using the bottom-up method. The script only provides the phonetic transcription and the interval of the segments within the reduced sequence. The second tier containing the token was added to add context. The sequence consists of 6 segments [spɥiko] that belong to the larger sequence "je pen**se puis co**mme" (*I think since*).

74

Figure 5 also presents another example of a reduced sequence detected by the system. In this case, the script identified a sequence containing 20 phonemes within a 639ms window. The detected sequence, [de ʒɑ̃k ave dez‿ɛkspeʁj] is part of a larger phrase "***des gens qui avaient des experiences***" (*People who had experiences*). The bolded letters indicate the segments included in the reduced sequence. This example highlights the system's ability to capture extended reduction zones, further demonstrating the variability and extent of phonetic reduction in casual speech.

It is important to emphasize that the number of detected reduced sequences is highly sensitive to the configuration parameters applied in the detection script. Adjustments to factors such as window size, phoneme count, or temporal thresholds can substantially impact which sequences are identified.

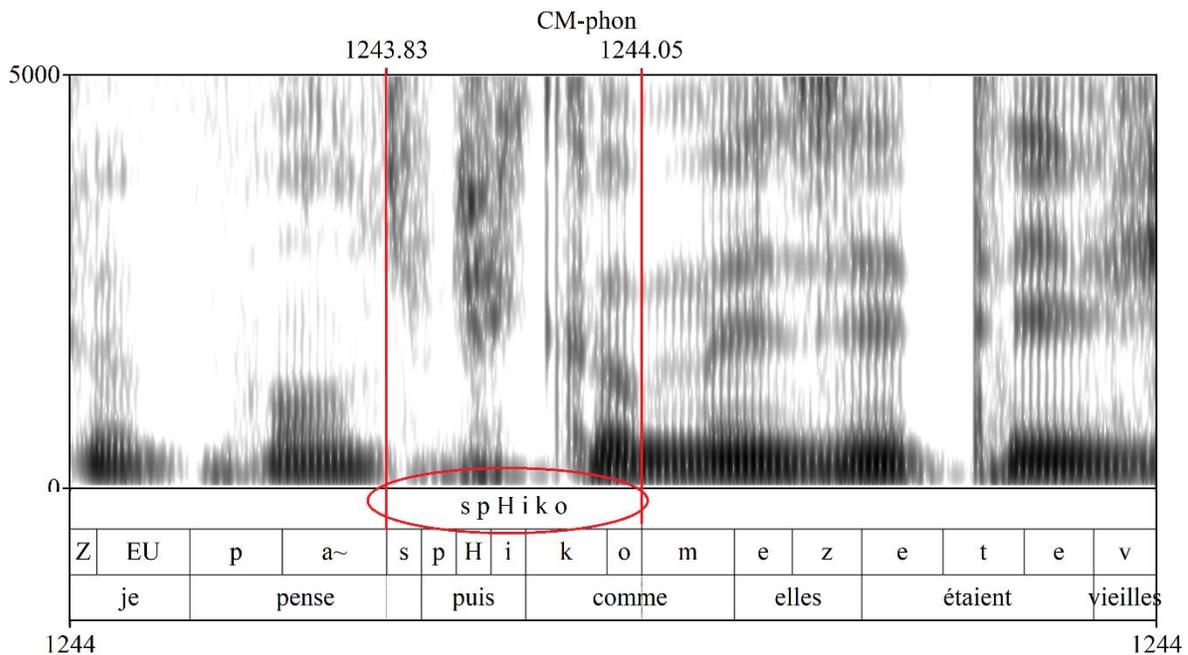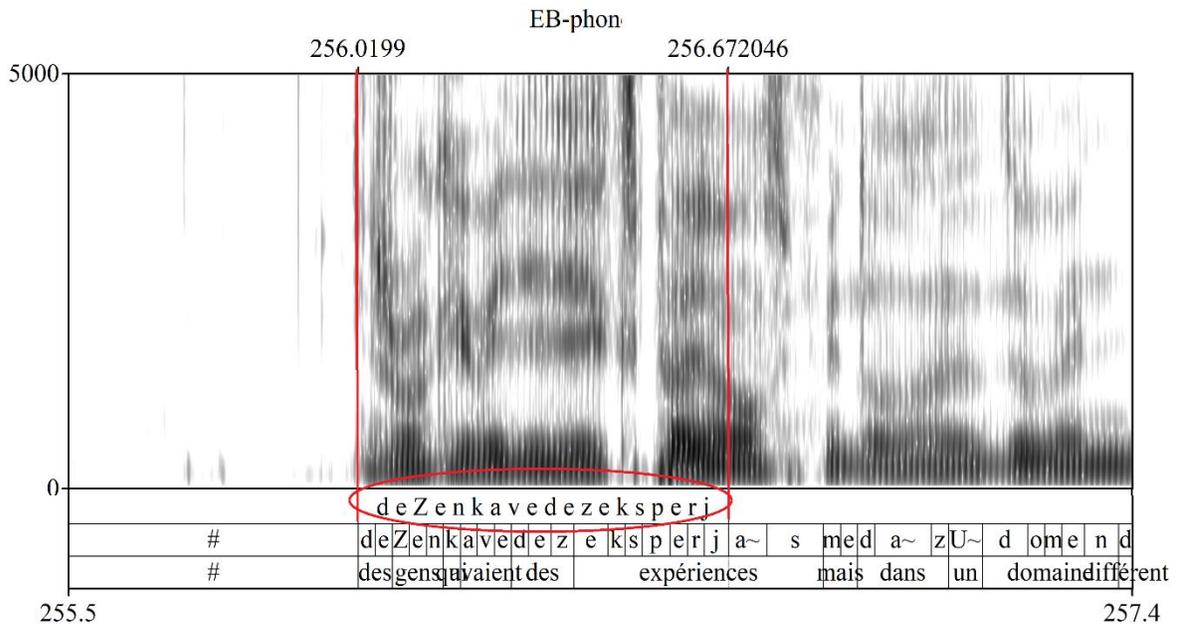Figure 5. Another example from a sequence detected using the bottom-up method. The script only provides the phonetic transcription and the interval of the segments within the reduced sequence. The second tier containing the token was added to add context. The sequence consists of 20 segments [deʒɑ̃kavedezɛkspeʁj] that belong to the larger sequence "Des gens qui avaient des expériences..." (*People who had experience*).

Crucially, the reductions captured by this script do not necessarily reflect fixed or regularly recurring phonetic patterns. Instead, they often emerge as context-dependent adaptations to various communicative and prosodic constraints. Factors such as speech rate, position within the prosodic domains, prominence, or even speaker-specific articulatory strategies can all contribute to the occurrence and degree of reduction. As such, the script is particularly valuable for identifying reduction phenomena that are shaped by these dynamic features of speech. This approach allows for an ecologically valid analysis of reduction, one that acknowledges the variability and fluidity in natural speech production, without only focusing on a set of more stable, repeatable reduction patterns.

To identify the words involved in the detected reduced sequences, a custom R script was developed to extract all lexical tokens whose temporal boundaries fall within the intervals marked as reduced. The script takes two input files: (i) the manually corrected output file from the reduction detection process, which specifies the time intervals of reduced sequences, and (ii) the token-level (word-level) annotation file provided with the corpus.

The script produces a .csv file listing each instance of reduction along with its corresponding start and end times, as well as the temporal boundaries of each word within that interval. In total, 2061 words were identified as occurring (even partially) within reduced sequences, representing approximately 3.3% of all words produced in the CID dataset.

This extracted dataset forms the basis for subsequent analyses aimed at exploring the distribution of morpho-syntactic categories in reduced speech, along with other qualitative and quantitative investigations. The results of this word-level extraction[2], as well as the results from the detection methods, are presented in Chapter 3.

---

[2] The extracted words and their corresponding counts are provided in the annex.

Chapter 3– Establishing reduction types: Lexicalized and Non-lexicalized reductions

3.1. Introduction

Reduction in spoken language is a multifaceted phenomenon that remains under-explored in several of its less perceptible forms. This chapter aims to lay the groundwork for a more nuanced understanding of reduction by introducing a distinction between two forms of reduction we propose: **lexicalized** and **non-lexicalized** reductions. These categories are defined and contextualized to provide a clear conceptual framework for the chapters that follow.

The impetus for this chapter arises from the observation that many reductions, particularly those that are subtle, difficult to identify, or occur across multi-word sequences, have received limited attention in literature as outlined in the Introduction and State of the Art sections. These forms might often evade detection by both human perception and traditional analytical methods. To address this gap, we adopt a dual-method approach, combining a top-down, linguistically informed analysis with a bottom-up, data-driven detection system. This enables the identification of reductions that are not lexically encoded or phonetically canonical, thereby expanding the scope of investigation beyond the segmental or word level. We focus in particular on examining temporally reduced sequences, which are context-sensitive and gradient in nature, making them especially challenging to detect, and often impossible to identify using a top-down approach. By shifting the analytical lens from isolated, easily recognizable reductions to broader, multi-word units, we aim to capture a richer array of reduction phenomena that more accurately reflect spontaneous conversational speech.

As detailed in the Methodology section (2.2), reduction detection is not a one size fits all process. Different reduction types may require different analytical strategies, and various techniques are capable of capturing different dimensions of reduction phenomena. To substantiate the lexicalized vs. non-lexicalized reduction distinction, we examine reductions identified through both linguistically informed (top-down) analysis and automated (bottom-up) extraction. This is operationalized through a detection script, detailed in the Methodology section (2.2).

The central objective of this chapter is to introduce and define non-lexicalized reductions, while exploring the extent to which their occurrence is influenced by specific linguistic and conversational factors such as morpho-syntactic category, phonemic content, or speaker-specific traits like articulation rate. Our hypothesis posits that lexicalized reductions tend to be more frequent and stable across speakers, whereas non-lexicalized reductions may be more variable and speaker-specific, reflecting idiosyncratic speech patterns.

This study thus aims to contribute not only to a theoretical distinction but also methodological innovation. The classification of reductions into lexicalized and non-lexicalized forms reflects their origins and discourse functions, while the computational tools developed provide a replicable framework for future research. The script, along with its extension for morphosyntactic feature extraction, offers a resource for investigating reduction across different languages and contexts.

Finally, we consider whether the tendency to reduce is itself a speaker-specific trait, whether individuals who frequently produce lexicalized reductions are also more likely to exhibit non-lexicalized ones, or whether these tendencies operate independently due to distinct underlying mechanisms. Addressing this question necessitates moving beyond traditional units of analysis and toward an integrated view of reduction as a dynamic interaction of prosody, syntax, and discourse structure.

The main study, based on the article *Exploring the Nuances of Reduction in Conversational Speech; Lexicalized and Non-lexicalized Reductions* (Bodur et al., *2025*), presents a robust framework for detecting and classifying reductions. In addition to the article-based study, this chapter includes an exploratory analysis of three frequent discourse markers

(*alors*, *enfin*, and *parce que*). This complementary analysis investigates temporal reduction patterns and inter-speaker consistency to explore whether high-frequency leads to more stereotyped distribution of reduction forms.

Together, these analyses argue for a more nuanced and speaker-sensitive account of reduction phenomena, with methodological innovations that can be applied across languages and corpora.

3.2. Article: Exploring the nuances of reduction in conversational speech: Lexicalized and non-lexicalized reductions

This chapter is based on the following journal article:

Bodur, K., Fredouille, C., Rauzy, S., & Meunier, C. (2025). Exploring the Nuances of Reduction in Conversational Speech: Lexicalized and Non-Lexicalized Reductions. In *Speech Communication*. doi: https://doi.org/10.1016/j.specom.2025.103268.
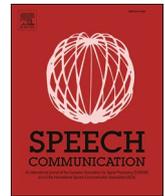
Exploring the nuances of reduction in conversational speech: lexicalized and non-lexicalized reductions

Kübra Bodur [a],[*] , Corinne Fredouille [b], Stéphane Rauzy [a], Christine Meunier [a]

[a] *Aix-Marseille Université, LPL, CNRS, France*
[b] *Avignon Université, LIA, France*

ARTICLE INFO

ABSTRACT

In spoken language, a significant proportion of words are produced with missing or underspecified segments, a phenomenon known as reduction. In this study, we distinguish two types of reductions in spontaneous speech: *lexicalized* reductions, which are well-documented, regularly occurring forms driven primarily by lexical processes, and *non-lexicalized* reductions, which occur irregularly and lack consistent patterns or representations. The latter are inherently more difficult to detect, and existing methods struggle to capture their full range.

We introduce a novel bottom-up approach for detecting potential reductions in French conversational speech, complemented by a top-down method focused on detecting previously known reduced forms. Our bottom-up method targets sequences consisting of at least six phonemes produced within a 230 ms window, identifying temporally condensed segments, indicative of reduction.

Our findings reveal significant variability in reduction patterns across the corpus. Lexicalized reductions displayed relatively stable and consistent ratios, whereas non-lexicalized reductions varied substantially and were strongly influenced by speaker characteristics. Notably, gender had a significant effect on non-lexicalized reductions, with male speakers showing higher reduction ratios, while no such effect was observed for lexicalized reductions. The two reduction types were influenced differently by speaking time and articulation rate. A positive correlation between lexicalized and non-lexicalized reduction ratios suggested speaker-specific tendencies.

Non-lexicalized reductions showed a higher prevalence of certain phonemes and word categories, whereas lexicalized reductions were more closely linked to morpho-syntactic roles. In a focused investigation of selected lexicalized items, we found that "tu sais" was more frequently reduced when functioning as a discourse marker than when used as a pronoun + verb construction. These results support the interpretation that lexicalized reductions are integrated into the mental lexicon, while non-lexicalized reductions are more context-dependent, further supporting the distinction between the two types of reductions.

1. Introduction

Spoken language is a fundamental means of human communication, enabling the exchange of information, ideas and the formation of social connections on a daily basis. Studying spoken language is of utmost importance for understanding human cognition and for developing technological tools that can effectively interact with humans. Despite the growing body of research on speech, the characteristics of speech in spontaneous conversations remain relatively underexplored due to their inherent complexity. Spontaneous speech often includes phenomena specific to casual dialogue, such as feedback signals, repetitions, incomplete utterances, and disfluencies (Prévot et al., 2022). A more comprehensive explanation of the intricate organization and dynamics of these features is essential to deepen our understanding of speech production, as they significantly influence the quality and properties of the resulting speech signal.

Variation in the speech signal is one of the intrinsic features of spoken language (Ernestus, 2000; Johnson, 2004; Meunier and Espesser, 2011). This variation arises as a result of the continuous adaptation of speech to contextual and situational demands (Lindblom, 1990). Speaking in interaction involves permanent adjustments at various stages of speech production, including convergence between interlocutors (Pardo, 2013; Guardiola and Bertrand, 2013). The analysis of large corpora of spontaneous speech, coupled with the use of automated

tools for their exploration, has facilitated the identification of variation patterns such as reduction, an area that offers valuable insights into both 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.

Speech reduction refers to 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 articulation. Indeed, the productions of speakers in spontaneous conversations are typically less clearly articulated compared to formal or scripted contexts (Ernestus and Warner, 2011). Phonetic segments may be weakened or even completely absent in some cases (Schuppler et al., 2011; Niebuhr and Kohler, 2011; Ernestus, 2014), yet interaction remains fluid between the interlocutors. In addition to absent segments, coalescence, where adjacent sounds merge into a single new sound, is also frequently encountered, complicating the identification of individual segments and suggesting that words in conversation are not always produced with clearly distinct segments. These reduced realizations can differ markedly from their counterparts in scripted speech (Prévot et al., 2022). Despite its prevalence, the process and distribution of reduction in spontaneous conversations remain insufficiently characterized, making it difficult to fully grasp how intelligible interaction is maintained when phonetic information is sparse.

Research has shown that speech reduction is influenced by a variety of linguistic and conversational factors. These include semantic and lexical predictability (Aylett and Turk, 2006; Clopper and Pierrehumbert, 2008; Turnbull, 2017), lexical frequency (with high frequency words being more prone to reduction) (Pluymaekers et al., 2005; Bybee et al., 2016), repetitive mention in discourse (Baker and Bradlow, 2009), speaking style (Gendrot et al., 2012), and the phonological properties of segments within reduced sequences (Pharao, 2010; Meunier and Bigi, 2016). Some of these factors are closely related to the lexical structure of language. Individual speaker characteristics also contribute, including linguistic proficiency, speech rate, familiarity with conversational partners, cognitive load, and gender. Higher proficiency is often associated with increased use of reduced forms (Greenberg, 1999), and faster speech rates encourage articulatory efficiency and therefore more reductions (Lindblom, 1990; Fosler-Lussier and Morgan, 1999; Bürki et al., 2011). Familiarity between speakers promotes the use of shared linguistic shortcuts (Fowler, 1988), and increased cognitive load tends to elevate reduction rates as speakers prioritize efficiency (Bell et al., 2009). Gender has also been proposed as a potential factor, with some studies indicating that male speakers may exhibit higher rates of reduction (Byrd, 1994; Bell et al., 1999; Keune et al., 2005). However,

other research has not found statistically significant gender differences (e.g., Binnenpoorte et al., 2005), suggesting that the relationship between gender and reduction may be shaped by additional interacting variables.

Determining the factors that influence speech reduction is a complex endeavor but so is characterizing the various forms that reduction can take. In all languages, speakers are able to recognize certain reduced forms, indicating that these reductions are, to some extent, conscious. For example, the French phrase "je ne sais pas" (*I don't know*), commonly produced as /ʃɛpa/, also appears in its reduced graphical form "chais pas" in text messages. However, the analysis of spontaneous speech, particularly conversational speech, reveals the existence of far less regular and less perceptible reduction.

To better understand this phenomenon, we propose a distinction between two types of reductions found in casual speech, which we refer to as *lexicalized* and *non-lexicalized* reductions. Lexicalized reductions refer to commonly reduced forms or word sequences that have become stabilized through repeated use (Bodur et al., 2022). These forms are well documented in literature and are often easily identifiable by both listeners and transcribers, who typically have mental representations of these reductions (see Fig. 1). In fact, some lexicalized reductions are included in Automatic Speech Recognition (ASR) dictionaries alongside their canonical forms, allowing systems to recognize and transcribe them accurately. Such reductions occur frequently across conversations and speakers, and can be found in many languages. Examples from English and French, based on Ernestus & Warner (2011) include:

- 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 as they lack regularity, reproducibility, or direct association with specific lexical units. These reductions are hypothesized to be influenced by, or to interact with, multiple layers of language and conversation, including prosody, syntax and the organization of interaction. They may span across several words (as shown in Fig. 2), occur only once, and affect any segment or word in the speech stream. Such characteristics contribute to the difficulty of their identification and collection for further analysis. For instance, Fig. 2 shows a temporally reduced sequence, "*qui avait amené*" (*who had brought*) that does not involve particularly frequent words and appears to feature phonemic coalescence, where segments are merged. Non-lexicalized reductions
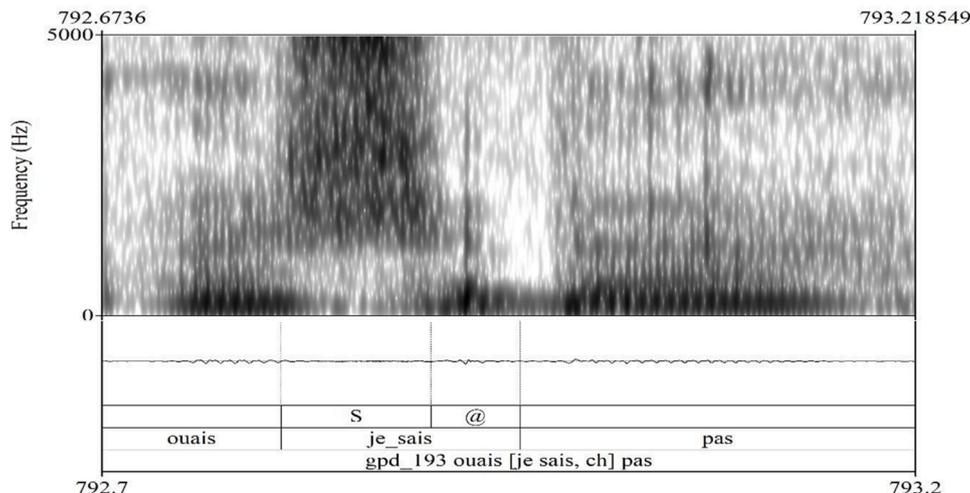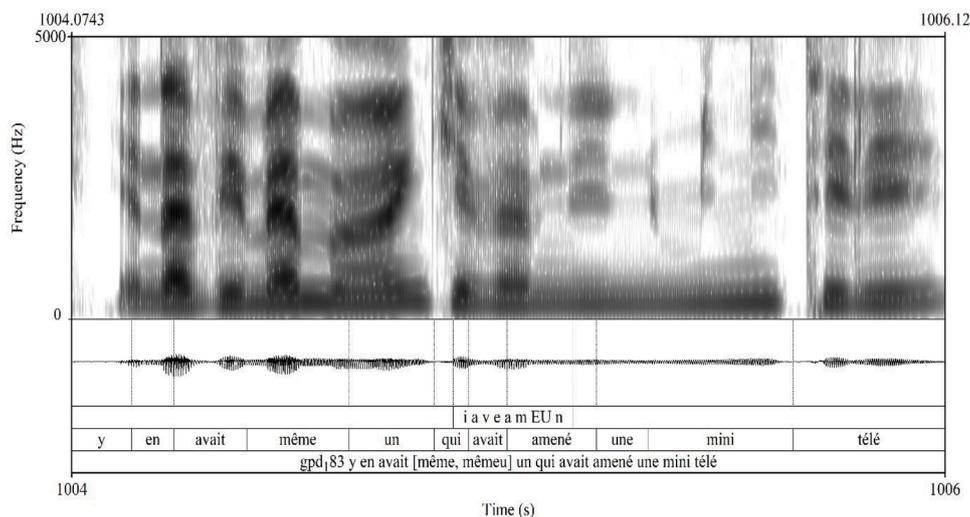

Fig. 1. An example of a lexicalized reduction "je sais pas (/ʃɛpa/, "I don't know")" (top), involved tokens (middle) and the IPU (bottom).

Fig. 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).

have not been as extensively studied as lexicalized ones, largely because the underlying processes and governing principles remain largely unknown. This gives rise to two major challenges: first, developing appropriate methodologies to collect examples of non-lexicalized reductions, which are difficult to detect due to their subtle acoustic properties; and second, determining whether a given sequence is indeed reduced by analyzing its phonetic and acoustic features. Tackling these challenges will require innovative approaches that go beyond current methods, such as top-down strategies that focus primarily on known, frequently reduced lexical items. Complementary methods are needed to detect more elusive reductions and better understand their role in conversational speech.

Previous research on speech reduction has primarily relied on manual detection by human experts or automatic systems such as forced alignment tools (e.g., Adda-Decker and Snoeren, 2011; Wu and Adda-Decker, 2020). However, manual detection through listening poses significant challenges; it is time consuming, and more importantly, subject to perceptual bias, as listeners' interpretations are influenced by their pre-existing linguistic knowledge (Elman and McClelland, 1988). Furthermore, missing phonemes may leave subtle acoustic traces in neighboring segments, complicating the annotation process.

By contrast, recent advancement in the analysis of large corpora of spontaneous speech, along with improved automatic tools that have led to their exploitation, have enhanced the efficiency and accuracy of studying variation and phonetic reduction. Automatic forced alignment, which aligns transcriptions with audio recordings to segment speech into words and phonemes, enables the detection of sequences with extremely short segments -potential indicators of temporal reduction. These reduced segments often result in transcription errors in automated systems, making mismatches between orthographic transcriptions and actual speech more readily identifiable. Despite these methodological improvements, comprehensive analyses encompassing the full range of reduction phenomena in conversational speech remain limited. The present study seeks to bridge this knowledge gap by shedding light on lesser-known reduction types and comparing them with more established forms.

1.1. The current study

This study investigates non-lexicalized reductions (NLRs) and compares them with lexicalized reductions (LRs) to assess whether the two function similarly. Using a corpus of spontaneous conversations between colleagues, it examines reduction patterns in informal, natural speech, which is likely to foster more frequent and varied reductions compared to structured speech genres. To capture a broader spectrum of reductions, the study employs two complementary approaches: one focused on well-documented LRs and the other aimed at identifying less-recognized NLRs.

The central question is whether NLRs and LRs differ in terms of their frequency, distribution, and phonetic characteristics. LRs are typically recognized and expected by both speakers and listeners, while NLRs are mostly opaque, affecting non-predetermined segments and words. We can therefore suggest that the production of these two types of reduction may rely on different underlying mechanisms. We explore this hypothesis by examining two key dimensions.

First, we examine how speaker-related factors—such as speaking time, articulation rate, gender, and overall reduction frequency—affect the occurrence of NLRs and LRs. Specific research questions include: (i) Does longer speaking time correlate with a higher incidence of reductions -especially LRs- given that extended speech offers more opportunities for reduction? (ii) Building on Lindblom's (1963) account of temporal compression and formant undershoot, is a faster articulation rate associated with increased reduction, particularly in the case of NLRs? (iii) Does gender influence reduction patterns differently for LRs and NLRs? and (iv) Are speakers who frequently produce LRs also more likely to produce NLRs?

Second, we explore how linguistic factors —such as phonemic composition, lexical frequency, and morphosyntactic category— influence reduction. We hypothesize that reduction rates vary across morphosyntactic categories. Specifically, LRs are expected to occur more frequently in discourse markers than in verb-pronoun constructions, consistent with prior research on reduction tendencies and lexical frequency. These analyses aim to offer a more nuanced understanding of reduction processes in casual speech.

2. Method & materials

2.1. Dataset

To examine reduction in spontaneous French conversations, we used the Corpus of Interactional Data (CID; Bertrand et al., 2008), which includes eight hours of audio-visual recordings featuring dyadic interactions between colleagues (10 women and 6 men, average age= 34). All participants were native speakers of French, with half originating from southeastern France and the other half from various regions. Each

conversation involved speakers of the same gender who shared a familiar relationship, a context conducive to greater variability and increased speech reduction (Gendrot et al., 2012).

The CID corpus includes phonetic annotations generated through automatic phonetization and alignment tools, designed for processing large volumes of speech data as well as providing consistent outputs for phonetic studies. The segmentation relies on Enriched Orthographic Transcriptions (Bertrand et al., 2006). Two versions of the phonetic annotation (v2 and v3) were available with the corpus. The v3 version contains the same orthographic tokens and phonemes as the v2 but the alignment was performed using SPPAS software (Bigi and Hirst, 2012) with extra manual corrections on the output of the alignment, for phonetization, tokenization and liaison. Both versions were tested for this study and will be further explained in the Bottom-Up Approach section. Lexical and morphosyntactic annotations are also available with this corpus (Bertrand et al., 2006).

2.2. Top-down approach for detecting lexicalized reductions

Frequently encountered lexicalized sequences in French were selected as a starting point, based on documentation in prior research (Adda-Decker and Snoeren, 2011; Wu and Adda-Decker, 2020). To expand this inventory, a small informal questionnaire involving five students was conducted to gather examples of commonly perceived and produced reductions. This inventory, therefore, reflects speakers' representations but is not intended to be exhaustive. It includes a total of 13 items spanning various morphosyntactic categories. The inventory is presented in Table 2.

A top-down, semi-automatic method was used to extract reduced variants of these items. Using Praat (Boersma and Weenink, 2022), target items were first located in the token tier of the CID corpus. Each occurrence was then manually reviewed using both the audio and visual signal. A sequence was labeled as reduced if it showed evidence of segment deletion, syllable omission, or coalescence (merged segments). For each reduced instance, the surrounding context -three tokens before and after- was also extracted. Reduced forms were annotated with phonetic transcriptions (see Fig. 1 for an example).

2.3. Bottom-up approach for detecting non-lexicalized reductions

As top-down methods are inherently limited to reductions that are known and perceptible, a bottom-up approach was adopted to detect reduction prone areas throughout the corpus, with a particular focus on NLRs. While some LR sequences may also be captured this way, the goal was to identify very highly compressed sequences of reduction that often go unnoticed by human listeners or transcribers.

To detect reduction-prone areas, two key parameters were defined: (i) an analysis window large enough to capture potential NLRs, and (ii) a phoneme count threshold within that window, sequences exceeding the threshold were flagged as potentially reduced.

After testing a number of window sizes, a 230 ms window with a six-phoneme threshold was selected. Other window sizes we tested —180 ms and 200 ms (keeping the phoneme count constant) — pointed to more highly reduced areas in the corpus. However, they resulted in a smaller number of detected reductions (see Table 1). The 230 ms window was critical for capturing longer, multi-word sequences, providing a

sufficient number of reductions for reliable analysis. That said, this window size inherently constrains the scope of detectable reductions, as our aim was not to create an exhaustive inventory but to adopt a practical method for identifying less-explored reductions. This approach enables the identification of areas marked by substantial reductions. Therefore, using a larger window size, such as 250 ms, would not effectively detect temporally reduced areas while maintaining the six-phoneme threshold.

2.3.1. Manual evaluation and detection errors

The precision and the quality of the input are of utmost importance for detecting reductions. Therefore, the performance of the automatic detection method was assessed for both v2 and v3 annotation versions. Sequences identified as reduced were manually cross-checked against the audio-visual signals in Praat.

Of the 1212 sequences detected using v3, 256 were marked as errors, yielding an error rate of 21 % (Table 1). In contrast, v2 produced more detections but also a higher error rate (36 %). Most errors stemmed from misalignment issues (e.g., inaccurate marking of phoneme duration or position) or inappropriate liaison (e.g., where the aligner inserted a liaison phoneme per French language conventions, even when the speaker omitted it). Manual evaluations confirmed that v3 offered more reliable results due to corrected segmentations and improved alignment. Consequently, all subsequent analyses were conducted using v3 data.

2.4. Extracting linguistic and speaker-related information

Phonemes, words, and morpho-syntactic categories within non-lexicalized sequences were identified using three custom R scripts developed for targeted data extraction. These scripts processed TextGrid files containing detected sequence intervals along with associated linguistic annotations.

Speaking time for each speaker (measured in seconds) was derived from phonetic and Interpausal Unit (IPU) annotations.

Gender information was obtained from CID metadata (Bertrand et al., 2008). To preserve anonymity, speakers are labeled S1 through S16 in this study, although speaker initials are available in the original corpus and annotations.[1]

Articulation rates, representing phonemes produced per second, were taken from Chardenon et al. (2022). These were calculated based on active speech durations, excluding pauses and laughter, to provide a reliable measure for analyzing temporal patterns in phonetic reduction.

3. Results

This section presents findings on the production of reductions, based on the two detection methods described in Section 2, along with an analysis of linguistic and speaker-related factors.

3.1. Proportions of reduced sequences identified through both approaches

Using the top-down approach, a total of 2566 reduced lexicalized targets were identified out of 3634 occurrences of the 13 target items in the CID corpus, indicating that 71 % of these items were produced in a reduced form. This represents a notably high proportion. The distribution of these reductions is presented in Table 2.

With the bottom-up approach, 956 reductions were correctly

Table 1
Number of automatically detected reductions across different window sizes and phonetic annotation versions alongside the error rates calculated for the sequences detected using the v2 and v3 for the 230 ms window size.

Window size	180ms/6 phon.		200ms/6 phon.		230ms/6 phon.	
Version	v2	v3	v2	v3	**v2**	**v3**
Detected reductions	456	95	933	429	2144	1212
Error Rate					36 %	21 %

---

[1] The corresponding labels used in previous studies on the CID (Bertrand et al., 2008):

Speakers	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
In the CID	AB	AC	AG	AP	BX	CM	EB	IM	LJ	LL	MB	MG	ML	NH	SR	YM

Table 2

Number of occurrences of the 13 selected lexicalized items in sequences detected using top-down and bottom-up approaches, along with their total occurrences in the corpus. The table also includes translations and morpho-syntactic categories.

Lexicalizd Item	Category	Detection			Lexicalizd Item	Category	Detection		
		Bottom-Up	Top-Down	Total			Bottom-Up	Top-Down	Total
alors (then)	adverb	32	132	400	*tu sais* (you know)	pronoun+verb	4	176	248
alors que (whereas)	conjunction	0	7	45	*je sais* (I know)	pronoun+verb	3	328	412
parce que (because)	conjunction	26	579	588	*enfin* (finally)	adverb	3	586	728
je suis (I am)	pronoun+verb	17	128	145	*je peux* (I can)	pronoun+verb	3	28	33
voilà (here, there)	adverb	14	108	277	*de toute façon* (anyway)	adverb	2	53	63
puis (then)	adverb	14	266	430	*puisque* (since, as)	conjunction	2	37	45
quand même (all the same)	adverb	6	138	220	**TOTAL**		126	2566	3634

detected. These reduced sequences represent 2.1 % of the total phonemes (6351 out of 297.703 phonemes) and 2.4 % of the total words (3058 out of 127.713 words) in the corpus. It is important to note that this method does not aim to identify all instances of NLR, but rather to extract those exhibiting the most significant degrees of reduction. Notably, the bottom-up method detected 126 instances of the 13 lexicalized target items, compared to 2566 instances identified by the top-down approach, indicating that the bottom-up method captured 4.9 % of the total reduced lexicalized items.

3.2. Speaker related factors

The distribution of reductions across speakers was highly variable. Speaker S7 produced the highest number of reductions (149 LR and 215 NLR), whereas S10 produced the fewest (68 LR and 7 NLR). While other speakers fall between these two extremes, S7 stands out with more than twice as many NLRs as the next highest speaker, S3.

The minimum observed reduction ratio for lexicalized items was 57 % (S8 and S15), indicating that even speakers who exhibited minimal reduction reduced at least half of their lexicalized targets. The reduction ratio, calculated as the number of reduced target items divided by the total number of occurrences, provides a more meaningful insight into a speaker's reduction patterns than raw counts alone (see Table 3 (LRR %) and Fig. 3). For example, S7 achieved a reduction ratio of 81.4 % (149 out of 183 target items), whereas S8, despite having the lowest reduction rate (57.2 %), reduced 139 out of 243 targets. Furthermore, variation in speaking time among speakers allows for significant reduction rates even with fewer target items, as seen with S10, who achieved a 73.9 % reduction rate. Conversely, S4, who produced the most reduced items overall (233), had a reduction rate of 74.4 %. These patterns indicate that item frequency in the corpus does not necessarily correspond to reduction ratios.

The proportion of reduced segments within NLR sequences for each speaker, as presented in Table 3, also highlights notable inter-speaker differences, while accounting for variations in individual speaking durations. Notably, the coefficient of variation showed that the presence of NLR varied more substantially across speakers (62.1 %) than lexicalized reduction (11.4 %).

Additionally, Pearson's correlation revealed a moderate but statistically significant relationship ($r = 0.60$, $p = 0.013$) between speakers' LR and NLR ratios, as shown in Table 3 and Fig. 3.

Given these speaker-level differences, three potential contributing factors were explored for the two types of reduction: speaking time, articulation rate and gender. Although a single, fully specified mixed-effects model capturing all relevant interactions would have been theoretically desirable, attempts to fit such models failed to converge,

likely due to the relatively small sample size (16 speakers) and model complexity. To address this limitation, we employed a Bayesian multi-level regression approach using the *brms* package. This framework enabled us to account for individual variability through random effects while benefiting from more robust uncertainty estimation (see for example Fong et al., 2010). As such, it allowed us to examine the influence of key predictors on reduction ratios without overfitting or compromising interpretability.[2]

3.2.1. Speaking time

Speaking times varied significantly among speakers,[3] with a coefficient of variation of 19.9 %. The speaker who produced the highest number of NLRs and had the highest LR ratio (S7, with 215 NLRs (5.5 % of his total phonemes) and an 81 % LR ratio) did not have the longest speech duration. Conversely, the speaker with the shortest speaking time (S10) recorded the fewest NLRs (7), but not the lowest proportion of reduced phonemes (0.6 %). This speaker also had 68 LR detections, yielding an LR ratio of 74 %. The speaker with the longest speaking time (S11) did not exhibit markedly different reduction patterns compared to other speakers.

The Bayesian regression model did not reveal strong or consistent effects of speaking duration on either NLR or LR when using ratio-based outcome variables. Specifically, the estimated effect of speaking duration on NLR was slightly negative (Estimate= -0.29), with a 95 % credible interval including zero ([-0.72, 0.15]), suggesting a lack of robust evidence for an association. A similar pattern emerged for LR (Estimate= -0.29; 95 % CI: [-0.74, 0.17]), indicating no statistically significant effect. The model summary is presented in Table 4.

A Pearson correlation analysis showed a statistically significant positive correlation between speech duration and the number of reductions detected using the top-down approach ($r = 0.59$, $p = 0.015$), indicating that longer speaking times are associated with more LR. In contrast, the correlation between speech duration and the NLR was weak and statistically insignificant ($r = -0.18$, $p = 0.48$), suggesting that NLR are relatively independent of speaking time.

3.2.2. Articulation rate

The speaker with the highest articulation rate (S7, 14.4 phonemes/

---

[2] The model used to test influencing factors is: brm(mvbind(NLR, LR) ~ speech + rate + gender + (1 | Speaker).

[3] Speaking times were calculated based on the IPU durations, which depend on the number of phonemes. The correlations were computed based on these durations. Readers can refer to the total phoneme numbers presented in Table 3 as an approximation for speaking times.

Fig. 3. Reduction ratios per speaker, ordered by LR ratio (divided by 10 to enable comparison on the same scale), are presented. The LR ratio is the number of reduced lexicalized targets divided by the total number of lexicalized targets, while the NLR ratio reflects the proportion of reduced phonemes to the total phonemes produced. Dotted lines represent the trendlines for LR and NLR ratios.

Table 4
Summary of Bayesian multilevel model estimates for NLR and LR as predicted by articulation rate, speech duration, and gender. The table includes fixed effects and distributional parameters. Estimates are presented with standard errors and 95 % credible intervals.

Parameter	Estimate	Est. Error	95 % CI	Rhat
NLR_Intercept	−11.62	2.89	[−17.19, −6.33]	1.02
LR_Intercept	1.66	3.04	[−4.54, 7.64]	1.01
NLR_articulation	1.07	0.23	[0.65, 1.52]	1.02
NLR_duration	−0.29	0.23	[−0.72, 0.15]	1.00
NLR_genderM	0.90	0.45	[0.00, 1.80]	1.01
LR_articulation	0.43	0.24	[−0.05, 0.93]	1.01
LR_duration	−0.29	0.22	[−0.74, 0.17]	1.01
LR_genderM	0.24	0.43	[−0.67, 1.10]	1.01
sigma_NLR	0.57	0.26	[0.15, 1.14]	1.04
sigma_LR	0.55	0.28	[0.08, 1.13]	1.01

phonemes, while "de toute façon" which contains 10 phonemes in its canonical form, was often encountered as [tfsɔ̃]. The phonetic and phonological properties of the canonical forms of LRs appear to influence the extent of reduction. For instance, although "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 during reduction. Furthermore, Fig. 5 illustrates that the number of phonemes and words in the target item shows no discernible relationship with the rate of reduction. The frequency of LR reduction was particularly influenced by schwa deletion, which often led to assimilation ("je suis" produced "chuis"; "je sais" produced "chais").

3.3.2. Lexical factors

3.3.2.1. Non-lexicalized reductions. The bottom-up approach extracted a list of 716 unique word tokens from various morphosyntactic
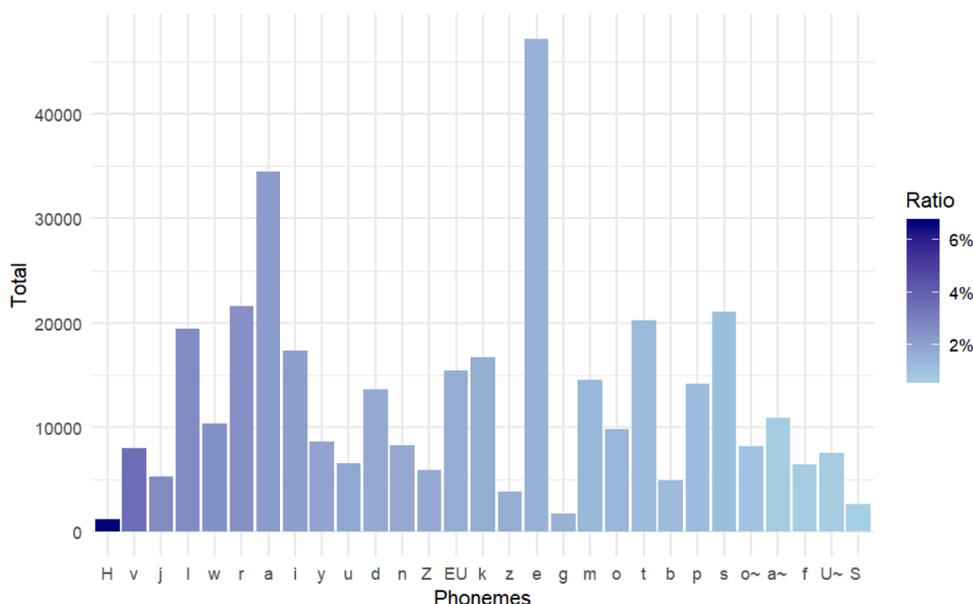
Fig. 4. Total number in the corpus (y axis) and the ratios (color intensity) for the reduced phonemes found in the non-lexicalized sequences calculated based on their total occurrence in the corpus.
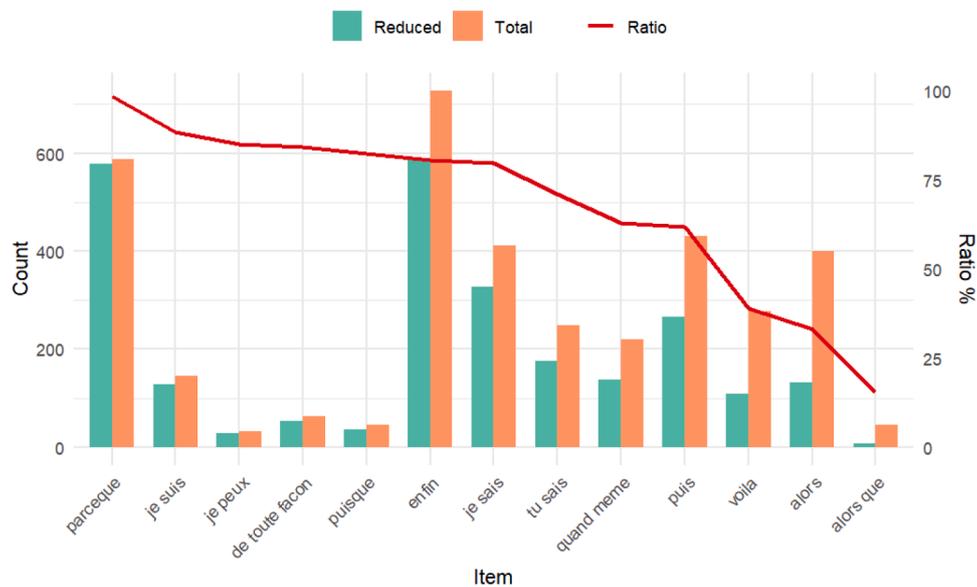
Fig. 5. Reduced (green) and total (orange) occurrences for each of the 13 selected items. The red line represents the ratio of reduction for each target item.

categories, occurring a total of 3058 times in reduced sequences, representing 2.4 % of the corpus' 127.713 total tokens.

The words identified within the reduced sequences demonstrated considerable lexical variability. Pronouns and verbs were among the most frequently occurring categories. However, when considering their reduction ratios relative to all corpus tokens, lower reduction rates were observed for these categories. For instance, the most frequent words "*je*" (5.7 %), "*tu*" (3.4 %), "*de*" (3.7 %), "*il*" (4.1 %), "*la*" (4.9 %), and "*que*" (5.1 %) exhibited relatively modest reduction rates.

In contrast, 68 words from different categories (including 30 nouns, 25 verbs, 4 adjectives and 9 disfluencies) were reduced in 100 % of their occurrences, though some appeared only once in the corpus. Examples include "*fourrure* (fur)", "*orthophoniste* (speech therapist)", "*signalisations* (signalings)", and "*SNCF* (French Railways)", among others.

An analysis of the number of words in the detected reduced sequences revealed that a reduced sequence could range from 1 to 10 words (Table 5). Typically, these sequences consisted of 2 to 4 words. Sequences containing >6 words were less common, accounting for <2 % of the detected sequences.

3.3.2.2. Morphosyntactic categories of the words found in the detected sequences. The analysis of morphosyntactic categories within NLR sequences revealed that reduction processes affected words across all categories, although certain categories were more prevalent. The 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 susceptibility of morphosyntactic categories to the reduction phenomenon.

We subsequently calculated the proportions of the reduced word categories relative to their overall occurrences in the corpus. The results, detailed in Table 6, showed that verbs (3.9 % of the total number of verbs in the corpus) and pronouns (3.2 % of the total number of pronouns in the corpus) remained the categories most frequently subject to reduction. However, adverbs (3.9 %) and conjunctions (3.4 %) also exhibited relatively high reduction ratios, suggesting that they are similarly prone to reduction compared to other morphosyntactic categories.

3.3.2.3. Lexicalized reductions. Our findings indicate that lexicalized sequences are frequently reduced (71 % of target sequences); though, reduction frequencies vary considerably across items. Fig. 5 presents the distribution of these items along with their respective reduction ratios.

In the CID corpus, the most frequently reduced lexicalized item was "parce que", which was reduced in 98 % of occurrences. Interestingly, "enfin" 'though the most frequently occurring among the 13 selected items-, showed a lower reduction ratio than some less frequent items. Therefore, an item may be rare in the corpus yet display a high proportion of reduction. For example, "puisque" occurred only 45 times but was reduced 82 % of the time. In contrast, "puis", a relatively frequent item, was rarely reduced. These findings suggest that the likelihood of reduction is influenced by intrinsic properties of the lexical item rather than frequency alone.

Table 5
The number of words comprising reduced sequences (left), the number (middle) and the frequency of occurrences in all of the detected reduced sequences (right).

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 6
The morphosyntactic categories of the words found in NLR sequences (bottom-up) and their proportions in the total of the corpus.

Category	Reduced	Total	Ratio
Verb	745	19,058	3.9 %
Pronoun	786	24,141	3.2 %
Auxiliary	89	2737	3.2 %
Noun	323	12,260	2.6 %
Determiner	279	10,334	2.7 %
Preposition	229	7405	3.1 %
Adverb	186	4806	3.9 %
Discourse Marker	161	22,168	0.7 %
Conjunctions	135	4005	3.4 %
Adjectives	93	3557	2.6 %
Disfluency	31	10,583	0.3 %
Interjection	1	103	0.97 %

3.3.2.4. The effect of morphosyntactic categories on the reduction of lexicalized items. To investigate whether lexicalized sequences are more susceptible to reduction in specific morpho-syntactic roles, we selected a subset of three items ("alors," "enfin," and "tu sais"), each of which can serve multiple syntactic functions. For example, "enfin" and "alors" can function as both adverbs and discourse markers, while "tu sais", typically a pronoun + verb construction, can also act 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 exhibit higher reduction rates when functioning as discourse markers compared to when they function as adverbs or verb phrases.

A logistic regression analysis was conducted to investigate the relationship between reduction (reduced vs unreduced) and morphosyntactic function for each item. In the case of "alors", which included 248 data points, none of the morphosyntactic categories exhibited a statistically significant effect on reduction. The coefficients for all categories (adverb, discourse marker, feedback, and disfluency), including the intercept was non-significant. However, "alors que", which had 48 data points, showed a notably low reduction ratio (15 %) and predominantly occurred as a conjunction. It appeared as a disfluency only twice. Here, the disfluency variable had a large but non-significant effect on reduction (coefficient: -15.848, $p = 0.995$), while the intercept was significant (coefficient: -1.717, $p < 0.001$).

For "enfin" (731 data points), no significant effect was found for morphosyntactic function, although reached statistical significance (coefficient: -1.557, $p = 0.991$).

Lastly, for "tu sais", which had 247 data points, the "pronoun+verb" category was significantly less likely to be reduced than when the expression functioned as a discourse marker (coefficient: -1.1722, $p < 0.001$). The intercept was also significant (coefficient: 1.1722, $p < 0.001$).

4. Discussion

This study presented an exploratory analysis of reductions in French conversational speech, aiming to distinguish easily identifiable reductions from those less perceptible to the human ear. We classified reductions into two types: lexicalized reductions (LR), identified via a top-down approach focusing on 13 common French sequences, and non-lexicalized reductions (NLR), detected through a novel bottom-up method. Although challenging, identifying NLRs is essential for understanding the broader spectrum of reduction processes. However, the number of detected reductions using this method is highly dependent on the chosen window size. The 230 ms window was deliberately selected to capture as many multi-word sequences as possible. In exchange, we were able to detect only a limited number of highly reduced zones that lack repetitive representations in the lexicon. Manual evaluations of the script's output revealed that input quality, particularly alignment accuracy, critically affects detection performance. Misalignments could lead to false positives, sequences flagged as reduced despite not meeting our criteria. This is why we observed a considerable improvement in output quality using the v3 (updated alignment that yielded better results (Bigi and Hirst, 2012)) of phonetic annotations, even though it reduced the sample size. One advantage of this bottom-up method lies in its ability to identify reduced forms that may elude manual annotation, which is often subjective and inconsistent across annotators. Automating the process thus enhances objectivity and reproducibility.

Our analysis revealed a heterogeneous distribution of reductions among speakers. While 71 % of the selected lexicalized items were reduced, suggesting a pervasive phenomenon, we observed considerable variability at the individual speaker level, an often overlooked factor in reduction research. Though previous studies have occasionally addressed the speaker aspect, primarily within the context of sociological factors (Labov, 1986; 2001), our findings suggest that even within homogenous socio-cultural groups, significant speaker-related variation persists. A modest correlation was found between LR and NLR ratios,

indicating that some speakers produce more reductions overall. However, a closer examination of individual speakers revealed variability in this pattern, suggesting contrasting tendencies in reduction behavior. One speaker (S7) emerged as a standout, producing 215 NLR sequences and an LR ratio of 81.4 %, whereas the others were below this production rate. In contrast, S8 had the lowest LR ratio (57.2 %, producing 139 reductions out of 243 targets) and the second lowest NLR count (15 sequences). Meanwhile, a speaker could produce relatively few target items (e.g., S10) while still maintaining a considerable LR ratio (73.9 %), even with the lowest number (7) of NLRs. These subtleties highlight how individual speaker traits and language preferences are interconnected, prompting further investigation into the underlying reasons for these differences.

One speaker-related factor we examined was gender. The significant effect of gender on NLR, with male speakers exhibiting more reductions than female speakers, warrants further exploration. This finding aligns with previous research suggesting that male speakers may be more prone to producing non-standard forms and reductions (e.g., Byrd, 1994; Labov, 1990, 2001). One potential explanation for this gender difference lies in the nature of NLR, which are often produced in temporally compressed zones of speech, where articulation is faster and more fluid. It is conceivable that male speakers, who may tend to speak at relatively faster rates than female speakers (e.g., Jacewicz, 2009), might produce more reductions as a result of these faster speech patterns. To investigate this, we conducted a post-hoc Bayesian analysis of the interaction between gender and articulation rate on NLR production. The results were not statistically significant (*Estimate*= *0.46; 95 % CI: [−0.46, 1.46]*), suggesting that the observed gender difference in NLR may not be driven by articulation rate alone. Interestingly, no significant gender differences were observed in LR production, possibly reflecting the more stable distribution of LRs across speakers, regardless of gender. This stability might indicate that LR is less influenced by speaker-specific factors like gender, suggesting a more uniform pattern of reduction in these contexts. It is essential to investigate additional speaker-related variables to further clarify the interaction between gender and reduction types.

Contrary to earlier hypotheses (Byrd, 1994, among others) that male speakers produce more reductions, our analysis found no significant differences between male and female speakers in their production of both LR and NLR. This suggests that gender is not a decisive factor in the variability of reduction rates. One such factor worth exploring is dialect. For instance, recent work by Clopper et al. (2023) highlighted that influence of the dialect on phonetic reduction in English (American Midwest) and concluded that, while dialectal influences may be subtle, they are not negligible. Given that our corpus primarily includes speakers from Southeastern France, the sample size may not be sufficient to assess dialectal influences on reduction patterns in depth. Future studies should consider testing these effects across various French dialects to determine if similar patterns emerge.

The reductions identified through the two methods appear to be influenced by distinct factors, reinforcing the hypothesis that they represent different types of reductions. Specifically, NLRs correlate with higher articulation rates, whereas LRs are relatively more strongly associated with longer speech durations. Initially, the model using ratio-based outcomes did not reveal a significant effect of speech duration on LRs. However, a follow up analysis using raw LR count data provided clearer insights. This post-hoc model, which reflects the cumulative nature of reduction opportunities, revealed a positive and statistically significant association between LR counts and speech duration (*Estimate*= *19.6; 95 % CI: [2.49, 36.68]*). These findings support our hypothesis that extended speaking time increases the likelihood of LRs, as it affords more opportunities for speakers to reduce frequently encountered lexical items. In contrast, a similar analysis with raw NLR counts showed no significant effect of speech duration (*Estimate*= *−2.66; 95 % CI: [−16.89, 11.49]*) suggesting that NLRs are less influenced by overall speech duration and may instead be governed by other linguistic

or conversational dynamics. The weak correlation between LRs and articulation rates further supports this distinction, indicating that LRs are less sensitive to spontaneous contextual factors such as speech rate and sequence duration. This finding aligns with Van Son & Pols' (1990, 1992) observations, which show that high speech rates do not necessarily lead to greater reduction, as some speakers are able to control their articulators in such a way that they can produce non-reduced pronunciations at even faster speech rates (e.g., speaker S14). In contrast, the strong correlation between NLRs and articulation rate is consistent with findings by Fosler-Lussier and Morgan (1999), who reported that higher articulation rates promote phonetic variation and reduction, contributing to the heterogeneity across speakers. This divergence may reflect the nature of LRs as lexicalized, stereotyped forms that are less influenced by phonological and prosodic contexts, whereas NLRs are more sensitive to such factors. While the use of a fixed window size in the detection method may introduce some artifacts, we argue that the observed relationship between articulation rate and temporal reduction reflects a genuine pattern rather than methodological bias. The strength of this correlation supports theoretical expectations that faster articulation leads to greater phonetic reduction, supporting an interpretation of these results as meaningful rather than artifact driven.

The duration-based method (230ms/6 phonemes) indicates that, although lexicalized items are sometimes detected within reduced sequences, it fails to capture all of the manually identified LRs. The limited overlap (4.9 %) between sequences identified through the top-down and bottom-up approaches suggests that many LRs do not consistently present with shorter durations. For instance, the reduced form /ʃepa/ for "je ne sais pas" may lack phonemes, but its duration can vary widely depending on conversational context. Bybee's Exemplar Theory (2002) proposes that the phonetic form of each token is recorded in the exemplar representation, enriched with phonetic and contextual details, allowing for a range of variation for each word or sequence. Frequent expressions like "je ne sais pas" thus accumulate a diverse array of reduced and full variants that speakers can access flexibly depending on situational demands. This perspective implies that reduced forms are not merely phonetic simplifications but complex, context-sensitive representations that reflect the nuances of spontaneous speech. Each occurrence of a reduced form is shaped by its usage in different contexts, leading to the creation of a rich database of variants in the speaker's mental lexicon. This perspective is particularly relevant for expressions like "parce que" and "je suis", where reductions seem not only phonetically predictable but also pragmatically motivated, aiding rapid processing of familiar sequences. These items are among the most frequently reduced in the corpus (e.g., parce que, 98 %; je suis, 88 %), while others like "alors" (33 %) and "voilà" (39 %) show lower reduction ratios. These figures are consistent with Wu and Adda-Decker (2020), who suggested that these reduced sequences often function as discourse markers in informal speech. Although dual-route models might predict stability in high-frequency sequences due to holistic storage and execution, our data reveal variability suggesting that these reduced forms are not merely automated articulations but meaningful, contextually adaptable units. This supports the view that reductions serve both lexical and pragmatic functions. Certain discourse markers exhibit flexible reduction patterns aligned with their discourse role, implying that speakers store reductions as context-dependent exemplars rather than fixed articulatory routines.

Importantly, reduction ratios for lexicalized items do not appear to correlate directly with their frequency in the corpus. For example, the sequence "*je peux*" (I can) has a high reduction ratio (84.8 %), despite only occurring 33 times, whereas "*alors*" (then) occurred 400 times but was reduced in only 132 instances (33 %). By focusing on corpus-specific lexical frequency rather than general frequency norms in French, we ensure contextual relevance. Lexical frequency varies substantially across speakers, topics and settings, so this approach provides a more accurate depiction of how reduction manifests in casual conversations, highlighting unique reduction patterns within everyday speech rather than broad generalizations.

Lexicalized items with the highest reduction ratios (parce que (98,4 %, conjunction), je suis (88 %, pronoun+verb), je peux (84.8 %, pronoun+noun)) may not always be considered as highly informative as they are easily inferable by the context (Bell et al., 2009). The "pronoun+verb" combinations such as "je sais", "je suis" and "je peux" showed higher reduction ratios than adverbs such as "alors" or "voila", with the pronoun often being the reduced component, probably due to their more predictable nature. To assess whether informativeness and morphosyntactic category influence the possibility of reduction, we analyzed a subset of three sequences (alors, enfin, tu sais) that can function either as discourse markers or as standard syntactic elements (adverbs and pronoun + verb constructions). Our analysis found no consistent relationship between morphosyntactic category and likelihood of reduction within this subset, except for "tu sais". The pronoun+verb configuration was significantly less likely to be reduced than the discourse marker usage. This suggests that reduction rates may vary by grammatical function and perceived informativeness, but are still largely lexically governed. This aligns with Schubotz et al. (2015), who proposed that more pragmatic, less compositional forms tend to be reduced more often and that such reductions aid comprehension by highlighting discourse structure. A future comparison between the durations of reduced and non-reduced variants of these items would allow us to draw more conclusions to complement these findings.

On the other hand, NLRs are not inherently tied to lexicalized sequences and may span multiple words and categories. Interestingly, many words with the highest reduction rates occurred only once in the corpus (e.g., orthophoniste, bouquinais), while the most frequently encountered words in reduced sequences often had lower proportions of reduction due to their high overall frequency. These finding challenges previous claims (Pluymaekers et al., 2005; Jurafsky et al., 2001) that frequent words are reduced more often. Rather, our data suggests that such findings may reflect raw frequency rather than intrinsic reducibility. Moreover, Jurafsky and colleagues' (2001) proposition that less informative parts of speech are more frequently reduced seems to be contradicted by our observation that many 100 % reduced words in our corpus were content words that carried significant information in context. That said, caution is warranted, as many of these words appeared only once, limiting our ability to draw more reliable interpretations.

Although verbs (27 % of all detected words) and pronouns (25.7 %) were more frequently represented (Table 6), the reduction ratios for adverbs (3.9 % of all adverbs in the corpus) and conjunctions (3.4 %) were comparably noteworthy. Interestingly, only 0.3 % of disfluent words appeared within reduced sequences. Given that disfluencies are often considered less informative and frequently involve repetitions —conditions thought to facilitate reduction (Fowler, 1988)— this is a surprising result. Nevertheless, it raises the question of whether reductions are more prevalent in utterances that include disfluencies. A previous study (Bodur et al., 2023) investigating the relationship between reduction and the prosodic position found that approximately 37 % of reductions occurred within disfluent units. These findings highlight the need for further investigation to determine whether disfluency itself, or specific morphosyntactic categories, predisposes sequences to reduction.

The length of the detected reduced sequences also varied in terms of the numbers of words involved. Most commonly, detected reduced sequences comprised three to four words, likely reflecting the constraints imposed by the selected detection parameters. Prior research (e.g., Moon and Lindblom, 1994; Baltazani, 2007) has suggested that longer sequences are more susceptible to both temporal and spatial (articulatory) reduction. In our data, sequences consisting of only a single word exhibited substantially lower reduction rates, likely due to this factor. However, extremely long sequences (e.g., eight to ten words) were also rarely influenced by reduction. One possible explanation is that in

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 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.

Articulatory characteristics of phonemes also appear to influence reduction likelihood (Meunier and Bigi, 2021); some phonemes appear more often in reduced sequences than others. Prior studies (e.g., Pluymaekers et al., 2005; Voigt and Schüppert, 2013) have reported a relationship between phoneme frequency and reduction rate. In our corpus (CID), the phonemes most frequently present in reduced sequences were /H/, /j/, /l/, /w/ and /v/, suggesting that the category of approximants (liquids and glides) appears more frequently in the detected sequences. In contrast, fricatives, plosives and nasals occurred less frequently in these sequences. Notably, these phonemes are also among the shortest in duration in French, further supporting the idea that phonetic properties —specifically segmental duration and articulatory ease— are potentially influential factors for reduction (Meunier and Bigi, 2021).

For the selected lexicalized sequences, phonetic compositions showed considerable heterogeneity. While neither the number of phonemes nor the number of constituent words show no discernible relationship with the rate of reduction, nature of the phonemes involved clearly influenced the forms LRs could take. For instance, reductions of "alors" often involved the deletion of the /l/, with merging of the vowels /a/ and /o/ into a prolonged vocalic segment. In the case of "enfin", the fricative /f/ was consistently preserved, with reduction commonly yielding the form as "fin. Conversely, in "parce que", both the /r/ and the schwa were systematically omitted, even in minimally reduced forms. These patterns suggest that the proportion of reduced material may be linked, in part, to the nature of the phonemes making up the target item. Though not systematically analyzed in this study, it is worth noting that "parce que" (which typically contains two schwas) was the most massively reduced item in the corpus. Given that the majority of speakers were from south-eastern France, where schwas are typically preserved, this result was unexpected and merits further investigation.

In conclusion, our examination of preliminary results reveals a striking degree of variability in reduction behavior across speakers, phonemes, and lexical items, encompassing both lexicalized and non-lexicalized sequences. These observations validate our methodological decision to distinguish between two types of reductions. While reduction is often conceptualized as primarily physiologically driven, emerging from efforts to minimize articulatory effort in the absence of cues promoting hyper-articulation (Lindblom, 1990), our results suggest a more nuanced perspective. Certain reduced forms may be integrated into lexical representations rather than arising solely from ephemeral articulatory adjustments. Notably, the production of lexicalized reductions appears largely independent of articulation rate, occurs consistently across speakers, and shows little indication of being purely physiologically motivated. These features suggest that LRs may be stored and retrieved from the mental lexicon as holistic forms, enabling variable pronunciation without strong contextual constraints. In contrast, non-lexicalized reductions are more context-dependent, affecting non-predictable words, spanning multiple word sequences, and strongly influenced by articulation rate, morphosyntactic and prosodic structure, and speaker-specific patterns.

Moreover, our bottom-up detection approach demonstrates significant potential for analyzing reductions in spontaneous speech, despite the inherent limitations of parameter-based methods. Future enhancements could include tailoring detection scripts to individual speaker profiles, particularly by integrating speaker-specific articulation rates. Such refinements could potentially enhance detection precision and reduce reliance on extensive manual validation, paving the way for more robust investigations into the dynamics of reduction in natural

language production. This study aimed to contrast distinct types of reduction and highlight the heterogenous nature of their occurrence. Our findings indicate that reductions arise from a complex interplay of phonological, lexical, prosodic, and contextual factors. This complexity may explain inconsistencies in prior literature concerning the roles of gender, articulation rate, and word frequency. Ultimately, our results contribute to a more nuanced understanding of reduction processes and underscore the potential for methodological innovations in the study of spontaneous speech production.

Data availability

CRediT authorship contribution statement

Kübra Bodur: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Corinne Fredouille:** Writing – review & editing, Methodology, Conceptualization. **Stéphane Rauzy:** Writing – review & editing, Methodology. **Christine Meunier:** Writing – review & editing, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adda-Decker, M., Snoeren, N., 2011. Quantifying temporal speech reduction in French using forced speech alignment. J. Phon. 39, 261–270.

Aylett, M., Turk, A., 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. J. Acoust. Soc. Am. 3048–3058.

Baker, R.E., Bradlow, A.R., 2009. Variability in word duration as a function of probability, speech style, and prosody. Lang. Speech 391–413.

Baltazani, M., 2007. Prosodic rhythm and the status of vowel reduction in Greek. Sel. Pap. Theor. Appl. Linguist. 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. J. Mem. Lang. 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. Trav. Interdiscip. Du Lab. Parole Lang. Aix-En-Provence (TIPA) 25, 25–55.

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priégo-Valverde, B., Rauzy, S., 2008. Le CID - Corpus of Interactional Data - annotation et exploitation multimodale de parole Conversationnelle. Rev. TAL 49 (3), 105–134, 2008.

Bigi, B., Hirst, D., 2012. SPeech Phonetization Alignment and Syllabification (SPPAS): a Tool For the Automatic Analysis of Speech Prosody. 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 Séquences 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).

Bürki, A., Fougeron, C., Gendrot, C., Frauenfelder, U.H., 2011. Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. J. Phon. 39 (3), 279–288.

Bybee, J., 2002. Phonological evidence for exemplar storage of multiword sequences. Stud. Second Lang. Acquis. 24 (2), 215–221.

Bybee, J., File-Muriel, R.J., de Souza, R.N., 2016. Special reduction: a usage-based approach. Lang. Cogn. 421–446.

Byrd, D., 1994. Relations of sex and dialect to reduction. Speech Commun. 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. Noirmoutier, p. 4.

Clopper, C.G., Burdin, R.S., Turnbull, R., 2023. Second dialect acquisition and phonetic vowel reduction in the American Midwest. J. Phon. 99, 101243.

Clopper, C.G., Pierrehumbert, J.B., 2008. Effects of semantic predictability and regional dialect on vowel space reduction. J. Acoust. Soc. Am. 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. Rethink. Reduct. 25–72.

Elman, J.L., McClelland, J.L., 1988. Cognitive penetration of the mechanisms of perception: compensation for coarticulation of lexically restored phonemes. J. Mem. Lang. 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. LOT, Utrecht.

Ernestus, M., Warner, N., 2011. An introduction to reduced pronunciation variants. J. Phon. 253–260.

Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian inference for generalized linear mixed models. Biostatistics 11 (3), 397–412.

Fosler-Lussier, E., Morgan, N., 1999. Effects of speaking rate and word frequency on conversational pronunciations. Speech Commun. 137–158.

Fowler, C.A., 1988. Differential shortening of repeated content words produced in various communicative contexts. Lang. 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).

Greenberg, S., 1999. Speaking in shorthand – a syllable-centric perspective for understanding pronunciation variation. Speech Commun. 29 (2–4), 159–176.

Guardiola, M., Bertrand, R., 2013. Interactional convergence in conversational storytelling: when reported speech is a cue of alignment and/or affiliation. Front. Psychol. 4.

Jacewicz, E., Fox, R.A., O'Neill, C., Salmons, J., 2009. Articulation rate across dialect, age, and gender. Lang. Var. Change 21 (2), 233–256. https://doi.org/10.1017/ S0954394509990093. Jul 1PMID: 20161445; PMCID: PMC2790192.

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. Typol. Stud. Lang. 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 Linguist. Linguist. Theory, 183–223.

Labov, W., 1986. The social origins of sound change. Dialect and Language Variation. Academic Press, pp. 524–541.

Labov, W., 1990. The intersection of sex and social class in the course of linguistic change. Lang. Var. Change 2 (2), 205–254.

Labov, W., 2001. Principles of Linguistic Change: Social Factors. Blackwell Publishers, Oxford.

Lindblom, B., 1963. Spectrographic study of vowel reduction. J. Acoust. Soc. Am. 35 (11), 1773–1781.

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. J. Phon. 271–278.

Meunier, C., Bigi, B., 2021. Variations temporelles des phonèmes en parole conversationnelle: propriétés phonétiques et facteurs lexicaux. Stud. Lingvist. 11, 11–38.

Moon, S.J., Lindblom, B., 1994. Interaction between duration, context, and speaking style in English stressed vowels. J. Acoust. Soc. Am. 96 (1), 40–55.

Niebuhr, O., Kohler, K.J., 2011. Perception of phonetic detail in the identification of highly reduced words. J. Phon. 319–329.

Pardo, J., 2013. Measuring phonetic convergence in speech production. Front. Psychol. 4.

Pharao, N., 2010. Consonant Reduction in Copenhagen Danish. Copenhagen: Unpublished Ph.D. Dissertation. University of Copenhagen.

Prévot, L., Bertrand, R., Blache, P., Meunier, C., Nguyen, N., Pallaud, B., 2022. Etudier la conversation pour mieux comprendre le langage. TIPA. Trav. Interdiscip. Sur Parole Lang. (38).

Pluymaekers, M., Ernestus, M., Baayen, R.H., 2005. Lexical frequency and acoustic reduction in spoken Dutch. J. Acoust. Soc. Am. 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. J. Phon. 96–109.

Turnbull, R., 2017. The role of predictability in intonational variability. Lang. 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. J. Acoust. Soc. Am. 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. J. Acoust. Soc. Am. 92, 121–127.

Voigt, S., Schüppert, A., 2013. Articulation rate and syllable reduction in Spanish and Portuguese. Phon. Eur. 317–332.

Wu, Y., Adda-Decker, M., 2020. Réduction temporelle en français spontané : où se cache-t-elle ? Une étude des segments, des mots et séquences de mots fréquemment réduits. In Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Journées d'Études sur la Parole, 1. ATALA et AFCP, Nancy, France, pp. 627–635.

3.3. Complementary analyses

3.3.1. Reduction patterns in three frequent items

To extend the main findings and test assumptions about frequency and reduction stereotypy, we conducted an additional analysis on a subset of three high frequency items that could also function as discourse markers: *alors (then)*, *enfin (finally)*, and *parce que (because).* These items were selected for their prevalence in the corpus and their relatively frequent occurrence in reduced forms. Their distribution in the CID is shown in Table 4.

Items	alors	enfin	parce que
Reduced	132	586	579
Total	400	728	588
Reduction Ratio	33%	80%	98%

Table 4. The total and reduced distributions of the three frequent lexicalized items: *alors, enfin* and *parce que* as well as their reduction ratios within the CID.

3.3.1.1. Reduction rates across speakers: *alors, enfin, parce que*

As presented in the main study, reduction patterns revealed notable inter-speaker variability. Regarding this subset of lexicalized items, *parce que* showed the highest and most consistent reduction ratios: 12 of the 16 speakers reduced it in 100% instances. In contrast, *enfin* displayed more uneven reduction patterns across speakers, while *alors* was the least reduced, with extreme speaker-specific variation (from 0% to 67%). It is also worth noting (although no sociophonetic analyses were conducted on this point) that *parce que* emerged as the most consistently reduced item in our corpus, despite containing two schwa vowels. While Southern varieties of French (half of the speakers in the CID corpus are from these regions) are often characterized by greater schwa preservation

compared to Northern varieties (Durand et al., 1987), this example shows that high-frequency expressions such as *parce que* can still undergo extreme reduction, suggesting that frequency effects can override regional tendencies toward schwa retention. Figure 6 presents reduction ratios for these three items across the speakers of the CID, visually confirming that *parce que* is reduced by nearly all speakers almost categorically, while *enfin* shows more variation, and *alors* is the least reduced and most speaker-dependent in its realization.
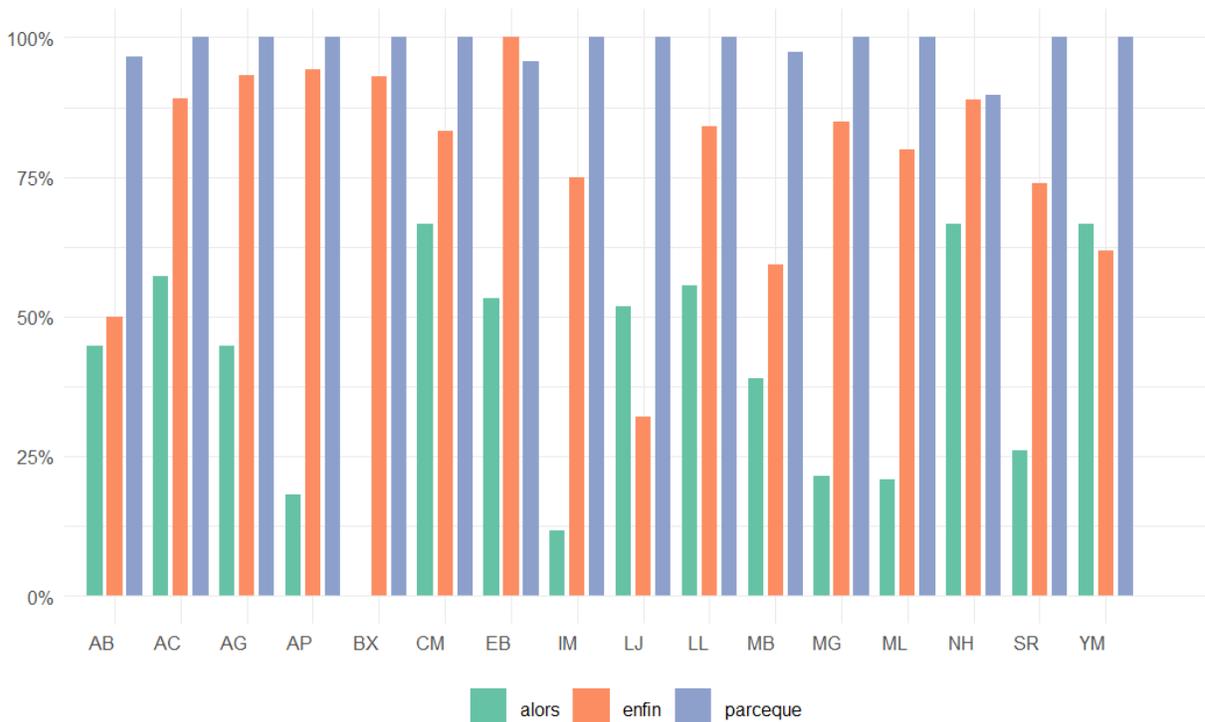
Figure 6. Distribution of the three frequent items across the speakers of the CID.

3.3.1.2. Duration-based evidence of phonetic material reduction: *alors, enfin, parce que*

Using a custom Praat script to extract item durations, we compared the durations of reduced and non-reduced occurrences of these three lexicalized items. The mean durations for their reduced and non-reduced occurrences are represented in Figure 7. This analysis revealed that reduced forms were, on average 36% to 38%, shorter than their non-reduced counterparts. A t-test revealed that the difference in reduced and non-reduced duration reached statistical significance only for *parce que* (p= 0.02).
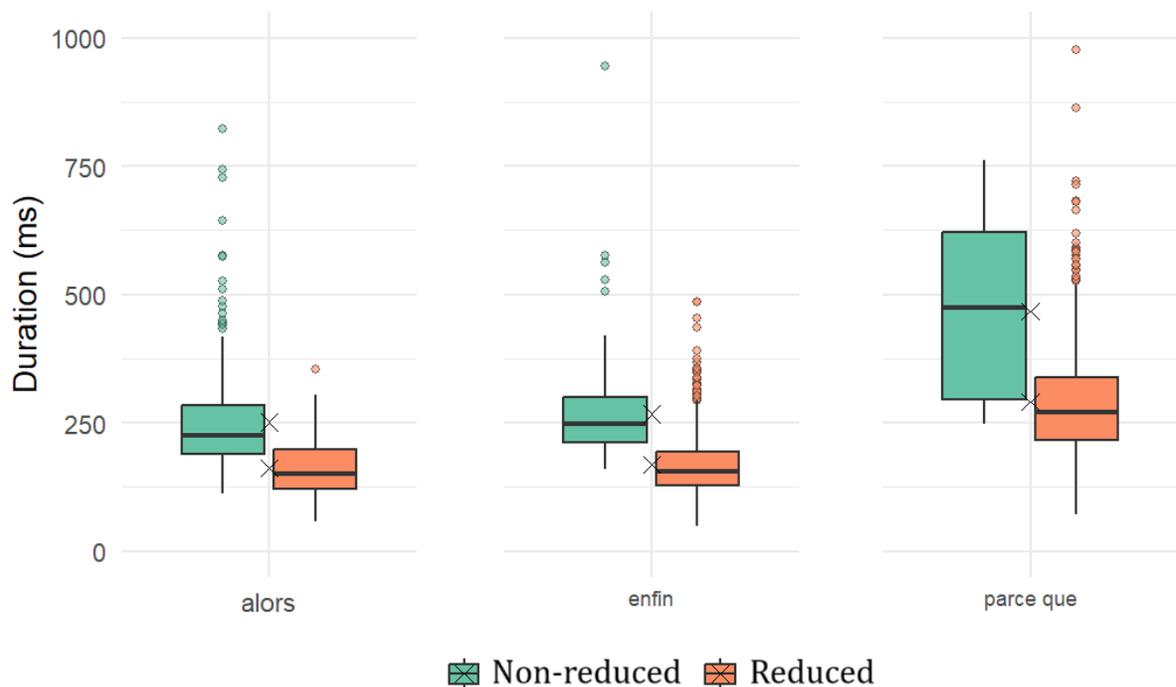
Figure 7. Mean durations (in seconds) of the three most frequent items under reduced and non-reduced conditions.

3.3.2. Exploratory analysis: convergence in reduction behavior

Beyond individual and lexical trends, we conducted a preliminary exploration of whether reduction behavior is shaped by interactional alignment between speakers, drawing on the theoretical framework of phonetic convergence (Pardo, 2013; 2006). Phonetic convergence refers to the phenomenon whereby interlocutors' speech patterns become more similar over the course of an interaction, often interpreted as a marker of social affiliation or coordination. While convergence has typically been studied through segmental or prosodic similarity, it could also be assumed to manifest through reduction behavior, particularly in spontaneous dialogue.

To investigate this possibility, we visualized the distribution of non-lexicalized reductions (NLRs) over time for each speaker in a dyad, using a custom Python script. These exploratory plots illustrate when and how frequently each speaker produced reductions during the conversation. The aim was to identify whether speakers showed temporal convergence, such as clustering of reduction events or parallel trajectories of reduction density over the interaction.
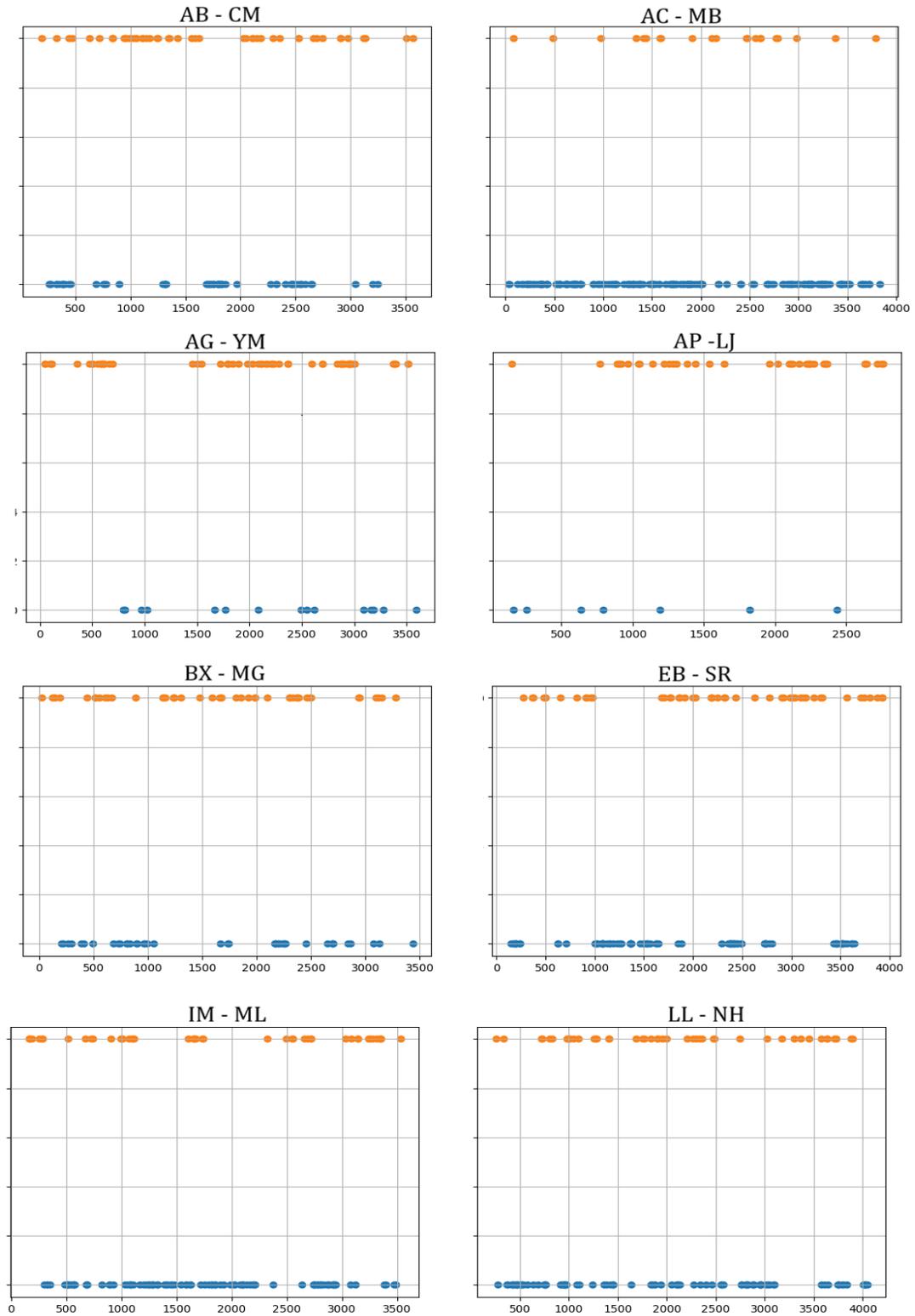
Figure 8. Each panel displays a dyad, with each detected NLR marked by a dot, orange for the first speaker (e.g., AB) and blue for the second (e.g., CM). The x-axis indicates the duration of the conversation.

97

Preliminary observations suggest variability across dyads (Figure 8). In some cases, speakers appeared to align in their reduction timing, producing NLRs in overlapping conversational windows. In other dyads, asymmetries persisted, with one speaker maintaining a more consistent reduction profile regardless of the partner's behavior. These patterns may reflect differing degrees of interactional coordination or sociolinguistic accommodation.

A Pearson correlation analysis revealed a significant negative correlation between interlocutors' NLR rates (per minute) ($r = –0.77$, $p = 0.026$), suggesting that within dyads, speakers who produced more reductions tended to be paired with partners who produced fewer. This preliminary observation contrasts with the expectation of phonetic convergence, where interlocutors' speech patterns become more similar over time, as suggested by Pardo's (2006) model. Instead, the observed divergence may reflect asymmetrical roles, differences in discourse functions, or speaker-specific adaptation strategies during spontaneous conversation. Given the small number of dyads, however, this result should be interpreted with caution. In future work, we plan to segment the conversations into temporal windows to examine whether convergence (or divergence) emerges progressively throughout interaction. This dynamic approach may provide more nuanced insight into how reduction behavior evolves over time and whether convergence is local, global, or absent altogether.

3.3.3. Interpretation

These findings underscore that frequency alone does not guarantee uniformity in reduction. Even among high-frequency items, the degree of reduction and its stability across speakers can vary widely. In the case of *parce que*, the significantly shorter durations in reduced occurrences could be explained by its phonological structure: as a longer sequence, *parce que* inherently contains more phonetic material, offering greater potential for reduction. The magnitude of duration difference may therefore reflect both its frequent use in reduced forms and the larger amount of phonetic substance available for compression. Nevertheless, this interpretation remains preliminary and should be tested on a larger set of lexical items in future work.

In addition to lexical and phonological factors, our exploratory analysis investigated whether speakers tend to align their reduction behavior with that of their interlocutors, an effect predicted by models of phonetic convergence (Pardo, 2006). A correlation analysis of NLR rates per minute between interlocutors revealed a significant negative relationship, suggesting that higher rates of reduction in one speaker were associated with lower rates in their conversational partner. While this may seem counter to convergence-based expectations, it points instead to asymmetrical adaptation patterns or role-based differences in turn-taking dynamics. These preliminary results hint at the complexity of interpersonal alignment in reduction behavior and warrant further investigation. In future work, we plan to segment conversations into temporal windows to analyze whether convergence (or divergence) patterns evolve over time.

Taken together, these complementary findings support the chapter's broader claim that phonetic reduction is shaped by the interaction of multiple factors such as frequency, lexical status, individual behavior, and discourse function—without being uniformly governed by any single variable.

3.4. Conclusion to the chapter

This chapter presented a dual-method exploration of phonetic reduction in French conversational speech, with a focus on distinguishing lexicalized (LR) and non-lexicalized reductions (NLR). By combining top-down and bottom-up approaches, the study offers a novel methodological contribution, enabling the identification of both frequent, predictable reductions and more context-sensitive, speaker dependent phenomena. The findings confirm that LRs are relatively stable across speakers and largely independent of articulation rate, consistent with their lexical integration. In contrast, NLRs exhibited greater variability, closely tied to articulation rate and individual speaker characteristics, such as gender. Importantly, the main study challenges established assumptions about the role of frequency and informativeness in predicting reduction. It also brings to light the influence of phonemic, morphosyntactic, and contextual factors, offering a more fine-grained perspective on how reduction operates beyond simple articulatory economy. The methodological distinction between

LRs and NLRs, supported by their limited overlap, represents a key contribution, demonstrating the need for differentiated analytical strategies. Moreover, the successful use of bottom-up detection method illustrates its potential for capturing subtle reductions often missed in manual annotation, despite sensitivity to parameter tuning.

The complementary analysis further enriches this typology by showing that even among high-frequency discourse markers, reduction is not uniformly patterned. Duration-based comparisons on three lexicalized items confirmed that lexicalized reduction can manifest without significant temporal compression. The significant inverse correlation observed between interlocutors' reduction rates suggests that interactional roles or speaker-specific strategies may mediate reduction alignment, inviting a closer examination of dynamic patterns over the course of dialogue.

Overall, this chapter advances the understanding of phonetic reduction by showing that it is not merely a mechanical outcome of fast speech or lexical frequency but rather a complex process shaped by lexical status, prosodic structure, and communicative context. These insights lay the groundwork for refining detection tools and expanding research into reduction patterns across different dialects and interactional settings.

Building on these contributions, the next chapter shifts the focus from categorical distinctions to continuous articulatory dynamics, examining how articulation rate affects vowel space characteristics and contributes to the occurrence of NLRs. This approach aims to offer a more detailed account of how speech tempo leads to formant undershoot and, ultimately, increases reduction probability, thus deepening our understanding of temporal influences on segmental realization.

Chapter 4– Reduction and vowel space characteristics

4.1. Introduction

This chapter aims to investigate the intricate relationship between vowel space properties and non-lexicalized reductions in spontaneous French speech. The central hypothesis posits that a reduced vowel system, marked by a shift from clearer, more distinct articulatory patterns towards more centralized and less distinct sounds, would correlate with higher rates of non-lexicalized reductions. Specifically, it is hypothesized that speakers with smaller vowel spaces, characterized by formant undershoot and less distinct vowel quality, would exhibit more frequent instances of temporal compression in their speech sequences. These reductions are typically manifested in the form of shorter segment durations, which often result in degraded F1 and F2 values, further contributing to the contraction of the vowel space and a reduction in vowel distinctiveness.

The relevance of this chapter lies in its potential to clarify the dynamic relationship between articulation rate, vowel space, and reduction patterns in speech. This chapter examines how speech rate affects the size and distinctiveness of the vowel space. In doing so, it offers insights into the articulation of temporally compressed speech. This is crucial for understanding phonetic reduction not as an isolated phenomenon, but as a result of underlying temporal and articulatory dynamics. Articulation rate also serves as a predictor of NLR, emphasizing the broader implications of speech tempo on phonetic reduction. This aligns with theories such as the Hyper- and Hypo-speech (Lindblom, 1990) framework, which propose that faster speech compromises articulatory precision and leads to reduced vowel clarity.

This investigation also highlights the variability of vowel systems within a spontaneous conversational context, using the CID dataset. Given that spontaneous

speech tends to prioritize communicative efficiency and naturalness over precise articulation, it serves as an ideal setting for studying reduction patterns. Moreover, the study aims to shed light on the complex interaction of speaker-specific articulatory habits, suggesting that individual differences contribute significantly to reduction patterns. These insights are crucial for a deeper understanding of phonetic reduction in naturalistic speech settings, and have implications for both theoretical phonetics and practical applications in speech processing.

This chapter presents the primary analysis reported in a peer reviewed conference publication (Bodur et al., 2025), along with additional findings not included in the original paper. These supplementary analyses provide important descriptive and comparative insights into vowel realization in spontaneous speech, offering a more nuanced perspective on how temporal and articulatory factors contribute to reduction. The chapter is therefore organized in three sections: (1) an overview of the vowels extracted from the corpus, (2) the main study as presented in the conference paper, and (3) supplementary analyses that extend and contextualize the core findings. These include an investigation of the Formant Centralization Ratio (FCR; Sapir et al., 2010), proposed as a relevant and complementary measure of reduction, and a comparison of the acoustic metrics used in the main study with those derived from more controlled speech contexts.

4.1.1. Overview of extracted vowels in the CID

Vowels in the CID corpus were extracted using a custom Praat script developed and described by Gendrot & Adda-Decker (2005). For each vowel token, the script extracts formant values at three temporal points: the onset, midpoint, and offset. Only the average of these three values will be used for the analyses in the subsequent sections. A total of 113.296 vowel instances were identified. To eliminate extreme outliers based on duration, only vowels with durations between 30 and 400ms were retained, yielding a filtered dataset of 112.346 vowels. Table 5 represents the number of extracted vowels for each vowel category.

Vowel	e	a	i	ø	o	y	u
Count	38173	28031	13946	12076	7907	6910	5303

Table 5. The distribution of vowel categories detected in the CID dataset.

Lobanov normalization (Lobanov, 1971) was applied to account for speaker-specific anatomical differences in vowel formant frequencies, thereby enabling more meaningful comparisons across speakers. This z-score–based method standardizes formant values within each speaker, reducing variability due to vocal tract size while preserving relative vowel distinctions.

The overall mean F1 across the CID corpus was 322.2 Hz (SD = 33 Hz), and the mean F2 was 1448.7 Hz (SD = 164.2 Hz), reflecting the general acoustic characteristics of the extracted vowel tokens. Table 6 presents the mean F1 and F2 values (in Hz), calculated separately for male and female speakers across each vowel category, to account for gender-based differences in vowel articulation. Table 7 summarizes the mean durations and standard deviations for each vowel category, offering a detailed overview of the temporal distribution of vowels in the corpus. These acoustic measurements form the basis for the vowel space-based metrics used in the main study.

	a	e	i	o	ø	u	y
$F1_{male}$	334	318	314	321	316	317	329
$F2_{male}$	1374	1512	1609	1240	1392	1281	1531
$F1_{female}$	342	319	308	318	314	313	312
$F2_{female}$	1385	1524	1571	1250	1405	1254	1503

Table 6. Mean F1 and F2 values (in Hz) for the extracted vowels, separated by male and female speakers from the CID dataset.

Vowel	a	e	i	u	y	o	ø
Mean duration (ms)	78	80	74	73	62	72	67
Standard deviation (ms)	50	57	52	54	44	48	53

Table 7. Mean durations (in ms) obtained for each vowel category in the CID as well as the standard deviations.

4.2. Article: Speech Reduction in French: The Relationship Between Vowel Space and Articulation Dynamics

This chapter is based on the following conference paper:

Bodur, K., Fredouille, C., & Meunier, C. (2025). Speech reduction in French: The relationship between vowel space and articulation dynamics. In *Proceedings of Interspeech 2025*, 17-21 August 2025, Rotterdam, Netherlands.

Speech Reduction in French: The Relationship Between Vowel Space and Articulation Dynamics

Kübra Bodur[1], *Corinne Fredouille*[2], *Christine Meunier*[1]

[1]Aix-Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[2]Avignon Université, LIA, Avignon, France

kubra.bodur@univ-amu.fr, corinne.fredouille@univ-avignon.fr, christine.meunier@univ-amu

Abstract

Reduction is an inherent characteristic of conversations, reflecting the dynamic adaptability of language. This study examined the link between vowel space and non-lexicalized reductions (NLR) in spontaneous French speech. The hypothesis posited that speakers with smaller vowel spaces —indicating centralized, less distinct vowels— would produce more NLR, defined as temporally compressed speech zones. Results showed that smaller vowel space (pVSA) predicted greater NLR only when articulation rate was considered, highlighting an interaction between spatial and temporal speech dynamics. Articulation rate was the strongest predictor, supporting theories that link faster speech to reduced articulatory precision. Vowel Distinctiveness Index (VDI) did not significantly predict NLR, suggesting it reflects broader systemic patterns rather than local reductions. These findings emphasized the interplay of temporal dynamics and individual articulation strategies in shaping reduction in speech.

Index Terms: speech reduction, vowel space area, vowel distinctiveness, spontaneous speech

1. Introduction

Spoken language is inherently shaped by variability in the speech signal, which arises from the continuous adaptation of speech production to contextual demands [1]. A key manifestation of this variability is speech reduction, where articulatory effort is minimized, often leading to shorter durations and less precise articulation. This process, particularly prominent in conversational speech, affects vowels significantly by altering their spectral and temporal characteristics [2].

Reduction can affect phonemes, syllables, and even entire words, resulting in diminished acoustic and phonetic substance [3] while still allowing a smooth interaction between speakers and listeners. Reduction often manifests as vowel centralization, resulting in smaller vowel space areas and less distinct vowel categories [4; 5]. Such centralization arises from formant undershoot during shorter vowel durations, which limits articulators from reaching vowel-specific acoustic targets. These changes can be observed as compressed F1 x F2 spaces, with deviations from prototypical vowel qualities [6; 7]. Factors such as speaking style, prosody, word frequency, and speech rate amplify these effects [8; 9; 10], making vowel space reduction a multifaceted phenomenon.

In spontaneous speech, reduction is not limited to vowels but occurs across various linguistic levels. Therefore, we propose a distinction between lexicalized and non-lexicalized reductions. Lexicalized reductions, such as regular examples in French like "*je ne sais pas*" [ʒə nə sɛ pa] reduced to [ʃɛpa] or [ʃpa], and "*je suis*" [ʒə sɥi] often produced as [ʃɥi], have been

more extensively studied due to their consistency and reproducibility across speakers and contexts as they involve fixed and predictable forms. Examples in English include "*going to*" pronounced as [ɡɑnə] and "*don't know*" produced as [dənoʊ].

However, a substantial number of reductions do not conform to specific lexical representations and are barely perceptible to listeners. These non-lexicalized reductions involve temporally compressed segments and lack stable or consistent forms across speakers and contexts. They are context-dependent, and significantly influenced by articulation rate, morphosyntactic structure, or prosodic patterns [11]. These reductions frequently obscure individual segments, compressing phonemes into shorter intervals while still enabling listener comprehension through contextual cues. For instance, the non-lexicalized sequence "J'ai l'impression, je crois qu'il" (*I have the impression, I believe that he..*) is produced as a temporally reduced stretch [ʒɛlɛ̃pʁɛsjɔʒɔkʁwaki] where 17 phonemes are compressed into a 537ms window[1], making it difficult to isolate as a sequence of discrete units (Figure 1). Such reductions reflect complex articulatory motor patterns where speakers economize effort while still providing sufficient material for listener comprehension.
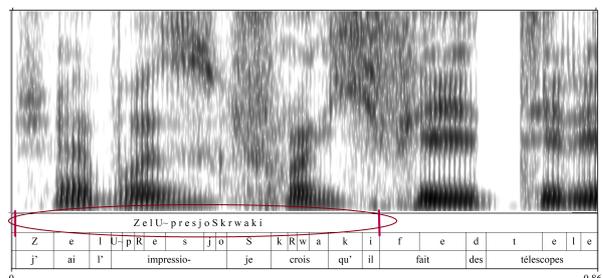
Figure 1: *An example of a non-lexicalized reduction: reduced sequence (top tier), segments corresponding to the sequence "j'ai l'impression je crois qu'il" [ʒɛlɛ̃pʁɛsjɔʒɔkʁwaki], (second), and the involved tokens (bottom). The sequence is made up of 17 phonemes produced in 537ms.*

This study examines the relationship between vowel space properties and non-lexicalized reductions. Specifically, it explores whether speakers with smaller vowel space areas (measured by pVSA) and lower vowel distinctiveness (measured as VDI) —both indicative of greater reduction— are more prone to produce non-lexicalized reductions characterized by temporally compressed segments, where vowel shortening is hypothesized to cause formant undershoot and reduced vowel spaces

---

[1]This sequence, along with other audio-visual examples of reductions, will be available in a repository upon acceptance.

[4]. Articulation rate, as a potential modulator of both pVSA and VDI, is also considered. By investigating these metrics, this study seeks to clarify the interplay between vowel reduction and non-lexicalized speech patterns, offering insights into how individual variability in vowel systems shapes broader phonetic and linguistic processes. Additionally, the bottom-up approach for detecting perceptually subtle, temporally compressed reductions helps expand the scope of reduction analysis.

2. Method and Materials

2.1. Dataset

The Corpus of Interactional Data[2] [12] was used to extract vowels and reductions. The dataset includes eight one-hour audio-visual recordings of conversations between same-gender pairs of colleagues (10 females, 6 males; $M = 34$ years), totaling eight hours of interaction. All participants were native French speakers. The corpus includes comprehensive multimodal annotations across linguistic levels, such as gestures, prosody, syntax, and phonetics. The aligner used for segmentation identifies only seven macro-classes of oral vowels, without distinguishing between mid-closed and mid-open variants. These macro-classes are represented as a, e [e, ɛ], o [o, ɔ], ø [ø, œ, ə], u, y, and i,[3], reflected in the corpus' phonetic annotations. For this study, we used an improved version (v3) of the phonetic annotations. This version includes better phoneme alignment with the help of SPPAS [13] and additional manual corrections.

2.2. Vowel Extraction

Our analysis focuses exclusively on French oral vowels, excluding nasal vowels. Vowel properties of the CID were extracted using the Praat script described in [9]. For each vowel, the script computed formant values at three points: start, middle, and end. Only the mean formant values across these points were used for the analysis. The script identified 113,296 oral vowels in the corpus. Following the application of a duration filter to retain only vowels with durations between 30 and 400ms (as such durations are unlikely to reflect meaningful linguistic reductions) a subset of 112,346 vowels were selected for analysis, of which 6373 were found in the detected reduced sequences.

Lobanov normalization, performed with the *norm.Lobanov* function from the *"vowels"* package [14], was applied to control for speaker variability and enable consistent inter-speaker comparisons. This method, chosen for its sensitivity to individual differences, ensures reliable metrics derived from normalized acoustic values.

2.3. Acoustic Metrics

pVSA. Vowel Space Area is a phonetic measure that quantifies the articulatory space occupied by vowels produced by a speaker. It reflects the dispersion and distinctiveness of vowel sounds, providing insights into a speaker's articulatory and phonetic characteristics. Vowel space area might be influenced by factors such as speaker anatomy [15], language and dialect [16], speaking style [17], age, gender [18], and speech rate [19]. The Pentagonal Vowel Space Area (pVSA) was computed based on five corner vowels /a, e, i, u, o/ (after the duration filter, a to-

tal of 93.361 corner vowels were identified in the CID). The detailed formula[4] used to compute the pVSA can be found in [20][5]. Smaller pVSA values indicate reduced vowel spaces, which we hypothesize to be correlated with higher rates of non-lexicalized reductions.

$$pVSA = \frac{1}{2} \sum_{i=v} (F1_i F2_{i+1} - F1_{i+1} F2_i)$$

VDI. Vowel distinctiveness index refers to the clarity and separability of vowel sounds, influenced by formant frequencies, vowel duration, and coarticulatory effects. While acoustic cues are essential, speaker-related factors such as dialect, accent and speech rate contribute substantially to distinctiveness. The VDI as described by [22] is based on the ratio of the overall vowel space to the dispersion of each vowel category, where dispersion is measured in terms of mean Euclidean distances between the vowel/system centers and their periphery. It reflects the relationship between vowel space size and category dispersion, with higher values indicating hyper-articulated speech and lower values suggesting hypo-articulation. The VDI scores of speakers were computed following the formula[6] detailed in [23].

$$\frac{D\_SYS}{d\_VOY} = \frac{\sqrt{(F1v - F1S)^2 + (F2v - F2S)^2}}{\sqrt{(F1v - F1V)^2 + (F2v - F2V)^2}}$$

Articulation Rate. Articulation rates, calculated as the number of phonemes produced per second (excluding pauses), were extracted by Chardenon and colleagues [24] for the speakers in CID.

2.4. Detecting Reductions

Temporally and phonemically reduced zones were identified using a bottom-up detection approach. A custom script [7] detected sequences containing at least 6 phonemes in a 230ms sliding window. Processing phoneme alignment files with temporal boundaries, the script counts phonemes from each onset. Once at least six phonemes are identified within the window, the sequence is recorded by shifting one phoneme, and the script goes on to search for the following 230ms window. If the threshold (6 phonemes) is exceeded, the window expands to include additional phonemes until the phoneme count in the sliding 230ms window falls below the threshold. The 230 ms window was selected after testing smaller sizes (e.g., 180ms, 200ms), which failed to provide a sufficient number of sequences for a reliable analysis.

A total of 956 highly reduced sequences were identified, corresponding to 2.1% of all phonemes in the CID (6,351 phonemes out of 297,703). Speaker reduction ratios (NLR%) were calculated as the proportion of phonemes in reduced sequences relative to the total phonemes produced by each speaker.

---

[2]The corpus, along with annotations, is freely accessible under open licenses and is available on the "Ortolang" repository: https://hdl.handle.net/11403/ortolang-000722/v1

[3]Detailed information regarding phonetic annotation and automatic alignment can be found in [12].

[4]The index i represents the position of a vowel in the ordered sequence of vowel categories used to define the vocalic pentagon.

[5]*convexHullArea* and *VowelMeansPolygonArea* functions from "*phonR*" package [21] were also computed to provide information regarding vowel space area. Both provided compatible results with our pVSA values.

[6]'v' refers to vowels (exemplar), 'V' refers to a vowel category while 'S' refers to the vowel system.

[7]To preserve anonymity, the detection script's source code will be available on GitHub upon acceptance.

Group	Male						Female									
Speaker	AG	AP	EB	LJ	SR	YM	AB	AC	BX	CM	IM	LL	MB	MG	ML	NH
pVSA	2722	5976	1561	4715	3742	1190	6430	2846	3550	2665	6486	5724	7178	9337	4321	7810
VDI	1,15	1,17	0,85	1,11	1,3	0,8	1,22	0,90	0,73	0,90	0,98	0,60	1,07	1,02	0,99	0,78
NLR	94	87	215	45	20	45	41	79	64	46	15	7	47	48	63	41
NLR%	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
Art.Rate	13,1	13,26	14,4	11,45	11,19	12,17	11,9	13,08	13,42	12,16	11,87	11,08	12,37	13,18	12,33	13,31

Table 1: *Values for individual speakers, including pentagonal vowel space area (pVSA), vowel distinctiveness index (VDI), counts and ratios of non-lexicalized reduction sequences (NLR and NLR%), and articulation rates measured in phonemes per second.*

3. Results

This section presents the pVSA and VDI values, followed by an analysis of the interaction between vowel space metrics and speakers' reductions. Individual values of pVSA, VDI and non-lexicalized reductions ratios (NLR%) are reported in Table 1.

The mean pVSA value was 5635 ($SD = 2233, range : 2665 - 9337$) for female speakers and 3318 ($SD = 1853, range : 1190 - 5976$) for male speakers. The mean VDI was 0.92 ($SD = 0.18, range : 0.6 - 1.2$) for female speakers and 1.05 ($SD = 0.2, range : 0.8 - 1.3$) for male speakers. While no significant difference was found in VDI values between genders ($Welch's t-test : t(9.66) = -1.30, p = 0.222$), pVSA values were significantly higher for female speakers ($t(12.31) = 2.24, p = 0.044$).

A Bayesian linear regression model was fitted to investigate the relationship between NLR and two predictor variables *pVSA* and *VDI*. The model also included random intercepts based on the variable gender to account for possible inter-speaker variation that might arise from gender differences in reduction patterns ($NLR \sim pVSA + VDI + (1|gender)$). Detailed results are presented in Table 2.

Parameter	Est	SE	95% CI	ESS	Rhat
Intercept	2.08	0.81	[0.26, 3.95]	288	1.02
pVSA	-0.63	0.41	[-1.49, 0.16]	1417	1.00
VDI	0.06	0.39	[-0.71, 0.87]	1428	1.00
sd(Intercept) (gender)	1.09	1.09	[0.03, 3.99]	658	1.00
sigma	1.33	0.29	[0.89, 2.03]	1403	1.00

Table 2: *Summary of Model 1 showing estimates (EST), standard errors (SE), 95% credible intervals, effective sample size (ESS: Bulk), and convergence diagnostic (Rhat). Includes multilevel hyperparameters for gender and residual standard deviation.*

The standard deviation of the random intercept for gender was 1.09, indicating moderate variability in NLR between the two gender groups. The fixed-effect intercept, which represents the average NLR when predictors are centered and gender variability is accounted for, was estimated at 2.08 ($Est.Error = 0.81$). The pVSA coefficient showed a negative association with NLR ($Estimate = -0.63$), though the credible interval included zero, suggesting the effect is not statistically significant. The VDI parameter showed a small, positive, but non significant effect ($Estimate = 0.06$). The residual variability (sigma) was estimated at 1.33, suggesting additional unexplained variation in NLR. All parameters exhibited good convergence ($\hat{R}$ values close to 1) and sufficient effective sample sizes ($Bulk\_ESS and Tail\_ESS > 1000$), ensuring robust posterior estimates. These results indicate that the predictors pVSA and VDI did not exhibit strong evidence of NLR.

Given that the primary predictors (pVSA and VDI) did not demonstrate significant effects, a follow-up model was fitted to further investigate these relationships. This revised model introduced articulation rate as another predictor variable, while keeping the random effect of gender constant ($NLR \sim pVSA + VDI + articulation + (1|gender)$), to explore whether articulation rate adds explanatory power to the frequency of NLR.

In Model 2, articulation as a fixed effect significantly improved the model, showing a positive and significant relationship with NLR ($Estimate = 1.12, CI = [0.76, 1.46]$). This suggests that higher articulation rate is strongly linked to increased NLR. A negative association was found between pVSA and NLR (-0.59), suggesting that smaller vowel spaces might correspond to higher NLR values. In contrast, VDI (0.25) had a small and non-significant positive effect, indicating no significant influence on NLR. The Intercept was estimated at -11.86, indicating a low baseline NLR, but this effect is also accompanied by substantial uncertainty ($95\%CI : [-16.35, -7.53]$). The variability due to gender *(sd(Intercept))* (0.79) reflected variability between male and female speakers in NLR. The addition of articulation led to a reduction in the random intercept variability for gender, suggesting that articulation may account for some of the variability previously attributed to gender. Additionally, the residual standard deviation (sigma) decreases from 1.33 in Model 1 to 0.59 in Model 2 (see Table 3), reflecting an improved fit of the model with articulation included.

Parameter	Est	SE	95% CI	ESS	Rhat
Intercept	-11.86	2.27	[-16.35, -7.53]	1178	1.00
pVSA	-0.59	0.19	[-0.98, -0.24]	577	1.01
VDI	0.25	0.18	[-0.13, 0.58]	1253	1.00
Articulation	1.12	0.17	[0.76, 1.46]	1518	1.01
sd(Intercept) (gender)	0.79	0.84	[0.02, 2.89]	328	1.01
sigma	0.59	0.15	[0.39, 0.96]	325	1.02

Table 3: *Summary of Model 2 showing estimates (EST), standard errors (SE), 95% CI, effective sample size (ESS: Bulk) and convergence diagnostic (Rhat) for the relationship between NLR and predictor variables (pVSA, VDI, articulation).*

Finally, a third Bayesian regression model was fitted to incorporate interaction terms between articulation, pVSA, and VDI, to see how articulation might influence the relationship between vowel space area (pVSA) and distinctiveness (VDI) in shaping NLR ($NLR \sim pVSA * articulation + VDI * articulation + (1|gender)$. The articulation variable remained a significant positive predictor of NLR in this model ($Estimate = 1.08, 95\%CI[0.78, 1.37]$). Interestingly, the pVSA showed a significant positive main effect ($Estimate = 4.72$), but its interaction with articulation (pVSA:articulation) has a small but significant negative effect on NLR (-0.41). This suggests that the effect of pVSA on NLR decreases as articulation rate increases. The main effect of VDI was negative but non-significant ($Estimate = -1.50$), and its interaction with articulation (articulation:VDI) was small and non-

significant ($Estimate = 0.15, 95\%CI : [-0.13, 0.42]$). This indicates that VDI contributed only minimally to explaining NLR variation. The intercept was estimated at -11.51 ($95\%CI : -15.53, -7.58$), indicating that when all predictors are at their mean, the baseline NLR is negative. The random effect of gender *(sd(Intercept))* (0.81) indicated some variability across genders, while the residual error *(sigma)* (0.46) suggested good model fit. This value decreased from 0.59 in Model 2 to 0.46 in Model 3, indicating that the addition of interaction terms improved the overall model fit. Overall, this model suggests that articulation is consistently the strongest predictor of NLR. While pVSA has a significant positive effect, its interaction with articulation reduces its impact.

Parameter	Est	SE	95% CI	ESS	Rhat
Intercept	-11.51	1.99	[-15.53, -7.58]	1982	1.00
pVSA	4.72	2.12	[0.41, 8.78]	1641	1.00
Articulation	1.08	0.15	[0.78, 1.37]	2240	1.00
VDI	-1.50	1.64	[-4.83, 1.79]	2033	1.01
pVSA:Articulation	-0.41	0.16	[-0.73, -0.08]	1642	1.00
Articulation:VDI	0.15	0.14	[-0.13, 0.42]	2028	1.01
sd(Intercept) (gender)	0.81	1.00	[0.02, 3.66]	814	1.01
sigma	0.46	0.14	[0.28, 0.81]	697	1.00

Table 4: *Summary of the Model 3 with articulation as an interaction term, summarizing estimates for fixed effects, interaction terms, and random effects predicting NLR. Includes estimates, standard errors (SE), 95% confidence intervals (CI), effective sample size (ESS: Bulk) and Rhat values.*

4. Discussion

This study investigated the relationship between vowel space properties and non-lexicalized reductions, hypothesizing that a reduced vowel system —characterized by a shift from clearer, distinct articulatory patterns towards more centralized and less distinct sounds, reflected in a smaller vowel space— would correspond to higher rates of non-lexicalized reductions. These reductions, identified as temporally compressed sequences of segments within the speech flow, were expected to be more frequent among speakers with smaller, less distinct vowel spaces. This prediction stemmed from the observation that shorter durations —typical of segments in non-lexicalized sequences— tend to result in degraded F1 and F2 values, further contributing to reduced vowel space size and distinctiveness. Additionally, articulation rate was considered as a key factor shaping this relationship, offering a dynamic view of how vowel space properties interact with speech patterns.

To test these hypotheses, the CID dataset was chosen for its familiar and spontaneous conversational context, which is thought to encourage speakers to prioritize communicative efficiency and naturalness over articulatory precision. Such a setting fosters greater speech variability and more frequent instances of reduction compared to read or controlled speech [25]. In this dataset, reduction rates and acoustic metrics revealed notable variability among speakers. Interestingly, some speakers exhibited large vowel systems with low distinctiveness (e.g., NH; $pVSA : 7810, VDI : 0, 8$) while others displayed relatively small vowel spaces with high distinctiveness (e.g., SR; $pVSA : 3742, VDI : 1, 3$). A post-hoc correlation test found no significant relationship between pVSA and VDI values ($p = 0.48$), indicating that vowel systems can vary in structure independently of overall distinctiveness.

Our findings highlight articulation rate as the strongest predictor of NLR across models, emphasizing the role of speech tempo in driving reductions. This aligns with the Hyper- and Hypo-speech framework [1], which posits that increased speech rate compromises articulatory precision, leading to formant undershoot and thus reduction in speech. Consistent with [26] and [27], this study underscores the strong relationship between temporal dynamics and vowel reduction. On the other hand, [28] found that while average vowel space size does not significantly differ between high and low speech rates, slow talkers exhibit greater inter-speaker variability and generally larger vowel spaces. This adds to how individual articulatory behavior —beyond articulation rate alone— may contribute to reduction patterns, underscoring the interaction of speaker-specific factors in shaping reduction phenomena.

Regarding pVSA, our findings demonstrate that smaller vowel space areas predict greater NLR, but only when articulation rate is taken into account in the model. This suggests that the impact of vowel space size on NLR is not static; instead, it interacts with speakers' temporal dynamics to shape reduction patterns. At slower articulation rates, smaller pVSA likely reflects biomechanical constraints, leading to reductions via reduced clarity of vowel articulation. However, at faster rates, reductions are dominated by the temporal compression itself. This finding extends the work of Perkell et al. [29], who argued that vowel space area reflects both cognitive and physical constraints in speech production which are modulated by temporal demands. By showing that pVSA's significance is contingent on articulation, these results extend prior research by highlighting the interaction between spatial and temporal dimensions of vowel production.

VDI, as a measure of vowel system dispersion relative to vowel category dispersion, did not consistently emerge as a significant predictor of NLR. As a theoretically robust index, VDI reflects the ratio of global vowel system variability to local category variability, providing insight into how distinct vowel categories are relative to the overall system. One explanation for its limited predictive power in this study may relate to the scope of NLR as a phenomenon. While NLR might reflect temporal compression and formant undershoot, VDI's theoretical underpinnings may be more sensitive to broader systemic patterns of vowel articulation rather than local reductions. Additionally, its relationship with NLR may depend on unmeasured factors such as phonological environment or individual speaking strategies. These results suggest that vowel distinctiveness might interact with other contextual variables to shape reduction.

Notably, gender effects diminished when articulation rate and vowel space measures were included in the model, suggesting that previously observed gender differences in reduction may be better explained by articulatory dynamics rather than inherent gender-based speech patterns. This aligns with previous research [30; 31], which emphasizes articulation rate and biomechanical constraints over categorical gender effects. This study reinforces the central role of articulation rate in non-lexicalized reductions, with vowel space size influencing reduction only when considered alongside speech tempo. These findings highlight the interaction between spatial and temporal aspects of speech rather than a direct effect of vowel distinctiveness on reduction patterns. Future studies should explore these metrics and their relationships with reduction behavior in larger, more diverse datasets to account for variability across languages, dialects, or speech contexts. Additionally, while this study considered articulation rate, other factors —such as cognitive load, task difficulty, and social influences— may also shape vowel reduction patterns and should be incorporated into future models.

References

[1] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," *Speech Production and Speech Modeling/Kluwer Academic Publishers*, 1990.

[2] M. Nakamura, K. Iwano, and S. Furui, "Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance," *Computer Speech & Language*, vol. 22, no. 2, pp. 171–184, 2008.

[3] C. G. Clopper, R. Turnbull, F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, and M. Zellers, "Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors," *Rethinking reduction*, pp. 25–72, 2018.

[4] B. Lindblom, "Spectrographic study of vowel reduction," *The journal of the Acoustical society of America*, vol. 35, no. 11, pp. 1773–1781, 1963.

[5] E. Flemming, "A phonetically-based model of phonological vowel reduction," *Manuscript. Cambridge: MIT*, 2005.

[6] S. H. Ferguson and D. Kewley-Port, "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," *The Journal of the Acoustical Society of America*, vol. 112, no. 1, pp. 259–271, 2002.

[7] K. Johnson, P. Ladefoged, and M. Lindau, "Individual differences in vowel production," *The Journal of the Acoustical Society of America*, vol. 94, no. 2, pp. 701–714, 1993.

[8] C. Mooshammer and C. Geng, "Acoustic and articulatory manifestations of vowel reduction in German," *Journal of the International Phonetic Association*, vol. 38, no. 2, pp. 117–136, 2008.

[9] C. Gendrot and M. Adda-Decker, "Impact of duration on f1/f2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German." in *INTERSPEECH*, 2005, pp. 2453–2456.

[10] Y. Wu and M. Adda-Decker, "Réduction temporelle en français spontané: ou se cache-t-elle? une étude des segments, des mots et séquences de mots fréquemment réduits." in *Proceedings of JEP-TALN-RECITAL, vol. 1: JEP*, 2020, pp. 627–635.

[11] K. Bodur, R. Bertrand, J. German, S. Rauzy, C. Fredouille, and C. Meunier, "Speech reduction: position within French prosodic structure," in *INTERSPEECH*, 2023, pp. 117–121.

[12] R. Bertrand, P. Blache, R. Espesser, G. Ferré, C. Meunier, B. Priego-Valverde, and S. Rauzy, "Le CID-Corpus of Interactional Data- annotation et exploitation multimodale de parole conversationnelle," *Revue TAL*, vol. 49, no. 3, pp. pp–105, 2008.

[13] B. Bigi and D. Hirst, "SPeech Phonetization Alignment and Syllabification (SPPAS): a tool for the automatic analysis of speech prosody," in *Speech Prosody*, 2012, pp. 19–22.

[14] T. Kendall, E. R. Thomas, and M. T. Kendall, "Package 'vowels'," 2018.

[15] S. Fuchs, R. Winkler, and P. Perrier, "Do speakers' vocal tract geometries shape their articulatory vowel space?" in *ISSP -8th International Seminar on Speech Production*, 2008, pp. 333–336.

[16] C. G. Clopper and J. B. Pierrehumbert, "Effects of semantic predictability and regional dialect on vowel space reduction," *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1682–1688, 2008.

[17] C. Gendrot, M. Adda-Decker, and C. Schmid, "Comparaison de parole journalistique et de parole spontanée: analyses de séquences entre pauses," in *JEP-TALN-RECITAL, vol. 1*, 2012, pp. 649–656.

[18] E. Jacewicz, R. A. Fox, and J. Salmons, "Vowel space areas across dialects and gender," in *ICPhS*, 2007, pp. 1465–1468.

[19] B. Siebenhaar and M. Hahn, "Vowel space, speech rate and language space," in *ICPhS*, 2019, pp. 879–883.

[20] C. Fougeron and N. Audibert, "Testing various metrics for the description of vowel distortion in dysarthria." in *ICPhS*, 2011, pp. 687–690.

[21] D. R. McCloy, "Vowel normalization and plotting with the phonr package," *Technical Reports of the UW Linguistic Phonetics Laboratory*, vol. 1, pp. 1–8, 2012.

[22] K. Huet and B. Harmegnies, "Contribution à la quantification du degré d'organisation des systèmes vocaliques," in *23èmes Journées d'Etude sur la Parole*, 2000, pp. 225–228.

[23] C. Meunier and A. Ghio, "Caractériser la distinctivité du système vocalique des locuteurs," in *32èmes Journées d'Études sur la Parole*, 2018, pp. 469–477.

[24] E. Chardenon, C. Meunier, and C. Fougeron, "Variations de débit articulatoire sur un corpus conversationnel avec différents types d'interactions," in *34ème Journées d'Etudes sur la Parole*, 2022.

[25] F. Torreira, M. Adda-Decker, and M. Ernestus, "The Nijmegen corpus of casual French," *Speech Communication*, vol. 52, no. 3, pp. 201–212, 2010.

[26] A. R. Bradlow, G. M. Torretta, and D. B. Pisoni, "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Communication*, vol. 20, no. 3-4, pp. 255–272, 1996.

[27] K. Johnson, "Massive reduction in conversational American English," in *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium.* Citeseer, 2004, pp. 29–54.

[28] Y.-C. Tsao, G. Weismer, and K. Iqbal, "The effect of intertalker speech rate variation on acoustic vowel space," *The Journal of the Acoustical Society of America*, vol. 119, no. 2, pp. 1074–1082, 2006.

[29] J. Perkell, M. Matthies, H. Lane, F. Guenther, R. Wilhelms-Tricarico, J. Wozniak, and P. Guiod, "Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models," *Speech Communication*, vol. 22, no. 2-3, pp. 227–250, 1997.

[30] A. P. Simpson, "Dynamic consequences of differences in male and female vocal tract dimensions," *The Journal of the Acoustical society of America*, vol. 109, no. 5, pp. 2153–2164, 2001.

[31] D. Byrd, "Relations of sex and dialect to reduction," *Speech Communication*, vol. 15, no. 1-2, pp. 39–54, 1994.

4.3. Complementary analyses: Enriching the understanding of vowel reduction dynamics

While the main study establishes a robust link between articulation rate and vowel space reduction, several additional analyses were conducted during the research project that might offer broader insights into the structure of the CID vowel system and reduction patterns. These complementary analyses, presented below, aim to address limitations of the article format and strengthen the empirical foundation of this chapter's core claims. This section presents (i) visualizations of vowel space areas, (ii) an analysis of the relationship between the Formant Centralization Ratio (FCR) and non-lexicalized reductions, and (iii) a comparison of acoustic metrics across controlled and conversational speech contexts.

4.3.1. Visualizing vowel space areas

To illustrate the variability between the vowel space size and structure across speakers, vowel space plots were generated, first grouped by gender, and then by individual speaker. These visualizations provide an intuitive and informative way to assess articulatory variation, complementing quantitative metrics like the pentagonal Vowel Space Area (pVSA). Plotting vowel space areas enables direct visual comparisons of articulatory range, highlighting patterns that may not be immediately apparent through numerical summaries alone (Bradlow et al., 1996; van Son et al., 2018).

To support the main analysis, vowel space plots were generated using the five French corner vowels /a, e, i, u, o/, which also served as the basis for calculating the pVSA values. Focusing on these corner vowels ensures comparability across speakers and follows established practice for capturing the articulatory boundaries of vowel systems (e.g., Bradlow et al., 1996).
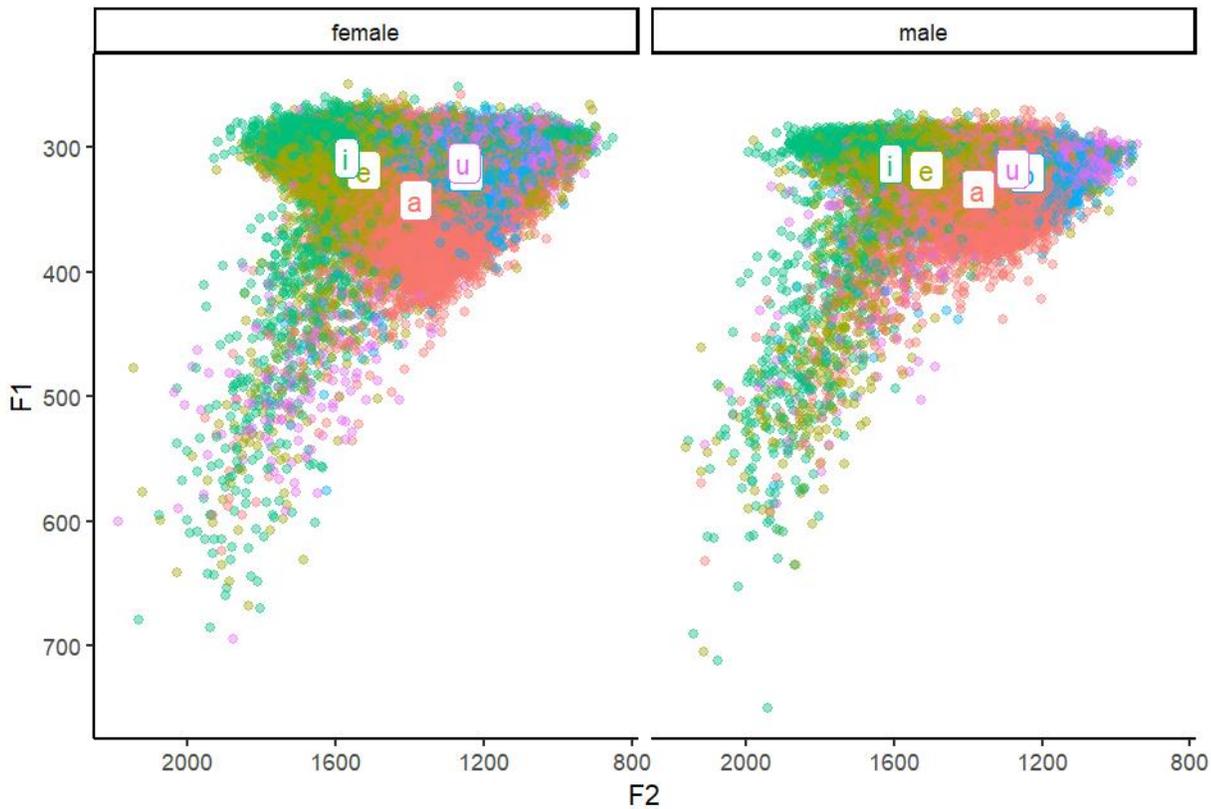
Figure 9. Vowel space plots for female (left) and male (right) speakers in the CID corpus. Each point represents a vowel token, color-coded by vowel category (a (*orange*), e (*yellow*), i (*green*), o (*blue*), u (*pink*)). Centroids indicate the mean F1 and F2 values per vowel category across all speakers of each gender.

For each plot, mean F1 and F2 values were calculated per vowel, either across all male and female speakers or within each speaker individually. The data were plotted in F1–F2 space using the *ggplot2* package in R, with vowel tokens shown as semi-transparent color-coded points and mean values overlaid as labeled centroids. Standard phonetic conventions were followed by reversing both axes, and plots were faceted by gender or speaker to facilitate side-by-side comparison. While the gender-based plots (Figure 9) provide a broad view of sex-related differences in articulatory space, the speaker-specific plots (Figure 10) offer a more fine-grained view of individual variation. Together, these visualizations enhance the interpretation of the acoustic data and support the broader goal of understanding how vowel articulation varies across speakers and contexts in spontaneous speech.
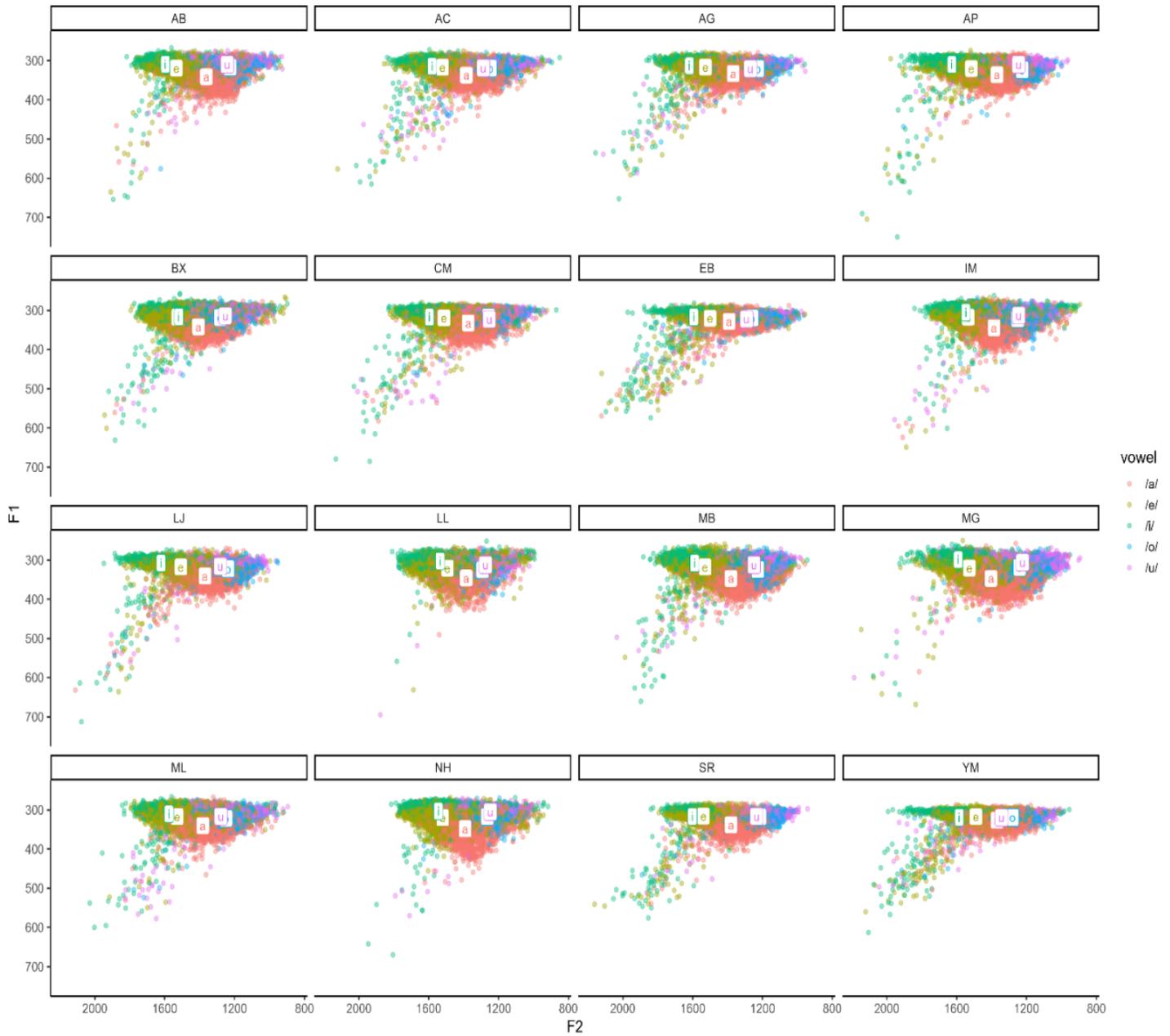
Figure 10. Speaker-level vowel space visualizations across the CID corpus. Each subplot shows a speaker's articulation of the five corner vowels, with color-coded points for individual tokens and mean formant values labeled at vowel centroids.

112

4.3.2. Formant centralization ratio (FCR): adaptation for French

Formant centralization refers to the tendency for vowel formant frequencies to converge toward a more neutral or centralized position within the vowel space. This phenomenon typically occurs in casual or rapid speech, resulting in reduced acoustic contrast between vowel categories. Seminal work by Peterson and Lehiste (1960) and Lindblom (1963) identified formant centralization as a hallmark of vowel reduction processes in connected speech, reflecting articulatory undershoot or decreased effort. Centralized vowels, being acoustically less distinct, contribute to a reduced vowel space and diminished vowel distinctiveness in conversational contexts.

Sapir and colleagues (Sapir et al., 2010) introduced the Formant Centralization Ratio (FCR), an acoustic measure designed to quantify the degree of vowel centralization more precisely. Originally developed for assessing dysarthric speech, the FCR captures vowel space reduction by comparing specific formant distances, making it a robust metric for studying both pathological and typical speech patterns. Incorporating measures like the FCR allows for a nuanced understanding of how articulatory characteristics influence speech clarity and reduction phenomena.

To quantify formant centralization within the present study, the vowel formant centralization ratio adapted for French (aFCR; (Audibert & Fougeron, 2012)) was employed. The aFCR provides a numeric estimate of system-level vowel compression by comparing formants that are expected to increase under centralization with those expected to decrease. Specifically, it captures articulatory undershoot effects, where vowels with typically extreme formant values (high or low) shift toward more centralized values (Sapir et al., 2010). The aFCR is considered sensitive to subtle articulatory variations such as those observed in clinical (Fletcher et al., 2017) or aging populations (Albuquerque et al., 2023), while remaining relatively robust to inter-speaker variability. It is calculated using the formant frequencies of the five peripheral vowels /i/, /e/, /a/, /o/, /u/, as follows:

$$\frac{F2u + F1i + F1u + F2o}{(increasing\ formant\ values)}$$
$$\frac{F2i + F1a + F2e}{(decreasing\ formant\ values)}$$

Equation 1. The formula for computing the Formant Centralization Ratio adapted for French.

A lower aFCR suggests greater distinctiveness between vowels, indicating clearer articulation. Conversely, a higher aFCR reflects a more centralized, compressed vowel system with less separation between vowel categories. In our dataset, the mean aFCR score was 0.91 ($SD = 0.02$), with no significant difference between female and male speakers ($t = 0.03$, $p = 0.97$).

To examine the potential relationship between vowel centralization and non-lexicalized reductions, we conducted a Bayesian linear regression predicting NLR ratios from aFCR. The effect of aFCR was highly uncertain (*Estimate= 5.86, 95% CI [-14.99, 26.97]*), with the credible interval including zero, suggesting little evidence for a reliable association between aFCR and NLR.

We then tested whether this relationship was modulated by articulation rate by fitting a second Bayesian regression model with an interaction term (aFCR × articulation rate). The results again provided no clear evidence for an effect of articulation rate on NLR (*Estimate = −11.14, 95% CI [−33.41, 11.61]*), nor for a significant interaction (*Estimate = 13.37, 95% CI [−11.24, 37.65]*). The main effect of aFCR also remained non-significant in this model (*Estimate = −159.35, 95% CI [−459.81, 138.12]*).

To further explore the articulatory underpinnings of vowel formant centralization (aFCR), we examined its relationship to two key measures: pentagonal vowel space area (pVSA) and vowel distinctiveness index (VDI). A moderate negative correlation was found between aFCR and pVSA (*r = −0.56, p = .023, 95% CI [−0.83, −0.09]*), suggesting that greater vowel centralization is associated with reduced vowel space area. This aligns with theoretical expectations, as a compressed vowel space reflects convergence of formant frequencies, which is captured by both aFCR and pVSA. More strongly, aFCR was highly negatively correlated with VDI (*r = −0.87, p < .001, 95% CI [−0.95, −0.66]*), indicating that

speakers with higher centralization ratios produced vowels that were less distinct from the central prototype. This finding supports the interpretation of aFCR as a robust marker of articulatory compression, with clear acoustic consequences for vowel dispersion.

While aFCR demonstrated robust relationships with articulatory structure as evidenced by its significant negative correlations with both vowel space area and vowel distinctiveness, it did not show a reliable association with reduction ratios (NLR%) in the current dataset. Additionally, no interaction with articulation rate was observed.

This lack of a direct link between aFCR and NLR does not undermine the validity of aFCR as a centralization measure; rather, it suggests that centralization and temporal reduction may reflect partially distinct aspects of speech production, or that the relationship between them may be more nuanced and influenced by other mediating factors. Together, these results highlight the importance of using multiple complementary articulatory and temporal measures when characterizing speech variability, and they contribute to a more fine-grained understanding of how vowel centralization relates to prosodic and temporal structure.

The preceding analyses focused on vowel formant centralization and its relation to vowel space structure and reduction in spontaneous conversational speech. While these findings offered valuable insight into the articulatory and acoustic patterns emerging in naturalistic conditions, speech production can vary considerably depending on task demands and context. To further contextualize the observed effects and to better understand how measures such as pVSA and VDI behave under more constrained articulatory settings, we next turn to speech produced in controlled conditions. Specifically, we examine a subset of speakers from the CID corpus who participated in both spontaneous and read-speech tasks. This comparison allows us to assess the extent to which vowel system characteristics observed in conversational speech persist or shift in more deliberate, carefully articulated contexts.

115

4.3.3. Comparison with controlled speech contexts

To complement the findings from conversational speech and assess vowel space measures across speaking styles, we examined vowel productions from a subset of CID speakers in a separate, controlled speech corpus (Meunier et al., 2003). This dataset includes three speakers (two female, one male: AC, LJ, and MB) who were also part of the main study and provides recordings under three distinct conditions: isolated vowels, monosyllabic words, and a monosyllabic text. By comparing vowel space metrics across these controlled tasks and conversational speech, we aimed to evaluate how articulatory precision and reduction vary across communicative contexts to draw attention to the degree of reduction in conversational speech.

4.3.3.1. pVSA across speech contexts

To assess the extent of vowel space reduction, we examined differences in vowel space area (pVSA) across four conditions: isolated vowels, monosyllabic words, monosyllabic text, and spontaneous conversation. Because pVSA reflects the acoustic dispersion of corner vowels, lower values typically indicate increased centralization and articulatory reduction.

The mean pVSA values showed a systematic decline from more controlled to more naturalistic speech. Specifically, average pVSA was 479286 (*SD = 163,10*) in the isolated vowel condition, 371562 (*SD = 108,553*) in the monosyllabic word condition, and 183861 (*SD = 57,982*) in the monosyllabic text condition. In spontaneous conversation, the mean pVSA was markedly lower at 78148[1] (*SD = 19,624*). To assess whether vowel space area (pVSA) differed across speech conditions, we fitted a linear mixed-effects model with Condition as a fixed effect and Speaker as a random intercept term: *pVSA ~ Condition + (1 | Speaker)*. The model showed substantial differences in pVSA across conditions (Figure

---

[1] The pVSA value for the conversational context was calculated using non-normalized formant values to ensure comparability with the controlled speech conditions, where formant values were also unnormalized. This approach was taken to maintain consistency in measurement across all speech contexts.

11). Compared to the conversational baseline, the vowel condition showed a robust increase in pVSA (*Estimate = 389,596, t = 6.22*), as did the word (*Estimate = 281,871, t = 4.50*) and text (*Estimate = 94,170, t = 1.50*) conditions. Although the contrast with text did not reach conventional levels of significance, the pattern suggests a gradation of vowel space reduction depending on the degree of speech spontaneity.
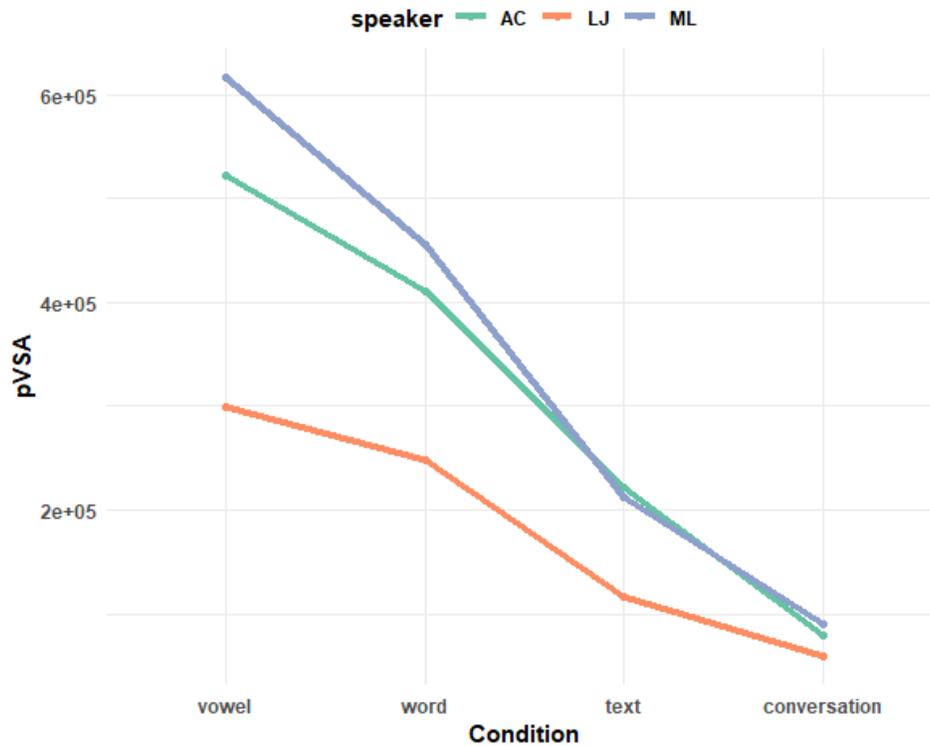
Figure 11. Variation in pVSA values by speaking condition. Each line represents an individual speaker, with colors distinguishing speakers. Speaking conditions are shown on the x-axis.

The model's residual standard deviation was estimated at 76,684, and the random intercept variance for speakers was 67,928, indicating speaker-level variability. These results were consistent with expectations: vowels produced in deliberate, clearly enunciated conditions occupy a larger acoustic space than those produced in casual, spontaneous contexts.

4.3.3.2.   VDI across speech contexts

To assess the consistency and separation of vowel categories in different speaking contexts, VDI values were also calculated for the same subset of speakers (AC, LJ, and ML) across the four conditions. VDI values decreased as speech became more spontaneous: the highest mean VDI was observed in the isolated vowel condition (*M = 3.98, SD = 0.63*), followed by word (*M = 3.00, SD = 0.74*), text (*M = 2.38, SD = 0.28*), and conversational speech[2], where vowel dispersion was the lowest (*M = 1.21, SD = 0.14*).
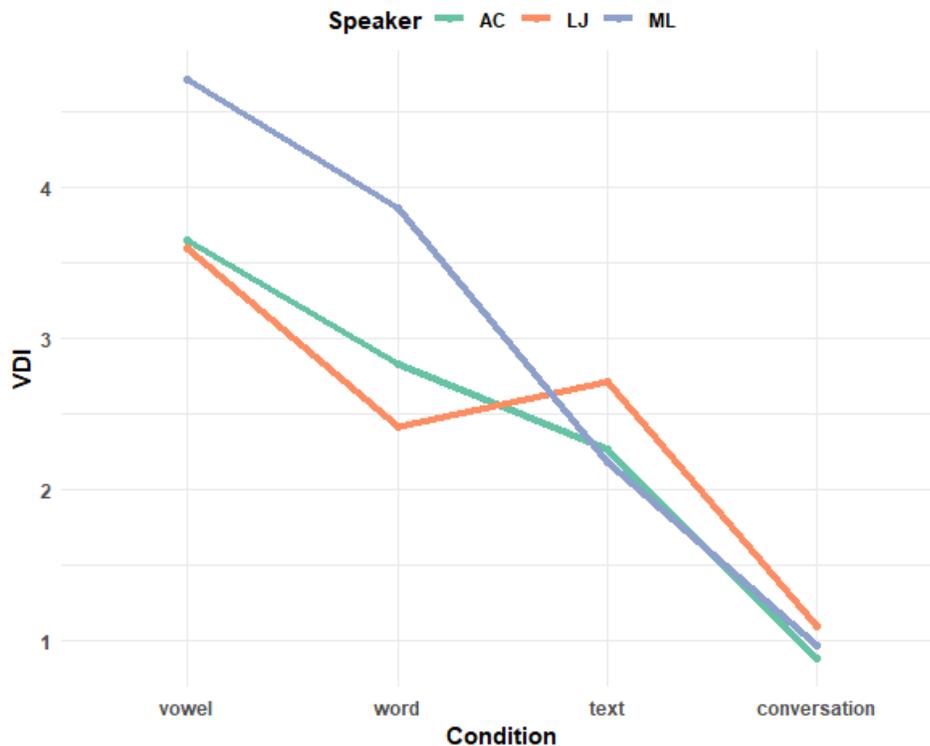
Figure 12. Variation in VDI values by speaking condition. Each line represents an individual speaker, with colors distinguishing speakers. Speaking conditions are shown on the x-axis.

---

[2] Again, the VDI scores were calculated based on the unnormalized formant frequencies to ensure a more reliable comparison with the controlled speaking situations.

118

A linear mixed-effects model confirmed significant differences in VDI across conditions (Figure 12). Compared to the conversational baseline, VDI values were significantly higher in the vowel (*Estimate = 3.00, t = 7.55*), word (*Estimate = 2.05, t = 5.16*), and text (*Estimate = 1.40, t = 3.52*) conditions. These findings suggest that as speech becomes less constrained and more spontaneous, the articulatory distinctiveness of vowel categories decreases, consistent with patterns of increased reduction and overlap in casual speech.

This analysis was conducted to illustrate the reduced articulatory characteristics of conversational speech compared to more controlled speaking conditions, reinforcing the notion that vowel articulation becomes less distinct as speech becomes more spontaneous. Interestingly, although VDI scores seemed more stable within conversational speech (Figure 11), greater intra-speaker variability emerged across the other conditions, suggesting that individual speakers may adjust their articulatory strategies more flexibly when shifting between structured and spontaneous speech.

4.4. Conclusion to the chapter

This chapter investigated the relationship between vowel space characteristics and non-lexicalized reductions in spontaneous speech. The main conclusion is that there is no significant correlation between vowel reduction metrics (such as pVSA and VDI) and speakers' non-lexicalized reduction ratios. This suggests that vowel reduction and NLR likely operate as distinct mechanisms, challenging oversimplified views that reductions share a common origin. Vowel reduction, as measured acoustically through vowel space metrics, is primarily influenced by articulatory dynamics like speech rate and clarity of articulation (Johnson, 2004), while NLR appears to be driven by temporal compression and other contextual factors.

A key contribution of this work is the combination of multiple vowel reduction metrics, which provides a more nuanced understanding of how vowel reduction manifests in conversation. By ensuring the complementarity of these metrics, we try to

avoid being misled by artifacts inherent to any single method, allowing for a more comprehensive analysis. Additionally, the comparison between read speech and spontaneous speech, though already well-documented in the literature, enriches our understanding of how speech style influences vowel articulation. The use of the same subset of speakers for both conditions controls for speaker variability, providing robust evidence of significant vowel reduction in spontaneous speech.

Furthermore, this study supports the conceptual proposition of non-lexicalized reductions, which contributes to the understanding of how different types of reductions operate in spontaneous conversation. While vowel reduction has often been treated as a monolithic phenomenon, our results indicate that non-lexicalized (also probably lexicalized) reductions are likely governed by distinct processes. This insight challenges traditional models of reduction, suggesting that these mechanisms may not be directly interrelated, but rather independent in their operation.

The findings may also raise important questions about the temporal and phonological asymmetry between lexicalized and non-lexicalized reductions. Lexicalized reductions, which tend to evolve diachronically, may reflect entrenched phonological patterns that do not align with the synchronic nature of vowel reduction, which occurs in real-time speech production. Non-lexicalized reductions, on the other hand, appear to be more closely tied to immediate processing demands, such as speech tempo and prosodic structure. This distinction helps explain the lack of significant correlation between vowel reduction metrics and the two types of reductions.

In conclusion, this study underscores the multidimensional nature of reduction in spontaneous speech and calls for a more sophisticated model that recognizes the distinct processes underlying phonetic and lexical reductions. It also highlights the need for further exploration into other potential predictors of vowel reduction, such as cognitive load and task-related factors, and suggests that future studies should refine measurement techniques to account for the variability in reduction behavior across different contexts and languages.

Having established that vowel reduction and non-lexicalized reductions operate through distinct mechanisms, the next chapter will turn to another important dimension of reduction: its relationship with the prosodic structure. While the previous chapter examined the temporal and spatial factors influencing vowel reduction, this chapter turns to the role of prosody, specifically how speech is structured into larger rhythmic and intonational units, which remains a key component of understanding the dynamics of speech reduction.

Chapter 5- Reduction within the prosodic structure

5.1. Introduction

This chapter extends the investigation of phonetic reduction initiated in the previous chapter by shifting from a focus on vowel space and articulatory precision to the role of prosodic structure in shaping reduction patterns. While Chapter 4 examined how reduced vowel systems interact with reduction, the present chapter examines whether and how these reductions are systematically influenced by the hierarchical organization of speech prosody.

Prosody, understood as the supra-segmental organization of speech, governs elements such as rhythm, intonation, stress, phrasing, and boundary marking (Himmelmann & Ladd, 2008). It provides a structural scaffolding that guides the timing and prominence of speech elements. One of the core theoretical assumptions of this chapter, following the Smooth Signal Redundancy Hypothesis proposed by Aylett and Turk (2004), is that prosodic structure plays a central role in regulating phonetic reduction. According to this model, speakers modulate the acoustic signal to preserve intelligibility and communicative efficiency. Reductions are more likely to occur in prosodically weak positions, such as unstressed syllables in non-prominent domains, where contextual redundancy is high and perceptual recovery is less at risk.

This prosody-based view of reduction aligns with the earlier findings from English (Bell et al., 2003b) and other languages, showing that phonetic reduction tends to be more common in weak prosodic environments and unstressed syllables, particularly those in medial positions of prosodic phrases. Research by Cho (2016) and Cole and Shattuck-Hufnagel (2016) further supports the view that acoustic reduction is modulated by

prosodic boundaries and prominence hierarchies, where domain-initial and domain-final positions are often enhanced while medial, non-prominent segments are more likely to be reduced.

This chapter examines these claims in the context of spontaneous French speech, a language whose prosodic structure has been characterized as distinct from stress-timed languages like English. French exhibits phrase-level prosody, where prominence is less lexical and more phrasal, organized around the Accentual Phrase (AP) and the Intonational Phrase (IP) (Jun & Fougeron, 1995, 2000). In this model, the AP is a key prosodic domain, marked by a characteristic pitch contour and prosodic cues such as final lengthening and initial boundary tones. Recent studies have proposed the existence of an Intermediate Phrase (ip) (Michelas & D'Imperio, 2010), a prosodic unit between the AP and IP that may further condition phonetic variation.

Despite this rich prosodic organization, phonetic reduction in French has received less attention than in English, particularly in spontaneous, conversational contexts. Most previous research has relied on read speech, limiting the generalizability of findings to more naturalistic settings (Ernestus & Warner, 2011). This chapter addresses this gap by analyzing a large corpus of spontaneous French conversations, offering a more ecologically valid investigation of reduction processes.

Building on the framework outlined above, this chapter seeks to answer the following central question: Are phonetic reductions more frequent in prosodically weak positions within the French prosodic hierarchy? By systematically examining reductions in relation to prosodic phrase position, boundary strength, and domain length, this study aims to provide new insights into the interaction between prosodic organization and speech reduction. In doing so, it highlights the central role of prosody as a mediating structure between linguistic form and phonetic realization, contributing to our understanding of speech variability and efficiency in spoken language.

This chapter is based on the study presented at *Interspeech 2023*, titled *Speech Reduction: Position within French Prosodic Structure* (Bodur et al, 2023) and aims to demonstrate the explanatory power of prosodic analysis for understanding reduction processes in naturalistic speech.

5.2. Article: Speech Reduction: position within French prosodic structure

This chapter is based on the following conference paper:

Bodur, K., Bertrand, R., German, J., Rauzy, S., Fredouille, C., & Meunier, C. (2023). Speech reduction: position within French prosodic structure. In *Proceedings of Interspeech 2023*, 20-24 August 2023, Dublin, Ireland.

Speech reduction: position within French prosodic structure

Kübra Bodur[1], Roxane Bertrand[1], James S. German[1], Stéphane Rauzy[1], Corinne Fredouille[2],
Christine Meunier[1]

[1]Aix-Marseille Université, LPL, CNRS, France
[2]Avignon Université, LIA, CNRS, France
kubra.bodur@univ-amu.fr

Abstract

Variation in the speech signal is a characteristic of spoken language, emerging partially as a result of interactions between various linguistic levels. One example of variation is phonetic reduction, where words are produced with missing or under-specified phonetic forms. Using a French conversational corpus, this paper focuses on the relationship between reduction and prosodic structure to see whether certain positions favor the occurrence of reduction. We annotated and observed the distribution of reduced sequences within specific prosodic domains (Intonational and Accentual Phrases). Preliminary analyses revealed that the detected reductions occur mostly mid-IP and very rarely at IP-final. However, this pattern may vary among speakers, as speakers have different patterns in terms of the number of reductions produced and their positions. It is also usually the case that the reduced sequences occurring mid-IP, coincide with the AP level boundaries, extending from one AP to another.

Index Terms: speech reduction, prosody, spontaneous speech

1. Introduction

Prosody can be defined as the supra-segmental organization of speech, encoding prominence, accent placement, phrasing, and speaking style. The way in which sounds are organized in speech, known as prosodic structure, follows a hierarchical order where larger units have more influence than smaller ones. The boundaries within prosodic structure are marked by changes in pitch, as well as variation in the acoustic/phonetic realizations of individual segments. At prosodic boundaries, the durations of segments might increase, gestures might be lengthened and this lengthening might increase as the prosodic boundary is hierarchically more important in the prosodic organization (e.g. [1, 2, 3, 4, 5]). The prosodic organization and its relationship with acoustic/phonetic realizations are crucial aspects to consider when exploring the role of prosodic structure in speech variation, even though these manifestations are specific to languages and their consonant categories. The influence of prosodic prominence might extend from individual segments to larger units such as syllables and words, contributing to a variation in speech signal.

Variation in the speech signal is an intrinsic feature of spoken language [6] as a result of the permanent adaptation of speech productions to the speaking context [7]. One way this variation might manifest itself in spontaneous conversations is through the hyper and hypo-articulation [7] of speech segments. In the case of the hypo-articulation, speech is not as clearly articulated as in formal or read speech [8, 9]. Segments (i.e. phonemes, syllables, words) may appear acoustically and phonetically weakened with shorter durations, reduced

vowel spaces and/or they might be completely missing in extreme cases of reduction [10, 11, 12, 13] all the while allowing the interaction to flow smoothly between speakers and listeners. Even though we know little about all the driving forces behind the reduction processes, previous research has suggested that reduction in speech is influenced by various linguistic and conversational factors such as semantic and lexical predictability [14, 15, 16], lexical frequency [17, 18], previous mention in discourse [19, 20], speaking style [9] as well as the phonological properties of the words composing reduced sequences [21, 22, 23]. It has been shown that prosodic structure is also one of the factors contributing to the level of reduction in speech [3, 19, 24, 25]. Aylett & Turk [3] claimed that reduction in speech occurs in order to preserve the smoothness/fluency of speech, suggesting that contextually redundant elements should be reduced through prosodic and phonetic means. They found that speakers tend to reduce the duration and amplitude of unstressed syllables within a word, but maintain greater precision in stressed syllables. Bell et al.'s [26] findings support this claim, as they found that vowels in unstressed syllables were more likely to be reduced than those in stressed syllables, and that the degree of reduction was influenced by the position of the syllable within the prosodic structure of the word. The same was proven for consonants in English [27] and both for vowels and consonants in Korean [28].

A Model of French Prosodic Structure
This study is based on the phonological model of French prosodic structure developed by Jun and Fougeron [29, 30]. In French, there are two main prosodic boundaries, the *Accentual phrase* (AP) [29] and the *Intonational phrase* (IP), which is a larger hierarchical structure that includes the AP. The AP is characterized by an initial boundary tone (LHi), a nuclear pitch accent, and a final tone (LH), and its location within the larger IP can influence the realization of the pitch accent and the overall f0 contour. The IP can consist of one or more APs and is delimited by a boundary tone that marks the end of the IP (L% or H%) and the beginning of a new one. The boundary between two IPs is marked by a larger change in pitch and duration than between accentual phrases. Additionally, research suggests that there is a third boundary level in French prosodic structure, the *intermediate phrase* (ip) [31, 32] which is larger than the AP but smaller than the IP and is characterized by a downstep (L-) and a final tone (either L- or H-). The downstep concerns the height of the peak of the accent of a non-initial AP, only if it is also non-final.

Our study aims to investigate the distribution of speech reductions within the prosodic structure of French and determine whether reductions are more likely to occur in certain positions.

Building on the work of Aylett and Turk [3], we make the assumption that reductions would be more prevalent in prosodically weaker positions, affecting unstressed syllables located in the middle of a prosodic domain for instance.

This study presents a novel approach to investigating reductions in spoken French by analyzing spontaneous conversations produced in a more natural context, in contrast to controlled and read laboratory speech. Additionally, this research focuses specifically on the French language, which has received comparatively less attention than English in terms of reduction studies. Moreover, we introduce a new method for detecting reductions in spontaneous conversations, enabling us to analyze reduced sequences involving multiple words (for more details, see Section 2). By combining these approaches, this study provides a descriptive examination of the relationship between prosodic structure and reductions in a set of conversations in French.

2. Methods

2.1. Corpus

For the study of reduction in spontaneous speech and its relation to the prosodic categories, we selected the *Corpus of Interactional Data* [33] as it includes phonetic and prosodic annotations performed both manually and automatically using tools that can process large amounts of speech data and provide relatively reliable outputs. The corpus consists of 8 audio-visual recordings (each lasting 1 hour) of conversational dyads between colleagues (10 females and 6 males, ($M = 34$ years old)) with relatively familiar relationships, which is thought to be a context enhancing the spontaneity of conversations.

2.2. Annotations

2.2.1. Detecting reductions

In order to study phonemic and temporal reduction in speech, we obtained reduced sequences by adopting a bottom-up approach. We used a script that detects sequences containing at least 6 phonemes in a 230ms window and continues to search for reductions through a sliding window. Taking individual alignment files containing phoneme sequences and their temporal boundaries as input, it counts from a phoneme's start time how many phonemes belong to the predefined window (230ms). If the threshold of the number of phonemes (6) is reached, the sequence is saved in a Textgrid and the script goes on to search for the following 230ms window. The script continues adding phonemes to the sequence as long as the threshold (6 phonemes) is exceeded. The 230ms window was selected after testing the script with other window sizes such as 180ms and 200ms, which pointed at more highly reduced (in terms of duration) areas in the corpus. The choice of 230ms is thus crucial, in that a larger window might contain more phonemes and words allowing us to obtain a sufficient number of reductions for a more reliable preliminary analysis.

2.2.2. Prosodic annotations

The prosodic annotation of the corpus was performed by semi-naive annotators following a perception based approach [34]. The annotation procedure follows Tones and Break Indices (ToBI) guidelines [35] in general by adapting the guidelines to the prosodic phrasing of French language following Jun and Fougeron's model [36]. The annotators labeled the ends of words from 0 to 3 according to the strength of their associations with the following word. In this annotation scheme, label 1 corresponds to the AP, 2 to the ip and 3 to the IP. However, this research is based on two main prosodic domains in French [30] since there is not enough evidence on the manifestation of the intermediate phrase [32] in conversational speech yet. Labels 2 and 3 were merged in the end, due to the lack of evidence for the presence of ip in spontaneous speech. Therefore, the coding scheme includes only AP and IP (see Figure 1). Additionally, the coding of numbers in the prosodic annotations "IP11" and "AP22" seen in the example are related to the inter-annotator agreement scores. For instance, when an accentual phrase is annotated as "AP02", 0 means no gold (no expert annotation) annotation while 2 means 2 naive annotators agreed upon the annotation. In the case of Figure 1, "AP22" means that all four annotators (2 expert and 2 naive) agreed on the annotation.
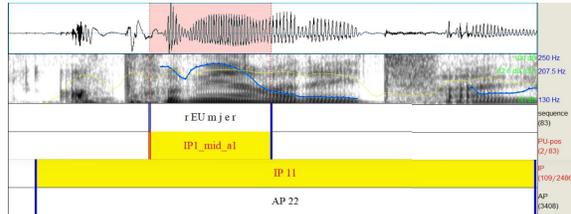
Figure 1: *An example of the annotation of reductions within prosodic domains. The first tier includes the reductions, the second shows how many APs form the IP (IP2), the position in the IP (mid), the AP containing the reduction (a2) and whether the reduction is on the frontier of a prosodic domain (Fa).*

2.2.3. Prosodic annotations and reductions

In order to analyze where in the prosodic domains the reductions occur, a second set of annotations based on the prosodic phrasing explained above was performed by the first author. The reduced sequences which had been detected automatically as well as the prosodic annotations of the corpus were brought together using Praat [37]. The annotation structure is as follows:

IP*n* (*n* = the number of APs making up the IP) - **position** in the IP (sta/mid/end/all)- **a*n***(n = the number of the AP within its IP).

Two complementary categories were also included in the annotation scheme: Fa and Fi. *Fa* is used when the reduced sequence crosses an AP boundary, meaning the sequence is stretching between two or more APs within the same IP, whereas *Fi* indicates that the reduced sequence crosses an IP boundary, extending to the previous or following IP (see Figure 1 for an illustration).

Disfluencies While annotating the positions of reductions within prosodic units, it was observed that there were many cases where the IPS contained disfluencies (repetitions, filled pauses, false starts) -which is a characteristic of spontaneous speech- as well as reductions. As disfluencies occurred quite frequently in the corpus we used (as frequent as 1 disfluency in every 15 words [38]), and as they interact with the prosodic structure [39, 40] thus making prosodic segmenting more difficult, we decided to exclude the annotations that coincide with disfluencies even if the distribution of reduction in disfluencies may be relevant. The disfluency annotations were already available for the corpus. This procedure resulted in the exclusion of 351 annotations (36,9 % of the reduced

sequences). However, the excluded data will be useful for further investigation of reduction processes.

3. Results

Having extracted the annotations, we used a custom-made script to exclude reductions coinciding with disfluencies (351 cases out of 950 detected reductions) and a case for which there was no prosodic annotation available. We ended up with 598 annotations for prosodic units and counted the occurrences of reductions at various positions (initial, middle, end, all) with respect to the IPs.

3.1. Distribution of reduced sequences within IPs

We first observed the overall distribution of reduced sequences at various positions within the IP. The results are shown in Table 1. Based on our annotations, reduced speech was produced mostly at IP-mid position (56%). It was also quite frequent at IP-initial (32%) while it was rare at IP-final (9%) or over all of the IP (3%).

Position	Count	Ratio
start	193	32%
mid	336	56%
end	53	9%
all	16	3%
	598	

Table 1: *The various positions where reduction occurred and their distributions within the IP.*

3.2. Distribution of positions among speakers

Figure 2: *The distribution of reductions in various positions within the IP per speaker. The bars show the ratios (%) of different positions where reductions occurred. The line shows the number of reductions produced by each speaker of the corpus.*

The distributions of above-mentioned positions for the annotated reduced sequences were observed separately for each speaker of the corpus. The general pattern is compatible with the overall results presented above. The reduced sequences occur mostly in the middle of an IP (see Figure 2 below). However, some speakers displayed different patterns. For example, the reductions produced by S10 tend to be found more frequently at IP-initial (40% of his reductions) than the other positions. Similarly, for S8, the reductions were more frequently at IP-final (46,68%) followed by IP-initial (26,66%) while they appear only 20% of the time in the middle. Figure 2 shows that there is no relation between the number of speakers' reductions and the specific distribution of their reductions.

3.3. Reductions on the boundaries

While annotating reduced sequences within prosodic domains, it has been observed that numerous sequences occurred on prosodic boundaries. It was usually the case that a sequence extended from the end of one IP to the start of another or from one AP to another in the middle of an IP (see Figure 3 for an example). Out of a total of 598 reductions, 116 (19,3%) appeared on AP boundaries, while 135 (22,5%) appeared on IP boundaries extending from one utterance to another.
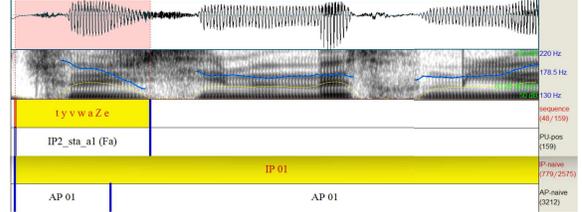
Figure 3: *An example of a reduced sequence occurring on two APs, the sequence is "tu vois/j'ai essayé de me mettre au milieu" ("you see/i tried to put myself in the middle").*

3.4. The length of the IPs containing reduced sequences

The total number of IPs found in the corpus being 31,119, only 2% of these were affected by reductions detected using the 230ms/6 phonemes parameter (after excluding the disfluent IPs). The annotation scheme allowed us to obtain the length of the IPs containing reductions, and the length of the IPs is calculated based on the number of APs found in them. The results suggest that, in the 598 detected sequences, reductions appeared more in shorter IPs (in terms of the APs they contain) than the IPs composed of 6 or 9 APs (see Table 2) which is rare but possible in the French prosodic structure.

IP length	Count	IP length	Count
IP1	309	IP5	6
IP2	195	IP6	2
IP3	59	IP9	1
IP4	26	**Total**	598

Table 2: *A summary of the IP length (based on the number of APs involved) containing reduced sequences.*

4. Discussion

The current paper investigates the relationship between reduction in spontaneous speech and prosodic structure, more specifically whether reductions occur in certain positions within prosodic domains. Based on a phonological model of French prosodic structure [30], it aims to investigate the distribution of reduced sequences within this structure.

We adopted a new and automatized method for the detection of reduced sequences to discover possible reductions in a corpus of conversational speech in French [33]. The selected window size (230ms/6 phonemes) focuses mostly on reductions involving multiple words -especially monosyllabic words-; suggesting that "smaller" reductions, affecting only one or two phonemes within a word, have not been taken into consideration in this study. We tried to analyze a reduction pattern in conversational speech using a bottom-up method. This is part of the originality of our study, compared to previous ones concentrating on reductions at the word level through a top-down

approach for detection of reductions (based on well known lex-icalized reduction forms).

For descriptive purposes, we observed where reductions oc-curred more frequently, whether there was a global tendency for speakers to produce reductions in similar positions and whether the length of a prosodic phrase played a role in the occurrence of reductions in these positions. As a characteristic of sponta-neous speech, some reductions co-occurred with interruptions in the speech flow, which are called "disfluencies". These were not included in the analysis of the data. However, if the reduced sequence was found in the *Reparans* which is the repairing part after the disfluency, we included that sequence in the analysis. The goal of this elimination process was to have a coherent pre-liminary analysis, since disfluencies are known to be able to cause ambiguity in separating prosodic units [41], as well as the segmentation of words making up the reduced sequences. This exclusion allowed us to have a first impression of how reduc-tions behave in speech when they are not disrupted by conver-sational factors. However, we believe that the study of the re-ductions containing disfluencies will also prove very useful for understanding the reduction mechanisms in the future, as there might also be an interaction between disfluency and reduction in spontaneous speech.

Looking at the distributions of reduced sequences at var-ious positions within the Intonational phrase (IP), the results suggest that reductions occurred mostly in the middle of the IPs (56% of the detected reductions) in the corpus. Compatible with our hypothesis that reductions would be more prevalent in mid-domain positions as these areas are susceptible to being prosod-ically weaker, thus less informative, this finding could be also considered compatible with previous research [3, 26] consider-ing the French prosodic structure. It is important to note that, although reduced sequences occurred preferably in the middle of an IP, no position was excluded for them to occur. Another finding is that, even if the distribution pattern of reductions in prosodic structure is quite consistent among speakers, it still displays variability to a certain degree. Some speakers tend to produce more reductions at the beginning or end of the IP. See-ing that the variation is also very high among the speakers in terms of the numbers of reductions produced in conversations (e.g., S7 produced 107 reduced sequences while S10 had only 5 -after the exclusion of disfluent IPs), these differences might be explained by individual characteristics of each speaker. Despite the fact that speaker related variability has rarely been addressed in previous research (except for Labov's work on sociological factors [42, 43]), it contributes significantly to the heterogene-ity in reduction rates; speakers behave differently regarding the number of reduced productions (see Figure 2 for an illustration). In a previous analysis, we found that articulation rate is one of the factors contributing to inter-speaker variation: a positive correlation was found between the articulation and reduction rates ($r(14) = 0.78, p = 0.0002996$). However, articulation rate itself is not enough for accounting for the differences between speakers in terms of reduction behavior, further investigation is required to see what might be causing these differences.

We also observed that some reductions can be realized at the junction between two APs. The AP ends with a pitch ac-cent realized by a movement of f0 (most often rising) and by a lengthening of the final full syllable of the lexical word. The ob-served reduced sequences concern cases where function words are often found at the beginning of the following APs (e.g., je crois/que...; chaque fois /qu'il y avait; tu vois/que... ;elle avait/un...). This finding is in line with the results obtained in previous studies, (not only on lexical predictability but also

on unstressed versus stressed syllables) since these reductions mainly concern function words found at the beginning of the following APs which can even be considered informationally redundant in many cases. For further analysis, it would be in-teresting to measure the durations of stressed syllables (from AP 1) which, although they are perceived as stressed since they are associated with an AP boundary, could nevertheless be re-duced compared to other syllables in this position which are not followed by reduction.

Furthermore, when we looked at the lengths of the IPs con-taining reduced sequences, we saw that shorter IPs were more common than the longer ones among the IPs containing reduc-tion (IP1 and IP2 make up 84,3% of the relevant IPs). On the other hand, the relationship between reduction and the length of an utterance can be very complex and may depend on a num-ber of factors, including the context in which the utterance oc-curs, the speaker's fluency, and the degree of stress or emphasis placed on certain words or phrases. Considering that disfluen-cies might cause an IP to be longer due to pauses, repetitions, and restarts, it can be thought that this dominance of the shorter IPs might be the result of the excluded IPs containing disflu-encies. However, when we compared the disfluency including and disfluency excluding datasets, the overall distribution did not change for this analysis. In this case, it is difficult to gen-eralize and to predict what type of IPs are more susceptible to be affected by reductions, since we do not have the information regarding the lengths of all the IPs comprising the corpus.

Our research being limited in terms of the number of re-ductions we were able to detect using the selected parameters (230ms/6 phonemes), we believe more spontaneous speech data would provide more opportunities to understand the relation-ship between reduction and its place in the prosodic structure. The fact that the evidence for the manifestation of the inter-mediate phrase [32] is not adequate for conversational speech may have limited our interpretation of the results, however, with more evidence in the future, this level could also prove to be useful in further accounting for the interaction between reduc-tions in French prosodic structure.

In conclusion, the preliminary results show that the distri-bution of reductions in various prosodic positions displays vari-ability as regards to the speakers and the length of the utter-ances. Even though we aim at exploring the relationship be-tween reduction and the prosodic structure, the prosody is not enough for accounting for the occurrence of reductions on its own. When enriched with more detailed acoustic and phonetic analyses, this preliminary research would serve as a valuable basis for further research to fully comprehend the mechanisms causing reduction during speech.

5. Acknowledgements

6. References

[1] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1707–1717, 1992.

[2] D. Byrd, J. Krivokapić, and S. Lee, "How far, how long: On the temporal scope of prosodic boundary effects," *The Journal of the Acoustical Society of America*, vol. 120, no. 3, pp. 1589–1599, 2006.

[3] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and speech*, vol. 47, no. 1, pp. 31–56, 2004.

[4] D. Byrd and S. Choi, "At the juncture of prosody, phonology, and phonetics—the interaction of phrasal and syllable structure in shaping the timing of consonant gestures," *Laboratory phonology*, vol. 10, pp. 31–59, 2010.

[5] J. Cole, "Prosody in context: A review," *Language, Cognition and Neuroscience*, vol. 30, no. 1-2, pp. 1–31, 2015.

[6] M. T. C. Ernestus, *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*. Utrecht: LOT, 2000.

[7] B. Lindblom, "Explaining phonetic variation: A sketch of the h&h theory," in *Speech production and speech modelling*. Springer, 1990, pp. 403–439.

[8] M. Ernestus and N. Warner, "An introduction to reduced pronunciation variants," *Journal of Phonetics*, vol. 39, no. SI, pp. 253–260, 2011.

[9] C. Gendrot, M. Adda-Decker, and C. Schmid, "Comparaison de parole journalistique et de parole spontanée: analyses de séquences entre pauses," in *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012*, 2012, pp. 649–656.

[10] B. Schuppler, M. Ernestus, O. Scharenborg, and L. Boves, "Acoustic reduction in conversational dutch: A quantitative analysis based on automatically generated segmental transcriptions," *Journal of Phonetics*, vol. 39, no. 1, pp. 96–109, 2011.

[11] O. Niebuhr and K. J. Kohler, "Perception of phonetic detail in the identification of highly reduced words," *Journal of Phonetics*, vol. 39, no. 3, pp. 319–329, 2011.

[12] M. Ernestus, "Acoustic reduction and the roles of abstractions and exemplars in speech processing," *Lingua*, vol. 142, pp. 27–41, 2014.

[13] C. G. Clopper, R. Turnbull, F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, and M. Zellers, "Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors," *Rethinking reduction*, pp. 25–72, 2018.

[14] M. Aylett and A. Turk, "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 3048–3058, 2006.

[15] C. G. Clopper and J. B. Pierrehumbert, "Effects of semantic predictability and regional dialect on vowel space reduction," *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1682–1688, 2008.

[16] R. Turnbull, "The role of predictability in intonational variability," *Language and speech*, vol. 60, no. 1, pp. 123–153, 2017.

[17] M. Pluymaekers, M. Ernestus, and R. H. Baayen, "Lexical frequency and acoustic reduction in spoken dutch," *The Journal of the Acoustical Society of America*, vol. 118, no. 4, pp. 2561–2569, 2005.

[18] J. Bybee, R. J. File-Muriel, and R. N. De Souza, "Special reduction: A usage-based approach," *Language and Cognition*, vol. 8, no. 3, pp. 421–446, 2016.

[19] R. E. Baker and A. R. Bradlow, "Variability in word duration as a function of probability, speech style, and prosody," *Language and speech*, vol. 52, no. 4, pp. 391–413, 2009.

[20] J. M. Kahn and J. E. Arnold, "Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground," *Language, Cognition and Neuroscience*, vol. 30, no. 1-2, pp. 103–119, 2015.

[21] N. Pharao, "Consonant reduction in copenhagen danish," *Unpublished Ph. D. dissertation, University of Copenhagen, Denmark*, 2010.

[22] C. Meunier and B. Bigi, "Répartition des phonèmes réduits en parole conversationnelle. approche quantitative par extraction automatique," in *Actes de la conférence conjointe JEP-TALN-RECITAL 2016*, 2016, pp. 615–623.

[23] ——, "Variations temporelles des phonèmes en parole conversationnelle: propriétés phonétiques et facteurs lexicaux," *Studii de lingvistică*, 2021.

[24] C. é Fougeron and S.-A. Jun, "Rate effects on french intonation: Prosodic organization and phonetic realization," *Journal of Phonetics*, vol. 26, no. 1, pp. 45–69, 1998.

[25] R. S. Burdin and C. G. Clopper, "Phonetic reduction, vowel duration, and prosodic structure." in *ICPhS*, 2015.

[26] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational english," *Journal of Memory and Language*, vol. 60, no. 1, pp. 92–111, 2009.

[27] L. C. Dilley and M. A. Pitt, "A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2340–2353, 2007.

[28] T. Cho and P. A. Keating, "Articulatory and acoustic studies on domain-initial strengthening in korean," *Journal of phonetics*, vol. 29, no. 2, pp. 155–190, 2001.

[29] S.-A. Jun and C. Fougeron, "The accentual phrase and the prosodic structure of french," in *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 2, 1995, pp. 722–725.

[30] ——, "A phonological model of french intonation," *Intonation: Analysis, modelling and technology*, pp. 209–242, 2000.

[31] A. Michelas and M. d'Imperio, "Durational cues and prosodic phrasing in french: evidence for the intermediate phrase," in *Speech Prosody*, 2010, p. 4.

[32] A. Michelas, "Caractérisation phonétique et phonologique du syntagme intermédiaire en français: de la production à la perception." Ph.D. dissertation, Université de Provence-Aix-Marseille I, 2011.

[33] R. Bertrand, P. Blache, R. Espesser, G. Ferré, C. Meunier, B. Priego-Valverde, and S. Rauzy, "Le cid-corpus of interactional data-: protocoles, conventions, annotations," *Travaux interdisciplinaires du Laboratoire parole et langage d'Aix-en-Provence (TIPA)*, vol. 25, pp. 25–55, 2006.

[34] L. Prevot, R. Bertrand, K. Peshkov, S. Rauzy, and P. Blache, "Prosody, Discourse and Syntax in French Conversations: Resource creation and Evaluation," Feb. 2023, working paper or preprint. [Online]. Available: https://hal.science/hal-01231884

[35] M. E. Beckman and G. A. Elam, "Guidelines for tobi labelling, version 3," *Ohio State University*, 1997.

[36] S.-A. Jun and C. Fougeron, "Realizations of accentual phrase in french intonation," 2002.

[37] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (version 6.0. 14)," *Retrieved from (last access: 29.04. 2018)*, 2016.

[38] B. Pallaud, R. Bertrand, L. Prevot, P. Blache, and S. Rauzy, "Suspensive and disfluent self interruptions in french language interactions," 2019.

[39] E. Shriberg, R. Bates, and A. Stolcke, "A prosody only decision-tree model for disfluency detection," in *Fifth European Conference on Speech Communication and Technology*, 1997.

[40] J. Beliao and A. Lacheret, "Disfluency and discursive markers: when prosody and syntax plan discourse," in *DiSS 2013: The 6th Workshop on Disfluency in Spontaneous Speech*, vol. 54, no. 1, 2013, pp. 5–9.

[41] L. Prevot, R. Bertrand, and S. Rauzy, "Investigating disfluencies contribution to discourse-prosody mismatches in french conversations," in *The 10th Workshop on Disfluency in Spontaneous Speech*, 2021.

[42] W. Labov, "Sociolinguistic patterns (university of pennsylvania, philadelphia)," 1972.

[43] ——, "Principies oflinguistic change: Social factors," *Malden, Massachusetts: Blackwell*, 2001.

5.3. Conclusion to the chapter

This chapter investigated the role of prosodic structure in shaping the distribution of phonetic reductions in spontaneous French speech. By adopting a bottom-up, data-driven approach and focusing on conversational data, the study contributes to a relatively more ecologically valid understanding of reduction phenomena. The findings largely confirm the central hypothesis: reductions are more likely to occur in unstressed and prosodically weak positions, especially in the medial portions of intonational phrases.

The first key contribution of this study lies in its empirical validation of the hypothesis that reductions tend to occur more in less informative zones of speech. By analyzing naturally occurring data from a corpus of spontaneous French, the study demonstrates that reductions are significantly more frequent in the initial and medial positions of intonational phrases, areas commonly associated with low prominence and limited information density (Aylett & Turk, 2004). This finding not only supports prior theoretical claims but also reinforces the functional link between prosodic organization and the speaker's strategic allocation of articulatory effort based on communicative relevance.

Second, this chapter offers a refinement to current prosodic models by applying the autosegmental-metrical (AM) framework to account for the specific structural contexts in which reduction occurs. While prior studies have largely emphasized reduction avoidance in phrase-final or lexically stressed positions (Fougeron & Keating, 1997), this study extends those insights by identifying which prosodic positions are most vulnerable to reduction, namely, those within intermediate and intonational phrase domains that lack prosodic prominence. Importantly, the findings also underscore the relevance of the intermediate phrase in the French AM model (Jun & Fougeron, 2000), a prosodic level often overlooked in reduction studies, yet shown here to somehow modulate articulatory patterns.

Finally, this chapter contributes to broadening the scope of reduction research by foregrounding the temporal dynamics of reduction within prosodic boundaries. While much existing literature has concentrated on segmental changes or stress-conditioned reduction (Kohler, 1990; Lindblom, 1990), this study highlights how reduction patterns also align with temporal structures in speech. The use of a detection script capable of identifying temporally compressed zones enabled a more detailed examination of how reduction unfolds over time in relation to prosodic units, revealing reduction as not merely a segmental phenomenon but one that is dynamically shaped by the timing and rhythm of speech.

Together, these contributions underscore the importance of prosodic structure in shaping reduction behavior, particularly in spontaneous, conversational contexts. By linking reduction to prosodic position, temporal organization, and informational structure, this chapter strengthens the theoretical integration between prosody and articulatory economy, opening avenues for future work on the interface between speech planning, rhythm, and reduction. At the same time, speaker-specific variation and the complexity of spontaneous speech highlight the need to consider a wider range of acoustic, contextual, and cognitive factors when analyzing reduction behavior.

While this chapter has focused on adult speech, the next chapter shifts attention to speech reduction in children, aiming to explore its developmental trajectory. Understanding how and when children begin to exhibit reduction patterns, and how these align with or diverge from adult norms, offers critical insight into the acquisition of prosodic competence and the gradual mastery of speech economy in natural communication. By comparing adult and child speech, the following analysis will deepen our understanding of how phonetic reduction emerges and evolves across different stages of language development.

Chapter 6- Reduction in children's speech

6.1. Introduction

This chapter concentrates on the role of speech reduction in spontaneous conversation, focusing on its developmental trajectory in middle childhood (ages 7 – 11). Reduction is a pervasive and systematic phenomenon that affects the production of phonemes, syllables, and entire lexical items across languages and speaker populations. Far from being "sloppy" or erratic, reduction is now widely recognized as a rule-governed process, sensitive to linguistic, prosodic, and informational structure (Aylett & Turk, 2004; Clopper & Turnbull, 2018). Despite a growing body of research on reduction in adult speech, relatively little is known about how these patterns emerge and evolve during childhood, especially during the middle childhood years, a period often overlooked in developmental phonetics and pragmatics.

This chapter addresses that gap by investigating how children produce lexicalized and non-lexicalized reductions in naturalistic caregiver-child interactions. The motivation for this study is twofold. First, it challenges the assumption that reduction is an automatized or low-skill process. On the contrary, we argue that reduction is a developmentally complex, skillful behavior, requiring the integration of phonological, articulatory, and prosodic knowledge. In this view, the ability to produce contextually appropriate reductions is not immediate but rather emerges gradually with increasing linguistic competence. We hypothesize that lexicalized reductions, which are more conventional, frequent in the input, and commonly used across speakers, emerge earlier in development. In contrast, non-lexicalized reductions, which require real-time planning and fine articulatory control, tend to appear later, as they place greater cognitive and motor demands on the speaker and also require mastery of the joint management of reduction, discourse and prosody.

132

Second, the study contributes to our understanding of middle childhood as a critical period for the refinement of conversational and phonetic skills. While children at this age have already acquired the fundamental structures of their language (e.g., phonology, syntax), they are still developing pragmatic and prosodic competence. This includes conversational abilities such as turn-taking (Garvey & and Berninger, 1981), conversational grounding (Hess & Johnston, 1988), and coherent contingent exchanges (Baines & Howe, 2010), all of which might influence reduction behavior. Studying reduction in this age group allows us to isolate the development of reduction mechanisms from basic linguistic acquisition, providing a clearer picture of how communicative efficiency and phonetic fine-tuning evolve during this important developmental stage.

6.2. Article: Speech Reduction in Middle-Childhood: the Case of Child-Caregiver Interactions

This chapter is based on the following article submitted to the First Language Journal.

Bodur, K., Fourtassi, A., Fredouille, C., & Meunier, C. (2025). Speech reduction in middle childhood: the case of child-caregiver interactions. *First Language*.

Speech Reduction in Middle Childhood: the Case of Child–Caregiver Interactions

Kübra Bodur[1] (iD), Abdellah Fourtassi[2],
Corinne Fredouille[3] and Christine Meunier[1]

Abstract

Reduction, a common phenomenon in everyday speech, affects phonemes, syllables, and words across various languages and speakers. Despite its ubiquity, reduction remains underexplored, particularly in children's speech. This study addresses this gap by analyzing reduction patterns in French-speaking children aged 7 and 11 during video call interactions with caregivers, focusing on both lexicalized and non-lexicalized reductions. We found that children exhibited similar patterns of lexicalized reductions to adults, based on 11 common French reduction forms, with no significant differences between the two-child groups. However, reduction rates varied by specific items, with some showing high reduction ratios despite low frequency. Non-lexicalized reductions, involving more complex linguistic mechanisms, were rare among children but varied among caregivers, suggesting that children had not yet mastered these advanced forms. There was no correlation between children's and caregivers' reduction strategies, indicating that caregivers' behaviors did not directly influence children's. Additionally, articulation rates differed significantly across groups, increasing with age and correlating with higher reduction rates. These results suggest that while 7- and 11-year-olds exhibit similar language and reduction patterns, they are still developing the more sophisticated reduction mechanisms seen in adults.

[1]Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
[2]Aix Marseille Univ, CNRS, LIS, Marseille, France
[3]Avignon University, LIA, Avignon, France

Corresponding author:
Kübra Bodur, Aix Marseille Univ, CNRS, LPL, 5 Avenue Pasteur, Aix-en-Provence 13100, France.
Email: kubra.bodur@univ-amu.fr

Introduction

Spoken language is a remarkably adaptable means of human communication, facilitating interaction across various contexts and among diverse speakers. This variability arises from the dynamic and responsive nature of language, where speakers are influenced by contextual factors and their audience during speech production (Arnold et al., 2012; Bell et al., 2009; Lindblom, 1990; Kahn & Arnold, 2015). This adaptability is evidenced by phenomena such as reduction, where linguistic elements may be simplified or modified in real time to improve communication efficiency. Conditioned by various linguistic, social, and cognitive factors (Clopper & Turnbull, 2018), reduction refers to the process whereby speech units contain relatively less acoustic and phonetic substance, leading to shorter durations and less precise articulations.

Universally present in all languages, and common in the speech of many speakers (Gendrot & Adda-Decker, 2007; Ernestus & Warner, 2011), this phenomenon still has not been fully explored in all its facets despite its extensive occurrence. Wu and Adda-Decker (2020) showed that in a corpus of casual French conversations, almost 14% of words were affected by reduction. Similarly, Johnson (2004) found that in American English, 6% of syllables are missing and segment deletions occur in roughly every fourth word, with function words being more frequently affected than content words. Several factors influence reduction in speech, including phonetic characteristics, lexical stress, prosody, and speech rate, with higher articulation rates inducing more reductions. Lindblom (1963) argues that reduction is primarily a consequence of constraints on articulatory speed, which result in formant undershoot. Social and contextual factors, like conversational setting and interlocutor familiarity, further modulate speech reduction patterns (Labov, 1972; Lindblom, 1990). The terms 'reduced' and 'reduction' imply a contrast with a canonical form. However, linguistic studies suggest that articulation does not follow a strict categorical opposition. Instead, it exists on a continuum, with reduced forms situated along this spectrum (Nolan, 1992; Zellers et al., 2018). This observation aligns with the hyper and hypo-articulation theory (Lindblom, 1990), which posits that articulation is shaped by two main constraints: the need to minimize the speaker's effort during production, and the necessity to ensure that the message remains comprehensible to the listener. Despite the term 'reduction' implying a degradation of speech, it involves a complex set of skills, as the message is still understandable by the listener even with the lack of acoustic substance.

To understand the linguistic and cognitive bases of speech reduction, it is essential to explore its development into adulthood. The relationship between phonetic and phonological development and speech reduction in children is complex, influenced by various factors. These include the growth of their vocabulary (lexical development), the maturation of their ability to produce speech sounds (speech motor development), the planning of these sounds (phonetic planning), and the capacity to hold and manipulate sounds in

their memory (phonological working memory). Together, these elements interact to shape how children develop and refine their speech patterns. Research indicates that phonological memory is crucial for speech development, with studies examining its relationship to spoken language development (Waring et al., 2019). Phonological working memory might play a role in the degree of speech reduction, as children with stronger working memory may show more efficient speech patterns, balancing between reduced forms for ease of production and clear articulation for comprehensibility. Articulation rate also influences reduction, as it directly impacts speech patterns (Redford, 2014). Children's slower, more variable speech movements with larger relative amplitudes (Cheng et al., 2007) contribute to a slower articulation rate, reducing gestural overlap in consonant-vowel and consonant-consonant sequences. This likely results in less reduction, as an increased overlap in adult speech promotes more efficient, compressed articulation. Since articulation rate is tied to motor control, these developmental differences reflect the refinement of speech coordination, with greater overlap and reduction emerging as children's motor skills mature. The degree of speech reduction may vary with the child's age or developmental stage, similar to many other variables observed in infant-directed speech (e.g., Stern et al., 1983). Typically, infants are exposed to adult speakers who tend to overemphasize segments to enhance their distinctiveness, supporting their linguistic development and adapting to the listener's needs (Lindblom, 1990). This emphasis and meticulous pronunciation suggest that infant-directed speech may include fewer reduced pronunciation variants compared to informal adult-directed speech.

In a prior study, Lahey and Ernestus (2014) investigated how young infants (11–12 months) are exposed to speech reduction to understand the development of phonological representations in the mental lexicon. They hypothesized that infants may acquire the full pronunciation of words before forming lexical representations for their reduced variants, depending on the type and frequency of reduced pronunciations in infant-directed speech. Analyzing two frequently reduced Dutch words '*allemaal*' and '*helemaal*', they found that these two words undergo frequent reduction in infant-directed speech, with reduction levels similar to those observed in casual adult conversations. This suggests that children may acquire lexical representations for reduced pronunciation variants very early in their linguistic development. Additionally, studies on children from 1 to 3 years (Bard & Anderson, 1994) and 2 to 4 years (Shockey & Bond, 1980) had revealed that children of these age groups were also exposed to highly reduced linguistic material.

Interestingly, while numerous studies have explored infants' and young children's exposure to reduced speech variants, there is a notable gap in understanding how children themselves produce reductions beyond early childhood, aside from studies addressing cluster reduction typical in toddlers. Existing research has largely focused on cluster reduction in toddlers aged 1.6 to 2.8, primarily among Dutch- and English-speaking children (e.g., Carter & Gerken, 2004; Gulian & Levelt, 2009, 2011; Pater & Barlow, 2003). Investigating weak syllable omission in young English-speaking toddlers, Carter and Gerken (2004) found that although children often omitted unstressed syllables, acoustic traces of the omitted material persisted. Pater and Barlow (2003) also found that reduction patterns often favor preserving the least sonorous segment, though individual variation might reflect different constraints. Additionally, Gulian

and Levelt (2011) found temporal traces of deleted segments in reduced /sC/ clusters, indicating that children's lexical representations are more detailed than their surface productions suggest, further corroborating the idea that reductions may arise during phonetic execution rather than phonological planning. While this body of work provides important insights into early speech production, much less is known about how these processes develop in older children. Research by Strik et al. (2008) suggests that speech reductions continue to evolve with age, as articulation rate and motor control shift over the lifespan; however, their study focused on adults aged 20 to 79, requiring complementary explorations on whether similar patterns occur in younger populations. Surprisingly, there has been limited research on how typically developing children, particularly those in middle childhood (ages 6–12), produce reduced speech. This developmental stage is crucial, as children in this age range have already developed foundational linguistic skills, including phonetics, phonology, and prosody, with adequate motor control. They might demonstrate fluency in everyday conversations, even though speech production is a neuro-motor behavior that continues to improve and evolve with practice into adulthood.

The present study aims to offer preliminary quantitative insights into reduction behavior during middle childhood. By examining reduction in children's speech, this research seeks to contribute to a deeper understanding of linguistic development and the underlying mechanisms of speech reduction. We focus on two types of reductions: first, lexicalized reductions (Bodur et al., 2025), which include frequently occurring reduced forms such as 'chui' [ʃɥi] for '*je suis*' [ʒə sɥi] and [f ɛ̃] for '*enfin*' [ɑ̃f ɛ̃] in French. These reductions have been extensively studied across languages (e.g., '*eigenlijk*' in Dutch [Ernestus & Smith, 2018]) and could be presumed to be stored in speakers' mental lexicon due to their frequent and regular exposure. They are perceptible by both speakers/listeners and even by Automatic Speech Recognition systems (Strik & Cucchiarini, 1999), facilitating the detection process.

Second, a substantial number of reductions in speech are more difficult to document because they often do not conform to specific sequences, lack regularity, and do not follow the same rules as lexicalized reductions. Consequently, we proposed the term 'non-lexicalized reductions' to describe instances of temporal compression within the speech flow. These reductions are influenced by various linguistic and conversational factors, such as phonemic properties, position within the prosodic structure, and articulation rate. Non-lexicalized reductions, as we define them, may include both segmental or syllabic deletions and compression of segment durations without full deletion, and may involve sequences of multiple words that span various morpho-syntactic categories, reflecting a broader linguistic phenomenon (see Figure 1). Within the context of continuous speech, it can be challenging to identify these reductions or determine which units in a given sequence are reduced, as they can be interpreted with the help of other contextual information. However, when heard in isolation, some examples can even be unintelligible.[1] Their production relies on fine motor skills, as increased overlap in speech segments allows for more efficient, compressed articulation. This not only conserves articulatory effort at higher articulation rates but also ensures intelligibility for the listener.

Within this framework, we hypothesize that children aged 7 to 11 will produce lexicalized reductions similarly to adults, although developmental differences might be
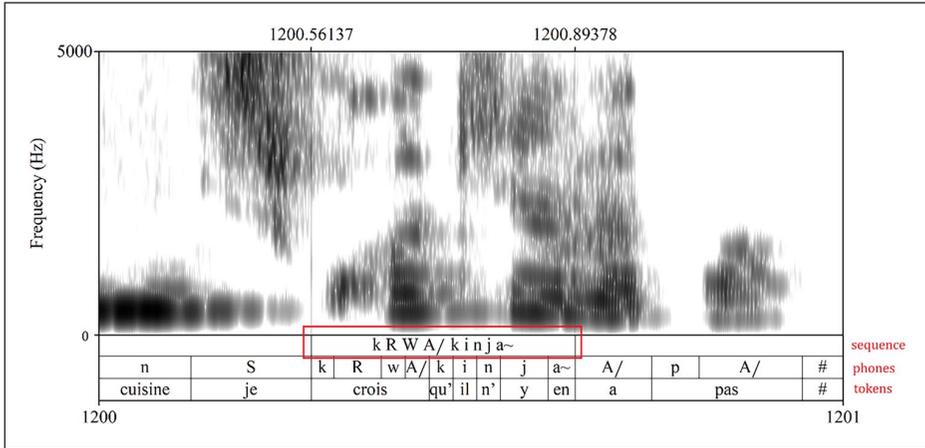
Figure 1. An example of a non-lexicalized reduction (from the speaker ID18 child: the detected sequences (top tier), the phonemes corresponding to the sequence 'je crois qu'il n'y en' [ʒə kʁwa kil ni ɛn], 'I don't think there are any..') (middle tier), and the tokens involved in the detected sequence (bottom).
Note. The sequence is made up of nine phonemes produced in 330 ms.

observed across age groups (older children producing more lexicalized reductions than younger ones) due to differences in language production. Furthermore, we expect lexicalized reductions to be more prevalent in children's speech than non-lexicalized reductions, as the latter require greater neuro-motor control to produce reduced sequences at higher speech rates while maintaining intelligibility for the listener – an ability that typically does not mature until early adolescence (Lee et al., 1999). Additionally, we hypothesize that caregivers, aiming to support children's linguistic and conversational development, will produce fewer non-lexicalized reductions due to slower articulation rates. Measuring the number of phonemes produced per second (excluding pauses), articulation rate serves as an indicator of developmental progress, typically exhibiting higher values in adults compared to children. In exploring how articulation rates vary across age groups, we hypothesize that developmental differences will manifest in articulation rates, with a correlation expected between articulation rates and reductions. Finally, we also hypothesize that the reduction ratios and counts between children and caregivers would be positively correlated. This expectation is based on the notion that children often learn language patterns from their caregivers. As a result, communicative behaviors, including speech reductions, exhibited by caregivers could influence those of the children, leading to similar reduction patterns between the two.

To investigate reduction in children's conversations, we conducted a corpus analysis using a French child–caregiver conversational corpus. This corpus provides a naturalistic and casual context, which is known to enhance reduction production in adult speakers (Torreira et al., 2010). In the following section, we present the corpus and the methods used to identify the two types of reductions that were analyzed.

Method and materials

Participants

This study included a total of 10 child–caregiver dyads, divided into two age groups for developmental comparison. Five children were 7 years old (2 F and 3 M, average age = 7.3, ±3.3 months), and the other 5 were 11 years old (2 F and 3 M, average age = 11.3, ±4.1 months) at the time of recording. Participants were grouped based on dyads, with each child paired with their corresponding caregiver. For analysis purposes, we refer to these as four groups: C7 (7-year-olds), C11 (11-year-olds), P7 (caregivers of 7-year-olds), and P11 (caregivers of 11-year-olds). This labeling allows for both within-dyad and between-age-group comparisons while maintaining the child–caregiver pairing structure. None of the participants reported speech or hearing disorders.

Dataset

The recordings for this study were retrieved from the Child Interpersonal Communication Analysis (CHICA) corpus (Goumri et al., 2024), which captures child–caregiver interactions in two settings: Zoom video calls and face-to-face interactions. The dataset comprises recordings of children aged 7, 9, and 11 years, engaging in conversations with a caregiver, aiming to obtain a more naturalistic corpus. Recordings took place in participants' homes using an intuitive elicitation task: a loosely structured word-guessing game where interlocutors switched roles upon correctly guessing a word, thereby helping balance the conversational dynamic. For the purpose of this study, only the 7 and 11-year-old groups were included to facilitate a more targeted comparison of developmental effects. Additionally, only the Zoom recordings were analyzed due to their provision of separate mono-channel audio tracks. Each recording lasted approximately 15 min, with the total Zoom dataset comprising about 4 hr of conversation, containing a total of 25,272 word tokens (children = 9171, caregivers = 16,101 words).

The corpus included orthographic transcriptions generated by WhisperX (Bain et al., 2023), which were initially corrected for major errors. To ensure highly accurate phonetic annotations and alignments, the first author then further refined and enriched the transcriptions produced by WhisperX. The correction involved carefully listening to the recordings multiple times and correcting several systematic omissions and inaccuracies. Specifically, WhisperX often ignored conversational elements such as repetitions, disfluencies (e.g., filled pauses like 'uh' or 'um'), and laughter, which are features that significantly impact the performance of phonetic alignment. These missing elements were manually reinserted to preserve the natural structure and rhythm of the speech. Additionally, lexical errors generated by WhisperX were corrected to ensure both semantic coherence and phonetic precision. These refinements were essential not only for accurate segment boundary detection and high-quality phonetic annotation but also for reliable reduction detection, which depends on the precise marking of phonetic events, including pauses and non-lexical vocalizations.

The words and utterances in the dataset were extracted using the SPPAS software (Bigi, 2015). SPPAS automatically detected Interpausal Units (IPUs), which approximate utterances, based on the audio signal. The detection parameters were set to a

minimum utterance duration of 180 ms and a minimum pause duration of 250 ms. Orthographic transcriptions were then aligned with the detected IPUs in Praat (Boersma & Weenink, 2009), with manual corrections made as necessary. Once the IPUs were populated, token normalization was performed in SPPAS to generate tokens, including their start and end points, which were subsequently used for phonetization and automatic alignment.

Detection of reductions

Reduction phenomena, influenced by diverse linguistic and conversational factors, might pose challenges for detection, necessitating distinct procedures. Our study adopts a top-down approach for detecting lexicalized sequences, while non-lexicalized reductions were identified using a bottom-up approach.

The detection of *lexicalized* reductions was performed manually, based on orthographic transcriptions and word-level annotations. The procedure entailed checking the word tier in TextGrid files to identify the occurrences of selected target sequences. Upon locating a target token, the audio–visual signal was manually analyzed using Praat (Boersma & Weenink, 2009) to determine whether the item was produced in its canonical or reduced form. Reduced items were then annotated in a separate tier labeled 'reduction'. For example, one of the most frequently reduced words in French, *enfin* is typically represented as [ɑ̃fɛ̃] in its canonical form and was considered reduced if produced/perceived as [fɛ̃]. Similarly, variations in the canonical form of *parce que* [paʁsə kə] were considered reduced when it was produced as [paskə], [paʁskə], and even [psk] in some extreme cases. Annotations were carried out by a Master's student. To ensure reliability, an inter-annotator agreement assessment was conducted on a sample dataset that mirrored the current recordings (Bodur et al., 2021). This assessment involved two annotators (the student and the first author) labeling reduced occurrences of target sequences in two recordings. Initial annotations yielded a Cohen's Kappa of 0.71 for caregivers and 0.65 for children. Following discussion on reduction criteria, a second set of annotations was conducted, resulting in agreements of 0.85 for children and 0.93 for caregivers, indicating a significant agreement. The target sequences were selected based on previous studies on frequently reduced forms (e.g., Adda-Decker & Snoeren, 2011; Bodur et al., 2022; Wu & Adda-Decker, 2020) and consist of the following 11 items:

- Adverbs: *alors (so), enfin (finally), puis (then), voila ` (there), quand même (anyway)*
- Pronoun+verb: *je peux (I can), je sais (I know), je suis (I am), tu sais (you know)*
- Conjunctions: *parce que (because), puisque (because)*

For *non-lexicalized* sequences, an automated bottom-up approach was chosen for detection, given that this type of reduction does not affect the same tokens regularly but rather is dependent on various factors. The method uses a sliding window of 230 ms duration, applied with a step size of one phone; that is, the window advances by one produced phoneme at each step. A sequence is flagged as reduced if at least six phonemes are detected within the 230 ms window starting from the boundary of the first phone. If

the following window (shifted by one phone) also satisfies this criterion, the reduced region is extended, and this process continues until the condition is no longer met. Importantly, in cases where consecutive windows overlap, only the newly included phones (i.e., those not already part of the previously flagged segment) are added to the growing reduction sequence. As a result, reduction zones can span longer than 230 ms when overlapping windows are continuously flagged as reduced. This method primarily targets and identifies temporally reduced zones in the continuous flow of speech, allowing for the detection of reduction patterns that are not lexically fixed but rather context dependent. Importantly, the approach is based on the duration of actually produced phonetic material rather than the canonical phonological forms. As such, it can highlight both compressions and reductions involving deletions, provided that the output sequence is sufficiently temporally compact. This method thus highlights regions of dense articulation that reflect reduction phenomena in real-time speech. Additionally, the choice of detection parameters, specifically the 230 ms window size and the six-phoneme threshold, was deliberate and strategically designed to capture multi-word sequences. After testing smaller window sizes (180 ms and 200 ms), we settled on 230 ms to capture a sufficient temporal span. This larger window helps identify longer sequences, which is crucial for detecting multi-word reductions, and reduces the risk of capturing overly short segments that do not reflect the intended context. By setting the phoneme threshold at 6, we further refined the detection to focus on meaningful reductions that span across multiple phonemes, thereby enhancing the method's sensitivity to longer, contextually significant sequences rather than brief, isolated reductions.

After detecting sequences for all of the 20 speakers, results were manually evaluated to ensure quality and reliability. Of the 180 non-lexicalized sequence detections (31 from children and 149 from caregivers), 119 (5 from children and 114 from caregivers) were accurately identified as reduced, resulting in a 66% accuracy rate. Excluded sequences were mainly due to phonetization, alignment, and liaison errors. An example from a detected sequence can be found in Figure 1.[2]

Results

This section presents findings on children's language production, emphasizing metrics such as word count, utterance count, and utterance length. These metrics reveal developmental patterns in language use, offering a foundational understanding of the progression in children's speech with age relative to adult language. Observing differences in language production is crucial for interpreting reduction performance and how it evolves with development. Subsequently, the distributions of speech reductions in children's and caregivers' language are examined and compared. Further analysis investigates the correlation between reductions observed in children's speech and those in caregivers' speech, potentially elucidating influences and learning mechanisms involved. Articulation rates are then assessed for their impact on the distribution of speech reductions, as understanding these rates is essential to evaluating how reductions vary across different speech contexts, contributing to a comprehensive view of language production dynamics during this stage of development. To account for variations in conversation length, all reported metrics were normalized by dividing the total values for each speaker by the duration of

Table 1. Mean word counts (normalized by dividing the total number of words produced by each speaker by the length of their respective conversation lengths) per minute and standard deviations for each group. The range of word counts (per minute) for each group is also provided.

Group	C7	C11	P7	P11
Word range	[27.9–53.6]	[33.5–57.3]	[55.8–79.5]	[57.5–83.9]
Word count	38.9	42.8	70	71.3
SD	11.6	9.9	10.2	10.3

their respective conversations in order to provide more reliable and comparable results. Metrics are presented as rates per minute (except for the IPU length and reduction ratios per item). Table 3 displays the raw (non-normalized) values for each speaker (independent from conversation lengths).

Language production in children and caregivers

Word production. Results show that caregivers (P7 and P11 combined) produced more words ($M=71$ per minute) overall compared to the children (C7 and C11 combined, $M=41$ per minute), indicating that adults generally led the conversation. We evaluated the statistical significance of the difference in speech output between children and caregivers using a paired *t*-test. The results revealed a significant difference in the number of words per minute ($t(9)=-9.06, p<.001$), indicating that caregivers produced a higher word count compared to children. The mean word counts for each group are presented in Table 1.

A Kruskal–Wallis rank sum test[3] was conducted to examine how development influences word production across the groups C7, C11, P7, and P11, identifying a significant overall difference between the groups ($\chi^2(3)=13.87, p=.003$). Post-hoc Dunn test revealed no significant difference between C7 and C11 ($Z=0.32$, *Bonferroni adjusted* $p=1$), indicating no significant difference between the word counts of younger and older children. However, significant differences were observed between C7 and P11 (*Bonferroni adjusted* $p=.02$) and between C7 and P7 ($p=.04$), highlighting a substantial difference in word production between younger children and adults. Additionally, Pearson's product–moment correlation analysis between the word production rates of children and caregivers showed a non-significant correlation ($r=.46, t=1.4796, p=.17$). Therefore, the data do not support the notion that children's word production is systematically related to their caregivers' word production.

IPU production. In CHICA, children produced slightly fewer IPUs per minute ($M=11.21$, $SD=1.97$) compared to caregivers ($M=11.83, SD=1.72$). However, a paired samples *t*-test revealed that this difference was not significant, ($t(9)=-1.12, p=.29$). Detailed group distributions are reported in Table 2.

Pearson's correlation analysis indicated a moderate positive correlation between IPUs of children and caregivers ($r=.56, p=.09$), but this was not statistically significant.

Table 2. Mean IPU counts (normalized by dividing the total number of IPUs produced by each speaker by the length of their respective conversation lengths) per minute and standard deviations for each group. The range of IPU values (per minute) for each group is also provided.

Group	C7	C11	P7	P11
IPU range	[8.32–13.56]	[8.97–14.95]	[10.67–14.62]	[9.90–14.30]
IPU count	11.2	11.2	12.2	11.4
SD	1.95	2.23	1.63	1.87

Note. IPU = Interpausal Units.

IPU length. With no significant difference in IPUs per minute between groups, we analyzed IPU length as a potential indicator of linguistic development across age groups. The average IPU length, calculated as the total word count divided by the number of IPUs per speaker, was 3.66 words per IPU for children and 6.05 words per IPU for caregivers. A paired t-test revealed a significant difference between children and adults ($t(9) = -7.48$, $p < .001$), with an estimated difference of 2.38 words. This suggests that while children produced nearly as many utterances as caregivers, their utterances were significantly shorter. Table 3 details individual distributions while group means are illustrated in Table 4.

A Kruskal–Wallis test analyzed differences in the number of words per IPU across four groups, accounting for speaker variability, and revealed a significant effect of group ($\chi^2(3) = 13.65$, $p = .003$). Post hoc comparisons between C11 and P11 (*Bonferroni adjusted p* = .045) and C7 and P11 (*Bonferroni adjusted p* = .014) were significant after adjustment, while all the other comparisons were not significant ($p \geq .05$). No significant difference was found between C11 and C7 ($Z = 0.37$, *Bonferroni adjusted p* = 1). These results suggest developmental differences in speech production, with adults showing greater verbal output.

Reduction in child–caregiver interactions

Lexicalized reductions. To study lexicalized reductions across age groups, the frequency of total occurrences of target words, their reduced occurrences, and the reduction ratio per item, speaker, and group were analyzed. Reduction ratios were calculated by dividing the number of reduced occurrences by the total occurrences for each item and for each speaker in order to allow comparability despite varying numbers of target and reduced occurrences. A total of 336 target items (based on the 11 selected forms) were produced in the dataset, of which 222 were reduced, resulting in a reduction ratio of 66%.

To ensure comparability, normalized values were used for total and reduced items. Caregivers produced more target words ($M = 1.48$, $SD = 0.7$) and reduced occurrences ($M = 0.95$, $SD = 0.42$) per minute compared to children (Target: $M = 0.84$, $SD = 0.54$; Reduced: $M = 0.57$, $SD = 0.38$). Among children, 7-year-olds produced fewer target tokens ($M = 0.75$) per minute than 11-year-olds ($M = 0.94$). Caregivers of older children (P11) produced slightly more target tokens ($M = 1.59$) than caregivers of younger

Table 3. Total number of words, IPUs, and average IPU length (as words per IPU) for each speaker. Children identified as ID1 through ID5 belong to the 7-year-old group, while those identified as ID15 through ID22 are in the 11-year-old group.

Group	Children										Caregivers									
	C7					C11					P7					P11				
Speaker	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22
Words	513	587	737	754	350	598	698	505	974	442	1027	1299	1103	1007	772	1267	906	866	1311	952
IPUs	153	171	207	164	146	168	182	159	182	131	204	239	202	150	143	155	174	154	201	153
WperIPU	3.3	3.4	3.5	4.6	2.4	3.5	3.8	3.2	5.3	3.4	5	5.4	5.4	7.4	5.4	8.2	5.2	5.6	6.5	6.2

Note. IPU = Interpausal Unit.

Table 4. Mean IPU length (measured as the number of words per IPU) and standard deviations for each group.

Group	C7	C11	P7	P11
IPU length range	[2.4–4.6]	[3.2–5.3]	[5–7.4]	[5.2–8.2]
IPU Length	3.47	3.86	5.74	6.35
SD	0.78	0.86	0.93	1.14

Note. IPU = Interpausal Unit.

children (P7, M=1.37). Similar patterns were observed in reduced word occurrences per minute, with means of 0.43 (C7), 0.71 (C11), 0.85 (P7), and 1.05 (P11).

Reduction ratios per speaker. Our results revealed variability among speakers. For instance, in the 7-year-old group, child ID1 had the highest lexicalized reduction ratio, with 5 reductions out of 8 target tokens (62.5%). In contrast, child ID4 had a reduction ratio of 62% with 18 reductions out of 29 targets. The child with the lowest reduction ratio (ID5) had only 6 target tokens and 2 reductions, resulting in a ratio of 33%. Caregivers also varied in their reduction performance, generally producing more lexicalized reductions overall. The total target and reduction occurrences for each speaker, along with their reduction ratios, are reported in Table 5.

Reduction ratios per age group. Reduction ratios were calculated based on raw total and reduced occurrences. Children had a mean reduction ratio of 0.64 (SD=0.15), while caregivers had a ratio of 0.66 (SD=0.17). A paired sample t-test revealed no significant difference between children and adults (t=−0.36, p=.72). Detailed reduction ratios for each group are presented in Table 6. To assess differences in reduction ratios across age groups, a Kruskal–Wallis rank sum test was conducted, showing no significant difference between groups ($\chi^2(3)$=7.24, p=.06). Post-hoc Dunn's test also found no significant pairwise difference with all Bonferroni adjusted p-values greater than .05, including between C11 and C7 (Z=2.48, p=.07).

Reduction ratios per item. Reduction ratios were computed for each target item to identify which tokens are most frequently reduced in child–caregiver interactions dataset. The items with the highest reduction ratios were '*puis*' and '*de toute façon*', exhibiting a 100% reduction rate (one reduced occurrence out of one total occurrence). Although '*alors*' had the highest number of total and reduced occurrences (Reduced: 69; Target: 120), its reduction ratio was only 57%.

When examining the groups separately, the reduction ratios of items varied between groups. Specifically, within the youngest group (C7), items such as '*je peux*' (two reduction out of two targets), '*parce que*' (five reductions out of five targets), and '*puisque*' (one reduction out of one target) demonstrated a reduction ratio of 100%, while some targets did not occur at all. Conversely, in the 11-year-old group (C11), items like '*je suis*' (four out of four targets) and '*puis*' (one out of one target) achieved a 100% reduction ratio. The other items displayed varying frequencies of reduction. It is important to note that some of the items with 100% reduction ratio occurred only once. Detailed information regarding groups' reduction patterns can be found in Figure 2.

Table 5. Children's and caregivers' production of reduced lexicalized items (LRperMin), the total number of targets (TargetperMin), and the proportion of lexicalized reductions within lexicalized targets (LR Ratio %).

Group	Children										Caregivers									
	C7					C11					P7					P11				
Speaker	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22
LRperMin	0.27	0.06	0.39	1.28	0.16	0.46	0.98	0.86	0.68	0.56	0.32	0.97	0.45	1.35	1.15	1.32	1.31	1.32	0.34	0.96
TargetperMin	0.43	0.12	0.65	2.06	0.49	0.72	1.39	0.99	0.88	0.72	0.92	1.28	0.72	1.56	2.38	1.72	2.62	1.85	0.39	1.36
LR Ratio (%)	62	50	60	62	16	63	70	86	77	77	35	76	63	86	48	76	50	71	87	70

Table 6. Mean reduction ratios for each age group in the dataset as well as their respective standard deviations.

Group Reduction range	C7 [0.33−0.62]	C11 [0.63−0.86]	P7 [0.35−0.86]	P11 [0.50−0.87]
Reduction ratio	0.53	0.75	0.62	0.71
SD	0.12	0.08	0.20	0.13
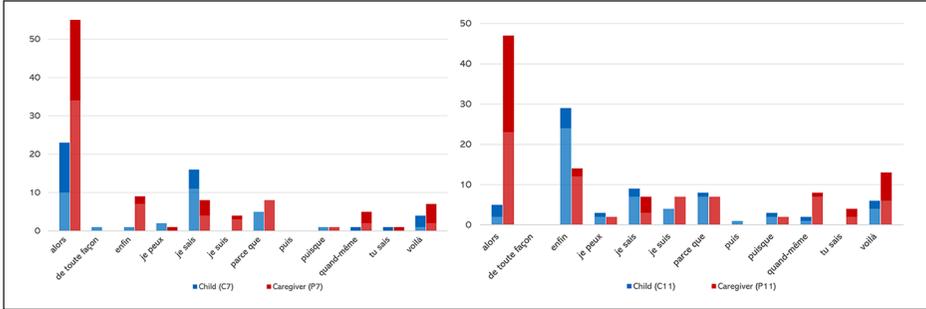
Figure 2. Distributions of target and reduced lexicalized items.
Note. The left panel shows the total number of target items for children (blue) and caregivers (red) in the 7-year-old group, while the right panel shows the totals for the 11-year-old group. Lighter shades indicate the number of reduced occurrences within the total targets.

Non-lexicalized reductions. Manually verified detections of the bottom-up method (described in section 'Detection of reductions') were analyzed to study the distribution of non-lexicalized reductions. This approach enabled assessing the frequency of non-lexicalized reductions across different age groups. As shown in Figure 3, the results reveal that only two children exhibited (a small number of) reductions (0.07 reductions per minute for ID4 (7 years old) and 0.14 for ID18 (11 years old), with no detections in the other children. Although the system flagged many potential reductions, errors in phonetization and alignment, often due to pronunciation variations, led to a high rate of incorrect detections for children. In contrast, caregivers displayed a wider range of non-lexicalized reduction patterns, with some speakers showing no reductions (ID5) and others reaching as high as 2.38 reductions per minute (ID15, 36 sequences in total). An example of a detected non-lexicalized sequence (from ID18, child) is presented in Figure 1. Statistical analysis revealed meaningful differences in the production of non-lexicalized reductions across speaker groups. A paired *t*-test comparing children and their caregivers showed that children produced remarkably fewer instances of non-lexicalized reductions ($t(9) = -3.28$, $p = .01$). A Kruskal-Wallis test revealed a significant difference in non-lexicalized reduction ratios across groups ($\chi^2(3) = 12.24$, $p = .006$), probably driven by higher reduction instances of caregivers. Post-hoc Dunn tests confirmed both child groups differed significantly from the P11 caregiver group, while no significant differences were found between the two caregiver groups (P11 ($M = 0.92$) and P7 ($M = 0.52$) $Z = 0.78$, *Bonferroni adjusted p = 1*) or between the child groups (C7 ($M = 0.01$) and C11 ($M = 0.03$) (*Bonferroni adjusted p = 1*). These results indicate that the observed group effect reflects a robust child–caregiver difference rather than an effect of age.
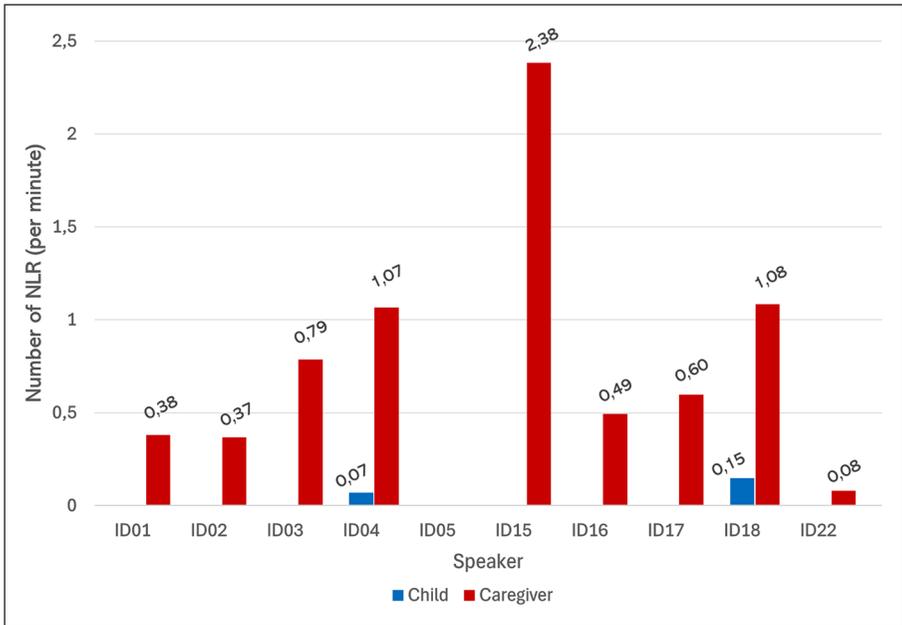
Figure 3. The number of detected non-lexicalized sequences per minute reported for children (blue) and caregivers (red) of each dyad.

Furthermore, the total number of phonemes and the number of reduced phonemes per minute were calculated for all speakers. These values were then used to determine the ratio of reduced phonemes. This approach allowed for accounting for inter-speaker variability in the number of detected sequences while controlling for the influence of articulation rate on the representation of non-lexicalized reductions. Caregivers exhibited a notable range of variability in their reduced phoneme ratios, spanning from 0% to 9%. On the other hand, children produced significantly fewer reduced phonemes, despite variations in their total phoneme counts, likely influenced by conversation length. This difference could suggest that children, still developing their linguistic abilities, did not employ reduced phonemes as frequently as adults did. Table 7 summarizes the number of total and reduced phonemes, in addition to the reduced phoneme ratios of speakers, highlighting the difference between children and caregiver groups.

The relationship between child–caregiver reductions

A Pearson's product–moment correlation was performed to evaluate the influence of caregivers' reductions on the frequency of children's reductions, considering both lexicalized and non-lexicalized types. The analysis revealed a moderate positive correlation between children and caregivers for lexicalized reduction ratios ($r=.33$, $p=.34$, $df=8$). Separate analyses for the two child groups (C7 and C11) revealed different levels of correlation with the P7 and P11 groups. For the C7-P7 dyads, the correlation was weak

Table 7. The number of reduced (PhoNLR) and total phonemes (PhonTot) per minute, along with the percentage of reduced phonemes relative to the total for each speaker (ratio %).

	Children										Caregivers										
Group	C7				C11						P7					P11					
Speaker	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22
PhoNLR	0	0	0	1.1	0	0	0	0	1	0	2.6	2.4	5.3	8.7	0	20.2	3.6	4.2	7.8	0.7
PhonTot	74	87	118	133	73	93	144	84	119	85	149	202	191	194	161	205	194	157	163	194
Ratio (%)	0	0	0	0.8%	0	0	0	0	0.8%	0	1.7%	1.1%	2.7%	4.4%	0	9.8%	1.8%	2.6%	4.7%	0.3%

Table 8. Articulation rates, measured as the number of phonemes produced per second for all speakers. Children with IDs 1 to 5 belong to the 7-year-old group, while those with IDs 15 to 22 are in the 11-year-old group.

Speaker	ID1	ID2	ID3	ID4	ID5	ID15	ID16	ID17	ID18	ID22
Children	6.6	6.5	7.5	7.2	6.2	7.2	7.8	8.7	8.6	7.5
Caregivers	8.6	9.7	10.5	10.4	8.2	11.2	8.7	9.1	10.2	8.9

($r=.20$, $p=.74$), lacking significance. The C11-P11 dyads also exhibited a weak positive correlation ($r=.15$, $p=.8$), which also failed to reach statistical significance. The analysis of non-lexicalized reductions (per minute) showed a non-significant correlation coefficient of $r=.25$ ($p=.48$, $df=8$), indicating a weak positive relationship, suggesting that higher frequencies of caregiver non-lexicalized reductions may be associated with higher child frequencies, though not significantly.

Articulation rate and reductions

Our study examined variations in articulation rates across different age groups and their relationship with speech reductions. Articulation rates for all speakers are presented in Table 8. A statistically significant difference in articulation rates was observed between children and caregivers, with caregivers demonstrating higher speaking rates. Kruskal–Wallis non-parametric analysis indicated significant group differences in articulation rates ($\chi^2(3)=13.32$, $p=.004$), primarily driven by the disparities between children and caregivers. However, no significant difference was found between the two-child age groups, C11 and C7 ($Z=1.28$, $p=1$). These results suggest that while age did not significantly influence articulation rates within these specific child groups, caregivers generally exhibited higher articulation rates compared to children.

A Pearson's correlation assessed the relationship between articulation rates of children and caregivers. A strong positive correlation was found between the C7 and P7 groups ($r(3)=.878$, $p=.05$), indicating that higher articulation rates in children are associated with higher rates in their caregivers (95% CI [−0.017, 0.992]) for this group. In the 11-year-old group, the correlation was weak and non-significant ($r(3)=−.230$, $p=.709$, [−0.925, 0.817]), indicating no reliable association and suggesting potential independence or other influencing factors.

Another Pearson correlation was conducted between articulation rate and lexicalized reduction ratio, revealing a significant moderate positive correlation ($r(18)=.55$, $p=.012$). For non-lexicalized reductions, the correlation showed a strong positive relationship ($r(18)=.80$, $p<.001$) with a confidence interval of [0.560, 0.919]. These significant relationships suggest that higher articulation rates are associated with higher reduction scores.

Discussion

This study analyzed speech reduction behaviors in children aged 7 to 11 using the CHICA corpus (Goumri et al., 2024), focusing on child–caregiver interactions. We examined

language production and speech reduction patterns across age groups (7 and 11 years, and caregivers) to understand the occurrence of reductions. Our goals were to investigate the distribution of lexicalized and non-lexicalized reductions, compare these to adult performance, and assess the influence of caregiver input and articulation rate on the production of reduced sequences.

The dataset comprised Zoom-recorded interactions between dyads engaged in a word-guessing game during video calls. This choice was motivated by several factors. First, there is a scarcity of studies and multimodal datasets focusing on middle childhood, with only a few existing examples (Bodur et al., 2021; Dorval et al., 1984; Sehley & Snow, 1992). Second, the recordings took place in a home environment, providing a natural and relaxed context expected to encourage informal speech and increase the occurrence of linguistic reductions. Third, the game design promoted more balanced interactions, as participants alternated turns. These factors collectively positioned the dataset as a valuable resource for analyzing naturalistic speech patterns in children.

The observed patterns of lexicalized reductions reveal important insights into language development in children. Caregivers produced more target and reduced words per minute than children, who had lower rates of both. Although older children exhibited slightly more target and reduced items than younger ones, the overall word output between these age groups did not differ significantly. This suggests that any apparent increase in lexicalized reductions in older children may reflect individual variation in word usage rather than a statistically supported developmental difference. Although there was considerable variability among individual speakers (Table 5), no significant differences in reduction ratios were found between children and caregivers, or between the 7-year-old and 11-year-old groups, indicating that there is no clear developmental effect on lexicalized reduction ratios in this context. This absence of age-related differences may also be attributed to the relatively narrow age range sampled. Both age groups fall within middle childhood, a period during which speech and language skills are still developing. Broader age comparisons, such as including preschool-aged children or adolescents, may reveal more distinct developmental trends in reduction use, particularly as speech planning and articulatory efficiency mature. These findings contrast with previous research on adult conversations, which reported higher counts of target and reduced tokens, along with significantly higher reduction ratios, – partly attributed to longer conversation lengths (Bodur et al., 2023). The similarity in reduction ratios between children and caregivers suggests that, in middle childhood, children are already capable of using lexicalized reductions in a manner similar to adults. This capability likely arises from their frequent exposure to such reductions in daily interactions, which they may internalize and store in their mental lexicons. The fact that children produce lexicalized reductions at levels comparable to caregivers highlights their ability to adopt complex linguistic features relatively early in development. Despite producing fewer total words, children appear to mirror adults' lexicalized reduction strategies, suggesting a high level of linguistic competence. Another explanation for the absence of significant differences between children and caregivers may be that caregivers might be adjusting their speech to better accommodate children, potentially standardizing reduction patterns across different age groups. This adjustment might level the linguistic playing field, minimizing developmental differences

and resulting in similar reduction patterns (thus ratios) among both younger and older children.

Additionally, our findings indicate that lexicalized reduction ratios are item-specific and not directly correlated with the frequency of occurrence of these items in the dataset. This aligns with earlier research (Bodur et al., 2022, 2023) and highlights that the attributes of individual words or sequences, along with their usage contexts, may be more crucial in influencing whether a reduction occurs than their frequency alone. For example, the token '*puis*' had a 100% reduction ratio despite appearing only once, while '*parce que*' achieved a 96% reduction with 27 reduced occurrences. In contrast, '*alors*', the most frequently occurring target token with 69 reductions out of 120 targets, had a lower reduction ratio of 57%.

By contrast, non-lexicalized reductions presented a very different pattern. As anticipated, these reductions, which involve more intricate phonological and motor processes, were almost absent in children's speech. Only two children, one 7-year-old and one 11-year-old, demonstrated this type of reduction. The remaining children showed no detectable non-lexicalized reductions using the bottom-up approach with the current parameters. In contrast, caregivers displayed a much broader range of non-lexicalized reductions varying from none (ID5) to 2.38 sequences per minute (ID15). This variability among adults, compared to the limited occurrences in children, highlights the more advanced linguistic and motor control skills required for such reductions. Despite the system identifying a number of potential reductions – especially for children, some were erroneous due to challenges in detecting them accurately, often caused by pronunciation variations. The limited occurrence of non-lexicalized reductions in children can be attributed to the complexity of phonological, conversational, and cognitive-neuromotor skills involved in producing such reductions. Non-lexicalized reductions typically occur in easily predictable (Jurafsky et al., 2001) or less informative speech contexts (when both semantic and prosodic information are comparatively less important), where the reduction of phonetic segments doesn't compromise the listener's understanding. Despite the reduction of many segments both temporally and acoustically – sometimes even leading to the creation of vocoids from the original phonemes –, sufficient acoustic and phonetic information must remain for the sequence to be perceptible and comprehensible within the flow of speech. This process demands phonetic and phonological mastery, which children are still developing. Children, as observed in our data, produced significantly fewer reduced phonemes compared to caregivers, a trend that aligns with prior research showing that phonetic reduction patterns, especially vowel reduction, are not fully mastered until late in linguistic development. For example, unstressed vowels in child speech are longer and less centralized than in adult speech (Redford, 2015). Pollock et al. (1993) similarly found that while unstressed syllable durations decrease over time in children's speech, stressed syllable durations remain constant. The developmental trajectory of speech reduction reflects a gradual mastery of these complex processes, with reductions increasing as children gain linguistic proficiency.

Adults, by contrast, not only reduce phonemes more frequently but also exhibit flexibility in when they reduce, exhibiting both hypo- and hyper-articulation depending on context (Lindblom, 1990), which is a hallmark of mature motor and linguistic control. Children, however, still lack this expertise, as seen in studies where they fail to

emphasize words after being misunderstood or do not systematically mark prosodic focus until around age 8 (Redford & Gildersleeve-Neumann, 2009; Chen, 2011). Our results support these findings, where non-lexicalized reductions among caregivers varied widely, but children displayed almost no non-lexicalized reductions, suggesting that these phonetic strategies are a product of more advanced linguistic and motor control systems, which develop with time and experience.

Children had significantly lower articulation rates, which likely contributed to the fewer non-lexicalized reductions observed in their speech. While lexicalized reductions can be produced even at slower speech rates, non-lexicalized reductions are more prevalent at faster rates, where phonetic compression and segmental overlap are more common. In our data, the frequency of non-lexicalized reductions per minute was strongly correlated with articulation rate, with caregivers primarily driving this effect. This supports the idea that higher speech rates facilitate the kinds of coarticulatory and temporal compression process (Johnson, 2004) underlying non-lexicalized reductions. Interestingly, articulation rate was also correlated with lexicalized reductions, though to a lesser extent. This might suggest that lexicalized reductions are possible to produce at slower rates, around hesitations and disfluencies, as these reductions are identified by the absence of specific phonemes/syllables or phonological changes, rather than the overall speed of speech. In contrast, non-lexicalized reductions are more influenced by overall speech speed and require more advanced coordination of linguistic and motor processes, which children are still developing. Additionally, our fixed-window detection method, validated on adult speech, may not have fully captured reductions in slower child speech, as fewer phonemes are produced per unit of time. This methodological constraint could have led to underestimation of non-lexicalized reductions in children.

In conclusion, our study reveals significant developmental distinctions between lexicalized and non-lexicalized reductions in children's speech in addition to notable differences in language production, reduction performance, and articulation rates of children and caregivers. While children as young as 7 years old demonstrate an ability to produce lexicalized reductions at levels similar to adults, non-lexicalized reductions remain a challenge, possibly requiring more advanced linguistic and motor skills. Lexicalized reductions, being tied to specific lexical items and stored in the lexicon, are likely acquired earlier in development, much like other vocabulary items. In contrast, non- lexicalized reductions involve coordinating multiple linguistic levels across broader speech units, making them more complex and reflective of mature speech planning and execution. These results suggest that while children are capable of adopting adult-like reduction patterns in some cases, they have yet to fully master the more sophisticated reduction strategies that come with advanced language proficiency and motor control.

Limitations and future directions. The present study contributes to the existing literature on speech reduction by examining its occurrence in child–caregiver interactions and shedding light on the largely underexplored phenomenon of reduced speech production in middle childhood. However, certain limitations should be acknowledged. The primary limitation is the small sample size ($n = 10$), which restricts cross-group comparisons. While children produced several lexicalized reductions comparable to those of caregivers, non-lexicalized reductions were notably absent. We attribute this to the articulatory demands of such reductions, which are closely linked to higher speech rates. Additionally,

the absence of age effects in our study may be due to the relatively narrow age range (7–11 years). A larger and more age-diverse sample would offer a clearer picture of developmental trends in speech reduction.

Another consideration is the interaction between articulation rate and reduction detection. Since speakers with higher articulation rates can produce more phonemes per unit of time, our fixed detection window may have underestimated non-lexicalized reductions in children, who typically speak at slower rates. While our methodology has been validated in previous studies on adult speech, future research could explore refinements that account for individual differences in speech rate, potentially adjusting detection parameters to capture reductions more dynamically across different articulation rates. This could provide additional insights into how non-lexicalized reductions emerge in developing speakers.

The study's conversational context is another factor to consider. Since interactions took place over Zoom during a loosely structured game, rather than in spontaneous, face-to-face conversations, the observed reduction patterns – or lack thereof – may have been influenced by the task setting. Although the task was designed to elicit casual, conversational speech through a word-guessing game, its turn-based nature may have nevertheless influenced the occurrence of speech reductions. Such settings may encourage more deliberate and careful articulation and may limit overlapping speech, reduce articulation rate/articulatory effort, and constrain conditions that typically favor reductions. Additionally, conducting the interaction via Zoom may have introduced subtle timing delays and reduced prosodic naturalness, further shaping how speech was produced. While lexicalized reductions may withstand such constraints, non-lexicalized reductions, which often rely on fast and fluid articulation, may have been more affected. Future research incorporating in-person interactions could offer valuable insights into how speech reduction develops in children and how different speaking contexts shape this process.

Despite these limitations and the small dataset, which can make generalization difficult, several key points warrant further investigation. Our findings contribute to the limited research on speech reduction in middle childhood, emphasizing the need for further studies with larger, more diverse datasets, broader age ranges, and different conversational settings to better understand the cognitive and developmental factors influencing reduction behavior.

ORCID iD

Kübra Bodur https://orcid.org/0009-0007-1221-3336

Ethical considerations

Data collection was approved by our university's Ethics Committee and registered by the Data Protection Officer, with all procedures strictly following European and national laws. Using the

standard, non-modifiable consent form, we obtained formal guardian consent for minors, who were informed in age-appropriate language.

Consent to participate

Participants understood that this was an experiment and that anonymized results would be published. All recording devices were non-invasive, and participants could stop recording at any time. Data storage and sharing fully comply with GDPR to protect anonymity, with all transcripts and derived data anonymized before sharing. The ethical approval file is available upon request.

Author contributions

Kübra Bodur: Conceptualization; Formal analysis; Investigation; Methodology; Writing – original draft; Writing – review & editing.
Abdellah Fourtassi : Data curation; Formal analysis; Supervision; Writing – original draft; Writing – review & editing.
Corinne Fredouille : Conceptualization; Methodology; Writing – review & editing.
Christine Meunier : Conceptualization; Supervision; Writing – review & editing.

Funding

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

Private access to raw videos will be possible by other researchers via means that are compliant with GDPR and the laws of the country.

Notes

1. Audio–visual examples of non-lexicalized reductions will be available upon request in the Nakala repository.
2. The utterance containing this reduced sequence is 'Tu peux en mettre dans la cuisine si t'as envie, et puis voilà. C'est tout à fait possible d'en mettre dans la cuisine, mais là, dans notre cuisine, **je crois qu'il n'y en** a pas'. (*You can put some in the kitchen if you want, and that's it. It's totally possible to put some in the kitchen, but here, in our kitchen*, ***I don't think there are any.***)
3. The Kruskal–Wallis test was chosen for across-group comparisons due to its non-parametric nature, eliminating the need for normality assumptions, which is particularly beneficial with the study's small sample size.

References

Adda-Decker, M., & Snoeren, N. D. (2011). Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics*, *39*(3), 261–270. https://doi.org/10.1016/j.wocn.2010.11.011

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*, 505–512. https://doi.org/10.3758/s13423-012-0233-y

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). Whisperx: Time-accurate speech transcription of long-form audio. arXiv preprint arXiv:2303.00747. https://arxiv.org/abs/2303.00747

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: Effects of referent availability. *Journal of Child Language*, *21*(3):623–648. https://doi:10.1017/S030500090000948X

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. https://doi.org/10.1016/j.jml.2008.06.003

Bigi, B. (2015). SPPAS-multi-lingual approaches to the automatic annotation of speech. *The Phonetician: Journal of the ISPhS*, *111*, 54–69.

Bodur, K., Fredouille, C., Rauzy, S., & Meunier, C. (2025). Exploring the nuances of reduction in conversational speech: Lexicalized and non-lexicalized reductions. *Speech Communication*, *17*, 103268. https://doi.org/10.1016/j.specom.2025.103268

Bodur, K., Meunier, C., & Fredouille, C. (2022). Formes réduites en conversation : caractéristiques des séquences et des locuteurs. In *34e Journées d'Etudes sur la Parole* (JEP2022), 202–211. ISCA. https://doi.org/10.21437/JEP.2022-22

Bodur, K., Nikolaus, M., Kassim, F., Prévot, L., & Fourtassi, A. (2021). Chico: A multimodal corpus for the study of child conversation. *Companion publication of the 2021 International Conference on Multimodal Interaction* (pp. 158–163). https://doi.org/10.1145/3461615.3485399

Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer* (version 5.1.13). [Computer software]. http://www.praat.org

Carter, A., & Gerken, L. (2004). Do children's omissions leave traces? *Journal of Child Language*, *31*(3), 561–586. https://doi.org/10.1017/S030500090400621X

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*(5), 1055–1083. https://doi.org/10.1017/S0305000910000541

Cheng, H. Y., Murdoch, B. E., Goozee, J. V., & Scott, D. (2007). Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, *50*(2), 375–392. https://doi.org/10.1044/1092-4388(2007/027)

Clopper, C. G., & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*, (pp. 25–72). De Gruyter Mouton. https://doi.org/10.1515/9783110524178-002

Dorval, B., Eckerman, C. O., & Ervin-Tripp, S. (1984). Developmental trends in the quality of conversation achieved by small groups of acquainted peers. *Monographs of the Society for Research in Child Development*, *49*(2), 1–91. https://doi.org/10.2307/1165872

Ernestus, M., & Smith, R. (2018). 5. Qualitative and quantitative aspects of phonetic variation in Dutch eigenlijk. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation* (pp. 129–163). De Gruyter Mouton. https://doi.org/10.1515/9783110524178-005

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants [Editorial]. *Journal of Phonetics*, *39*(SI), 253–260. https://doi.org/10.1016/S0095-4470(11)00055-6

Gendrot, C., & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: An automated formant analysis from eight languages. *Proceedings of the 16th international congress of phonetic sciences* (pp. 1417–1420). Institute of Phonetics, Saarland University.

Goumri, D. E., Agrawal, A., Nikolaus, M., Vu, H. D. T., Bodur, K., Emmar, E., Armand, C., Mazzocconi, C., Gupta, S., Prévot, L., Favre, B., Becerra-Bonache, L., & Fourtassi, A. (2024). CHICA: A developmental corpus of child-caregiver's face-to-face vs. video call conversations in middle childhood. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino, Italy (pp. 3153–3164).

Gulian, M., & Levelt, C. (2009). An acoustic analysis of child language productions with reduced clusters. *BUCLD 33 Online Proceedings*.

Gulian, M., & Levelt, C. (2011). Temporal Measures of Reduced /sC/-clusters in Toddler Speech: Evidence for a Detailed Lexical Specification. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong, China (pp. 787–790).

Johnson, K. (2004). Massive reduction in conversational American English. *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, Tokyo, Japan (pp. 29–54).

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.13jur

Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience*, *30*(1–2), 103–119. https://doi.org/10.1080/01690965.2013.848989

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular* (*Vol. 3*). University of Pennsylvania Press.

Lahey, M., & Ernestus, M. (2014). Pronunciation variation in infant-directed speech: Phonetic reduction of two highly frequent words. *Language Learning and Development*, *10*(4), 308–327. https://doi.org/10.1080/15475441.2013.860813

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*(11), 1773–1781. https://doi.org/10.1121/1.1918816

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-2037-8_16

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. In G. J. Docherty & D. R. Ladd (Eds.), *Gesture, segment, prosody* (pp. 261–289). Cambridge University Press.

Pater, J., & Barlow, J. A. (2003). Constraint conflict in cluster reduction. *Journal of Child Language*, *30*(3), 487–526. https://doi.org/10.1017/S0305000903005658

Pollock, K. E., Brammer, D. M., & Hageman, C. F. (1993). An acoustic analysis of young children's productions of word stress. *Journal of Phonetics*, *21*(3), 183–203. https://doi.org/10.1016/S0095-4470(19)31332-4

Redford, M. A. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *The Journal of the Acoustical Society of America*, *135*(5), 2952–2963. https://doi.org/10.1121/1.4869820

Redford, M. A. (2015). The acquisition of temporal patterns. In M. A. Redford (Ed.), *The Handbook of Speech Production* (pp. 379–403). https://doi.org/10.1002/9781118584156.ch17

Redford, M. A., & Gildersleeve-Neumann, C. E. (2009). The development of distinct speaking styles in preschool children. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1434–1448.

Sehley, S., & Snow, C. (1992), The conversational skills of school-aged children. *Social Development*, *1*, 18–35. https://doi.org/10.1111/j.1467-9507.1992.tb00132.x

Shockey, L., & Bond, Z. (1980). Phonological processes in speech addressed to children. *Phonetica*, *37*(4), 267–274. https://doi.org/10.1159/000259996

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15. https://doi.org/10.1017/S0305000900005092

Strik, H., & Cucchiarini, C. (1999). Modeling pronunciation variation for ASR: A survey of the literature. *Speech Communication*, *29*(2–4), 225–246. https://doi.org/10.1016/S0167-6393(99)00038-2

Strik, H., van Doremalen, J., & Cucchiarini, C. (2008). Pronunciation reduction: How it relates to speech style, gender, and age. *Proceedings of the Interspeech*, *2008*, 1477–1480. https://doi.org/10.21437/Interspeech.2008-425

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication*, *52*(3), 201–212. https://doi.org/10.1016/j.specom.2009.10.004

Waring, R., Liow, S. R., Eadie, P., & Barbara, D. (2019). Speech development in preschool children: evaluating the contribution of phonological short-term and phonological working memory. *Journal of Child Language*, *46*(4), 632–652. https://doi.org/10.1017/S0305000919000035

Wu, Y., & Adda-Decker, M. (2020). Réduction temporelle en français spontané: ou se cache-t-elle? une étude des segments, des mots et séquences de mots fréquemment réduits. In *Actes de la 6e conférence conjointe Journées d'études sur la Parole* (JEP, 33e édition). Volume 1 : Journées d'études sur la Parole, pages 627–635.

Zellers, M., Schuppler, B., & Clayards, M. (2018). Introduction, or: Why rethink reduction? In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Ed.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation* (pp. 1–24). De Gruyter Mouton. https://doi.org/10.1515/9783110524178-001

6.3. Conclusion to the chapter

This chapter has presented a comprehensive investigation of speech reduction in children during middle childhood. By focusing on naturalistic caregiver-child interactions, this study provided complementary insights into how reduction mechanisms, particularly lexicalized and non-lexicalized forms, emerge and are used during this phase of linguistic development.

First, this study significantly extends the scope of reduction research to include children's speech, a domain that has historically been overshadowed by adult-oriented studies. Contrary to assumptions that reduction mechanisms are fully established by early adolescence (Redford, 2019b), our findings show that children continue to develop advanced reduction forms well into middle childhood. This supports the growing view that speech reduction is a developmentally sensitive phenomenon, shaped by cognitive, phonological, and motor skills' maturation (M. A. Redford, 2015b).

Second, the findings reinforce the importance of differentiating between lexicalized and non-lexicalized reductions (LR and NLR). While children exhibited adult-like patterns for commonly lexicalized reductions, likely due to their frequency and entrenchment in the input, non-lexicalized reductions were rarely observed. This divergence supports the hypothesis that temporal and articulatory reductions involve greater articulatory and cognitive demands and thus emerge later in development (Green et al., 2000; M. A. Redford, 2015a). Such a distinction adds granularity to current theoretical models by suggesting separate acquisition paths for different reduction types, moving beyond binary frameworks of "present" or "absent" reduction behaviors.

Third, the study offers empirical evidence linking articulation rate to reduction behavior. Articulation rate increased with age and positively correlated with overall reduction frequency, underscoring the role of speech motor development in facilitating reduction. This aligns with broader developmental research showing that increased motoric efficiency supports more fluent and reduced speech production over time (Flipsen, 2003; Haselager et al., 1991; M. A. Redford, 2019b). These developmental

increases reflect not only motoric maturation but also growing fluency in speech planning and execution, both crucial for the production of phonetic reductions. Our findings suggest that as children's articulation becomes faster and more efficient with age, their capacity to produce reductions, particularly non-lexicalized forms, might also improve later in their developmental trajectory. This supports the view that reduction is not merely an automatic byproduct of casual speech but is tied to the development of articulatory skill and fluency. Consequently, articulation rate can be seen as a useful proxy for the underlying speech motor control that enables increasingly adult-like reduction behavior.

Fourth, the lack of a strong correlation between children's reduction patterns and those of their caregivers suggests that reduction may not be directly shaped by caregiver input, at least not in a straightforward, imitative manner. This challenges traditional input-driven models (e.g., (Kuhl, 2004)) and supports theories that posit a larger role for internal developmental trajectories in the emergence of spontaneous speech features.

Finally, the observation that 7- and 11-year-olds showed similar patterns for LR but struggled with NLR identifies a developmental milestone in the acquisition of spontaneous speech. This provides a framework for future work examining the longitudinal trajectory of reduction behavior across childhood and adolescence, and how it intersects with broader dimensions of linguistic and cognitive growth.

In sum, this chapter contributes to a more developmentally-informed understanding of speech reduction, offering robust empirical evidence that reduction is not a monolithic or automatic process, but a learned and gradually acquired behavior. It underscores the importance of differentiating between types of reduction, highlights the role of motoric and cognitive factors, and challenges simplistic models of phonetic acquisition. These findings open up new avenues for studying how children become proficient speakers, not only through mastering formal structures of language, but also through acquiring the subtle, context-sensitive phonetic behaviors that define natural, fluent speech.

Chapter 7- General Discussion

This dissertation set out to investigate speech reduction as a complex, structured, and context-sensitive phenomenon, moving beyond the view of reduction as a uniform or purely phonetic process. The primary goal was to understand how and why speech reductions occur in casual, spontaneous conversation, and to explore their interaction with different linguistic domains, including prosody, morphosyntax, and language development. A central hypothesis was that speech reduction would encompass distinct types that are shaped by different cognitive, linguistic, and contextual factors. To address this, the dissertation proposed a novel distinction between two types of reductions: lexicalized and non-lexicalized, each with its own distributional and functional characteristics.

The four studies presented in this thesis provide detailed empirical support for our hypothesis. The first study (Chapter 3) introduced the distinction between what we call lexicalized and non-lexicalized reductions. Lexicalized forms appeared consistently across speakers, were unaffected by articulation rate, and were not tightly linked to contextual variation. These patterns suggested that such forms might be restored in the mental lexicon and retrieved as whole units. In contrast, non-lexicalized reductions (characterized by temporally compressed zones in the speech flow) were relatively more variable. They occurred across different word combinations, showed strong sensitivity to articulation rate, and were influenced by morphosyntactic and prosodic structure. These reductions seem to arise from real-time articulatory planning and reflect more context-dependent processes.

The second study (Chapter 4) examined vowel space properties and their relationship to non-lexicalized reductions. Speakers with smaller vowel spaces and less distinct articulation were more likely to produce temporally reduced sequences, especially when speaking at faster rates. Articulation rate consistently emerged as the strongest predictor of non-lexicalized reductions across multiple models. Importantly, the effect of vowel

space size on non-lexicalized reductions was only significant when articulation rate was included. This points to an interaction between temporal and spatial constraints in speech production. Gender, in contrast, was not a significant predictor once these articulatory factors were accounted for, suggesting that previously observed gender effects may reflect differences in articulation rate rather than inherent gender-based patterns.

The third study (Chapter 5) explored the prosodic positioning of non-lexicalized reductions in French. Reductions occurred most often in the middle of Intonational Phrases (IPs), where prosodic prominence tends to be lower. This supports the idea that prosodically weaker positions favor reduction. Reductions also appeared frequently at the junction between Accentual Phrases (APs), especially when function words were involved. These words were often informationally redundant and prone to being compressed or merged with surrounding material. While the general distribution was consistent, speaker-specific variation was notable. Some speakers reduced far more than others, even when articulation rate was comparable, suggesting that individual speaking styles also play a role.

The fourth and final study (Chapter 6) examined how speech reduction manifests itself in children aged 7 to 11, comparing their production patterns to those of their caregivers during naturalistic interactions. The results revealed a clear developmental contrast between lexicalized and non-lexicalized reductions. Lexicalized reductions -those tied to specific word forms- were already well established in children's speech. Both age groups produced these reductions at rates similar to caregivers, with no significant developmental differences. This suggests that children internalize these forms early, likely through frequent exposure, and store them as part of their mental lexicon. In contrast, non-lexicalized reductions, which require more complex phonetic and prosodic adjustments, were virtually absent in children but varied widely among adults. These reductions correlated strongly with articulation rate, which was significantly lower in children. This supports the view that non-lexicalized reductions demand mature speech planning and motor control, still developing in middle childhood. Taken together, these findings suggest that while lexicalized reductions emerge early, non-lexicalized reductions reflect a later stage of linguistic and articulatory development.

Building on the findings across these four studies, this dissertation advances the claim that speech reduction is not a unitary phenomenon, but a structured, multifaceted process shaped by lexical, prosodic, cognitive, and developmental constraints. A central theoretical proposition that emerges from this work is that not all reductions are of the same kind and therefore cannot be studied through a single analytic lens. To address this heterogeneity, we proposed a heuristic distinction between Lexicalized Reductions (LR) and Non-Lexicalized Reductions (NLR); not as a rigid binary classification, but as a conceptual framework for exploring diverse mechanisms that underlie different types of reduction. This distinction is inherently exploratory and aims to broaden existing perspectives on reduction by acknowledging the multiplicity of forms it may take and the layered sources (representational, contextual, articulatory) that shape their realization, rather than offering a definitive typology. To empirically explore this distinction, we implemented two complementary detection methodologies; a top-down, representation driven approach, and a bottom-up, signal driven approach. Each of these is underpinned by distinct theoretical assumptions about the nature of reduction, which could enrich the ongoing discussions about whether and to what extent reductions are stored, planned, or computed in real-time.

The top-down detection approach is premised on the supposition that certain reduced sequences, especially those that are frequent, perceptually more robust, and pragmatically routinized, may be represented holistically in the mental lexicon, consistent with usage-based theories of language and the exemplar model of phonological representation (Bybee, 2002a; Pierrehumbert, 2002). Importantly, the very fact that native speakers could not only recognize these reductions but also actively suggest plausible reduced variants (see Methodology section for the selection of the 13 target items) points to the existence of pre-compiled or at least highly entrenched representations. This aligns with the usage-based models of language (Bybee, 2002a), where frequently encountered and predictably reduced forms may become conventionalized and stored as exemplar clusters; rich memory traces that encode phonetic detail, frequency, contextual information, and social-indexical cues.

In the exemplar framework, linguistic knowledge emerges from rich episodic traces of previous experiences with language use. Frequent sequences such as [ʃɛpa] for *je ne sais pas* [ʒə nə sɛ pa] may thus become stored as units not merely because of categorical rules

164

but through recurrent exposure to specific phonetic variants across communicative contexts. Indeed, in the exemplar theory (Bybee & Torres, 2008; Pierrehumbert, 2008a), lexical representations are thought to emerge not from abstract rules but from the accumulation of detailed phonetic episodes. Over time, high-frequency, prosodically cohesive sequences (particularly those in functionally redundant or predictable contexts) may form dense exemplar clouds, stabilizing their phonetic shape and facilitating rapid access and production. This model accommodates the presence of reduced forms like [ʃɛpa] in the lexicon without needing to posit categorical reanalysis; rather, lexicalization is a gradient outcome of usage patterns and memory entrenchment. Crucially, this implies that LR can be considered as not just articulatory shortcuts but stored linguistic units, subject to frequency effects, priming, and possibly even semantic bleaching.

Moreover, this perspective does not deny the role of articulatory constraints; rather, it emphasizes that such constraints are rarely the sole explanation for reduction; except perhaps in cases of extreme speech rate. Instead, articulatory pressures often act as a vehicle for expressing other types of constraints such as those related to predictability, attention, or communicative intent. Over time, patterns of phonetic realization, when frequent, predictable and socially shared, can be learned and represented in long-term memory. As Pierrehumbert (2002, 2008b, 2016) has argued, such storage is not constrained to discrete phonemic forms but may still include parametric detail and contextual metadata. Admittedly, there are strong typological tendencies in principles of phonetic realization, often grounded in constraints of motor control and auditory perception. However, the exact relationship between phonological categories and phonetic realizations is language-specific (Mielke, 2008). For example, although there is nearly a universal tendency for vowels to be nasalized before nasal stops, the time course and extent of such coarticulatory effects might vary across languages, and even over time within the same language (Beddor, 2009). What allows these phonetic regularities to shape lexical storage is the capacity of memory to encode gradient, multi-dimensional experiences. According to the episodic model, speech processing is driven not by idealized forms but by a memory system capable of retaining vast, fine-grained information, including time, speaker identity, social context, and emotional valence. Even though much of this process is unconscious, the memory system supports detailed representations of spoken utterances, which over time may give rise to conventionalized lexical items that retain traces of their experiential origins. In the context of lexicalized

reductions, such a mechanism may be conceptualized in two ways: either as a rich accumulation of episodic memories, each encoding temporal, social, and affective information; or as a parametric representation of phonetic distributions, continuously updated with exposure (Pierrehumbert, 2002). While these approaches differ in granularity and mechanism, they share the assumption that gradient phonetic detail is encoded in memory, and that phonological representations interface with more continuous, multidimensional maps of the phonetic space.

Nonetheless, the assumption that a reduced form is stored must be treated with caution. Not all frequent reductions might be necessarily lexicalized; frequency alone is insufficient to guarantee entrenchment or mental representation. Experimental validation, such as perception studies, priming tasks, or neuroimaging, would be required to demonstrate whether specific reduced forms are accessed as wholes rather than assembled on the fly. Some studies suggest that listeners often reconstruct reduced forms online by relying on contextual cues, syntax, and word knowledge (Kemps et al., 2004; Mitterer & Ernestus, 2008). Therefore, while our top-down method is suitable for identifying highly routinized, perceptually accessible reductions, it may overestimate the degree of lexicalization without further psycholinguistic evidence.

In contrast, the bottom-up method implemented in this dissertation provides a complementary perspective, one that is agnostic to lexical status and sensitive instead to finer phonetic and contextual variation. By using an automated, phonetically driven detection algorithm, we identified reduction events that were not easily recognizable to speakers in isolation and often involved syntactically diverse words. These findings challenge the assumption that all reductions are stored forms; rather, they point to a class of reductions -NLRs as we proposed- that are shaped by local processing constraints, especially articulation rate, prosodic boundaries, and morpho-syntactic structure. Previous research has shown that increased articulation rate leads to gestural overlap, segmental undershoot, and increased coarticulation (Aylett & Turk, 2004; Gahl et al., 2012). Our findings support this view, as temporally compressed sequences were significantly more correlated with higher articulation rates. One of the most significant findings of this dissertation, presented in Chapter 4, concerns the role of articulation rate in shaping NLRs in speech. These results challenge the common assumption that segmental cues such as vowel centralization, reduced vowel space, or diminished vowel

distinctiveness are direct or sufficient indicators of reduction. While such acoustic properties have traditionally been regarded as hallmark indicators of phonetic reduction (Johnson, 2004; Lindblom, 1990), our findings show that they do not consistently correlate with the occurrence of temporally compressed sequences that qualify as NLRs. That is, although vowel centralization is often interpreted as a proxy for articulatory undershoot or target weakening, it does not reliably predict the presence of NLRs. This result questions reductionist accounts that rely on segmental metrics in isolation and underscores the need to consider suprasegmental and temporal factors, such as articulation rate, in explaining the emergence of NLRs.

Critically, however, these segmental features do not operate in a vacuum. Our data indicates that while reduced vowel space and diminished vowel contrast were not directly correlated with increased NLR occurrence, they interact systematically with articulation rate, which emerged as a robust predictor of NLR. Specifically, speakers with large vowel spaces and high vowel distinctiveness, traits that are often associated with hyper-articulated, intelligibility-focused speech (Bradlow et al., 1996; Ferguson & Kewley-Port, 2002), may still exhibit elevated rates of NLR when speaking at high articulation rates. This suggests that NLRs arise not from a speaker's phonetic base level of precision, but from dynamic adjustments to speech tempo, where coarticulatory overlap and gestural compression become more likely under increased temporal constraints (Byrd & Choi, 2010). This dissociation underscores an important theoretical point: reduction is not reducible -ironically- to formant shifts. Segmental properties such as vowel space area or vowel centralization must be interpreted within the broader prosodic and temporal structure of the utterance. Therefore, while vowel space area has been linked to speech clarity and intelligibility (Munson & Solomon, 2004), it should not be treated as a causal factor or predictor of reduction in spontaneous speech. It is important to clarify that these features are rather outcomes, or surface manifestations of reduction. In the case of NLRs, which involve temporal compression and articulatory overlap, the observed narrowing of the vowel space probably reflects the phonetic realization of these processes, not an independent factor driving them. In this sense, reduction in vowel space area co-occurs with NLRs but does not predict or explain their presence on its own. These patterns are shaped by broader speech-planning parameters, such as utterance length, lexical predictability, and prosodic boundary strength (Bell et al., 2009; Pluymaekers et al., 2005). This view aligns with critiques of over-reliance on

formant-based metrics in spontaneous speech analysis. As Zellou and Tamminga (2014) argue, the time course and functional role of segmental changes in natural speech are deeply conditioned by language-specific patterns and speech rate variability. Additionally, while motor constraints may yield universal tendencies toward coarticulatory reduction, the manifestation of these patterns is shaped by the language's prosodic and lexical architecture (Mielke, 2008). Consequently, our data advocate for a holistic framework in which reduction is examined not just as vowel undershoot or target loss, but as the emergent outcome of competing demands on planning, intelligibility, and efficiency within connected speech (Hall et al., 2018).

These findings echo and extend calls from Scarborough (2013) and Ernestus and Warner (2011) for a multi-level approach to reduction; one in which formant-based cues are understood as downstream consequences of more global factors like prosodic boundary strength, lexical predictability, or articulation rate. In this model, formant metrics lose their diagnostic power when decoupled from their contextual occurrence.

Moreover, this perspective also aligns with Turk and Shattuck-Hufnagel's (2020) argument that reduction emerges from competition between prosodic structuring, lexical access, and motor timing, rather than as a unidimensional acoustic artifact. In this framework, high articulation rate acts as a rate-dependent trigger for compression, producing NLRs not as byproducts of vowel reduction per se, but as the temporal reorganization of articulatory gestures. This interpretation resonates with findings in the motor control literature, where increased speech rate induces gestural overlap and truncation effects (Gahl et al., 2012), and where coarticulatory dynamics override segmental identity in shaping speech output. Thus, our findings suggest a conceptualization of NLR as fundamentally a prosodically modulated, rate-sensitive phenomenon, rather than a segmentally driven one. In this light, classic vowel reduction markers should be treated not as universal indexes of reduction, but as context-sensitive acoustic tendencies whose functional role depends on broader speech planning parameters. Importantly, this perspective supports a growing body of work that critiques the sufficiency of vowel-based metrics for capturing reduction in spontaneous, connected speech (Smiljanić & Bradlow, 2005).

Interestingly, the NLR detected through bottom-up method were less likely to recur in stable, repeatable forms across speakers and more likely to involve novel or lower-frequency lexical combinations. This pattern suggests that such reductions are probably generated dynamically in response to online production demands. In particular, they appear to arise from real-time articulatory optimization and cognitive load management, where speakers might compress speech to meet timing constraints or maintain fluency under pressure. This aligns with models that emphasize situational and efficiency-based adaptations over lexicalized storage, such as the principle of communicative efficiency (Piantadosi et al., 2011), where speakers economize effort in contexts where predictability or redundancy permits greater phonetic lenition. Moreover, the emergence of NLRs appears to be highly speaker-dependent, both in form and frequency. That is, different speakers showed idiosyncratic reduction patterns, often targeting different word combinations. Such variability suggests that these reductions and their frequency are shaped not only by universal articulatory or cognitive constraints, but also by individual differences in speech planning strategies, articulatory habits, and prosodic preferences. This observation aligns with prior work showing substantial inter-speaker variability in reduction strategies, coarticulatory behavior, and linguistic performance (Gahl & Garnsey, 2004). In this view, NLRs do not represent stored, frequency-driven templates, but rather reflect a speaker-specific recalibration of motor plans in response to contextual and prosodic constraints. Their episodic and transient nature makes them less amenable to lexicalization, and less likely to be perceptually salient or easily recognized across speakers.

A result that further underscores the potential divergence between these two reduction types is that the bottom-up, phonetic signal-based system did not detect a significant number of the reductions identified through the top-down approach. That is, many of the reduced occurrences of sequences proposed by native speakers as plausible or familiar reduced forms did not emerge through the automated detection algorithm. This discrepancy likely stems from a methodological asymmetry: the bottom-up approach was calibrated to detect extreme reductions (e.g., sequences of six phonemes produced in under 230 milliseconds). However, many LRs are shorter and do not meet this strict temporal threshold, which means the bottom-up system would not be able to capture them. Had the detection script been set to identify shorter sequences, for example, two phonemes in 50ms or three in 80ms, it may well have retrieved a greater number of LRs.

This methodological consideration invites caution in interpreting detectability as a direct reflection of cognitive or phonological status. Still, the fact that LRs tend not to exhibit the same degree of segmental compression as NLRs may indicate that they differ not only in origin but also in their typical acoustic and prosodic profiles. Rather than arising from real-time articulatory compression, many LR forms appear to follow more stable, routinized phonetic patterns that do not rely on maximally compressed or segmentally ambiguous realizations. Instead, they may achieve perceptual coherence through regularized, possibly stored templates in the mental lexicon. In that sense, their reduction may be phonologized in ways that make them less rate-sensitive than NLRs, which are more likely to emerge under transient production pressures such as increased articulation rate or prosodic weakening. This pattern supports the idea that LR and NLR reductions often reflect distinct tendencies in the speech production system, though the boundary between them remains methodologically and conceptually porous.

The developmental findings of this dissertation further support the proposed distinction between LR and NLR by highlighting the different acquisition trajectories and cognitive demands associated with each type. Notably, children in our study produced LR forms at rates comparable to adults, suggesting that such reductions are acquired early, likely through frequent exposure and entrenchment in the input. This is consistent with usage-based models of language acquisition, in which high frequency, routinized sequences are internalized as dictionary items through repeated experience (Bybee, 2002b; Cameron-Faulkner et al., 2003). These stored forms require minimal online computation and can be retrieved holistically, making them accessible even to children with still-developing phonological and motor control systems. The exemplar model (Pierrehumbert, 2001) provides a robust theoretical framework here: through exposure to rich, detailed phonetic input, children accumulate token-specific memory traces that, over time, stabilize into conventionalized, frequency-driven forms. This process allows children to reproduce LR variants such as [ʃɛpa] for *je ne sais pas* without needing to derive them from scratch.

In contrast, we suggest that the production of NLR is more cognitively and motorically demanding. Unlike lexicalized forms, which are more likely to be stored as holistic, routinized units, NLRs are more likely to emerge from dynamic articulatory processes

such as gestural overlap, prosodic compression, and context-sensitive variation. Successfully producing such reductions requires fine-tuned temporal coordination and anticipatory planning across segments and words, skills that are known to develop gradually over the course of childhood (Redford, 2015a; Smith & Goffman, 2004). One major developmental constraint is articulation rate: children typically speak more slowly than adults, reflecting not only their emerging motor control but also their relatively limited planning efficiency (Redford, 2014). Given the strong association increased articulation rate and NLR occurrence in our data, this slower tempo may help explain why children struggle to produce NLR forms: the temporal demands that facilitate coarticulatory reduction and prosodic compression often exceed their available cognitive and motor resources.

This raises a broader question: are NLRs merely mechanical by-products of fast speech, or do they reflect higher-order, linguistically structured processes? Our findings support the latter interpretation. The relative absence of NLRs in children's speech suggests that these forms rely on advanced planning capacities and sensitivity to prosodic structure, abilities that are not yet fully available in early development. In contrast, children's successful production of LRs implies that such forms depend less on online computation and more on entrenched representations. This developmental asymmetry reinforces the distinction between stored and dynamically generated reductions and supports the view that speech reduction operates across multiple levels of representation and control. Crucially, this perspective situates reduction as **a learned skill, acquired gradually and shaped by cognitive, motoric, and prosodic development**. The LR/NLR distinction thus provides not only a useful framework for analyzing adult speech but also a developmental lens through which to trace the acquisition of spoken language competence, revealing how growing fluency, rate control, and planning sophistication mediate the emergence of different reduction patterns.

Yet the proposed LR/NLR distinction represents only a starting point; a single dimension in what is likely a far more complex typology of reduction phenomena. Future work should expand this typological framework to encompass other organizing axes, such as sociolinguistic variation (e.g., register, formality, identity-indexing), discourse factors (e.g., information structure, topic continuity, predictability), and cognitive mechanisms. By moving beyond a solely phonetic or articulatory perspective, and

toward a model that incorporates social, structural, and communicative constraints, we gain a more complete picture of reduction as a strategic linguistic process. This broader typology not only enhances our theoretical understanding of language use but also opens avenues for cross-disciplinary integration with fields such as sociolinguistics, psycholinguistics, and pragmatics.

In summary, this dissertation offers both a conceptual and methodological shift in the study of speech reduction. Theoretically, it advances the LR/NLR distinction and reframes reduction as a multi-level, context-sensitive phenomenon, by providing a theoretical lens to interpret different manifestations of speech reduction. Rather than attempting to categorically separate reductions, this framework encourages researchers to consider the spectrum of reduction mechanisms, ranging from stored exemplars to online phonetic optimization. Methodologically, it introduces a two-step detection protocol that scales to naturalistic corpora and accommodates both stored and emergent forms. Empirically, it demonstrates how reduction patterns interface with prosody, segmental properties, and developmental progression. Developmentally, it provides strong evidence that reduction is not innate but acquired, highlighting its dependence on maturing cognitive and motor systems. Finally, this work lays the groundwork for future research in psycholinguistic testing, experimental design, and cross-linguistic comparison. By foregrounding the tension between efficiency and structure, fluency and control, this thesis positions speech reduction as a key to understanding how language is planned, produced, and ultimately learned.

Conclusion

This dissertation advances a reconceptualization of speech reduction as a structured, multi-layered phenomenon that transcends the earlier views of reduction as a purely phonetic or articulatory artifact. Drawing on evidence from the four studies proposed in the thesis, the findings consistently demonstrate that reduction in speech is governed by a constellation of linguistic and cognitive constraints, including lexical representation, prosodic structuring, and developmental maturation. Rather than occurring randomly or solely in response to articulatory ease, reduction patterns systematically aligned with utterance position, lexical frequency and routinization, and the gradual acquisition of speech planning mechanisms in childhood. These results converge to support the view that reduction reflects a form of linguistic planning embedded within the broader architecture of speech production (Aylett & Turk, 2004; Bell et al., 2009; Wagner & Watson, 2010). Crucially, reduction operates at the interface between phonetic realization and higher-order processes such as prosodic encoding, lexical access, and communicative efficiency, thereby challenging reductionist accounts that localize it solely at the phonetic or articulatory level (e.g., Lindblom, 1990). By foregrounding the structured nature of reduction across developmental and contextual dimensions, this dissertation positions it as a revealing locus for understanding the coordination of form and function in spoken language.

Bibliography

Adda-Decker, M., & Lamel, L. (1999). Pronunciation variants across system configuration, language and speaking style. *Speech Communication*, *29*(2), 83–98. https://doi.org/10.1016/S0167-6393(99)00032-1

Adda-Decker, M., Boula de Mareüil, P., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication, 46*, 119–139. https://doi.org/10.1016/j.specom.2005.03.006

Adda-Decker, M., Gendrot, C., & Nguyen, N. (2008). Contributions du traitement automatique de la parole à l'étude des voyelles orales du français. *TAL: traitement automatique des langues, 49*, 13-46.

Adda-Decker, M., & Snoeren, N. (2011). Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics*, *39*(3), 261–270. https://doi.org/10.1016/j.wocn.2010.11.011

Albuquerque, L., Oliveira, C., Teixeira, A., Sa-Couto, P., & Figueiredo, D. (2023). A Comprehensive Analysis of Age and Gender Effects in European Portuguese Oral Vowels. *Journal of Voice*, *37*(1), 143.e13-143.e29. https://doi.org/10.1016/j.jvoice.2020.10.021

Audibert, N., & Fougeron, C. (2012). Distorsions de l'espace vocalique: Quelles mesures? Application à la dysarthrie. *Actes Des 29èmes Journées d'Etudes Sur La Parole/JEP-TALN-RECITAL 2012*, 217–224. https://hal.science/hal-02436294

Aylett, M., & 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*, 31–56. https://doi.org/10.1177/00238309040470010201

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). *WhisperX: Time-Accurate Speech Transcription of Long-Form Audio* (arXiv:2303.00747). arXiv. https://doi.org/10.48550/arXiv.2303.00747

Baines, E., & Howe, C. (2010). Discourse topic management and discussion skills in middle childhood: The effects of age and task. *First Language*, *30*(3–4), 508–534. https://doi.org/10.1177/0142723710370538

Ballard, K., Djaja, D., Arciuli, J., James, D., & van Doorn, J. (2012). Developmental Trajectory for Production of Prosody: Lexical Stress Contrastivity in Children Ages 3 to 7 Years and in Adults. *Journal of Speech, Language, and Hearing Research*, *55*, 1822–1835. https://doi.org/10.1044/1092-4388(2012/11-0257)

Bartkova, K., & Sorin, C. (1987). A model of segmental duration for speech synthesis in French. *Speech Communication*, *6*(3), 245–260. https://doi.org/10.1016/0167-6393(87)90029-X

Bates, S. (1995). *Towards a definition of schwa: An acoustic investigation of vowel reduction in English.* The University of Edinburgh.

Beddor, P. S. (2009). A Coarticulatory Path to Sound Change. *Language*, *85*(4), 785–821. https://doi.org/10.1353/lan.0.0165

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. https://doi.org/10.1017/S004740450001037X

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, *113*(2), 1001–1024. https://doi.org/10.1121/1.1534836

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. https://doi.org/10.1016/j.jml.2008.06.003

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*, *25*, 25–55.

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., & Rauzy, S. (2008). Le CID - Corpus of Interactional Data—Annotation et Exploitation Multimodale de Parole Conversationnelle. *Revue TAL : Traitement Automatique Des Langues*, *49*(3), pp.105-134.

Bigi, B. (2015). SPPAS - MULTI-LINGUAL APPROACHES TO THE AUTOMATIC ANNOTATION OF SPEECH. *The Phonetician. Journal of the International Society of Phonetic Sciences*, *111–112*(ISSN:0741-6164), 54–69.

Binnenpoorte, D., Cucchiarini, C., Boves, L., & Strik, H. (2005). Multiword expressions in spoken language: An exploratory study on pronunciation variation. *Computer Speech & Language*, *19*(4), 433–449. https://doi.org/10.1016/j.csl.2004.11.003

Blache, P., Bertrand, R., Ferré, G., Pallaud, B., Prévot, L., & Rauzy, S. (2017). The Corpus of Interactional Data: A Large Multimodal Annotated Resource. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 1323–1356). Springer Netherlands. https://doi.org/10.1007/978-94-024-0881-2_51

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listenersa). *Journal of the Acoustical Society of America*, *112*(1).

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(4 Pt 1), 2074–2085. https://doi.org/10.1121/1.427952

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3–4), 255–272. https://doi.org/10.1016/S0167-6393(96)00063-5

Brand, S., & Ernestus, M. (2018). Listeners' processing of a given reduced word

pronunciation variant directly reflects their exposure to this variant: Evidence

from native listeners and learners of French. *Quarterly Journal of Experimental

Psychology*, *71*(5), 1240–1259. https://doi.org/10.1080/17470218.2017.1313282

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening.

Quarterly Journal of Experimental Psychology, *64*(1), 74–95.

https://doi.org/10.1080/17470218.2010.499174

Broś, K. (2017). Some remarks on stress and reduction correlations in Spanish. *Linguistica

Copernicana*, *14*, 223–246. https://doi.org/10.12775/LinCop.2017.013

Browman, C. P., & Goldstein, L. (1990a). Gestural specification using dynamically-

defined articulatory structures. *Journal of Phonetics*, *18*(3), 299–320.

https://doi.org/10.1016/S0095-4470(19)30376-6

Browman, C. P., & Goldstein, L. (1990b). Tiers in articulatory phonology, with some

implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in

Laboratory Phonology* (1st ed., pp. 341–376). Cambridge University Press.

https://doi.org/10.1017/CBO9780511627736.019

Browman, C.P., & Goldstein L. (1992) Articulatory Phonology: An

Overview. *Phonetica* 31 December 1992; 49 (3-4): 155–

180. https://doi.org/10.1159/000261913

Bürki, A., Ernestus, M., & 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. https://doi.org/10.1016/j.jml.2010.01.002

Bürki, A., Fougeron, C., Gendrot, C., & Frauenfelder, U. H. (2011). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics*, *39*(3), 279–288. https://doi.org/10.1016/j.wocn.2010.07.003

Bybee, J. (2002a). PHONOLOGICAL EVIDENCE FOR EXEMPLAR STORAGE OF MULTIWORD SEQUENCES. *Studies in Second Language Acquisition*, *24*(2), 215–221. https://doi.org/10.1017/S0272263102002061

Bybee, J. (2002b). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290. https://doi.org/10.1017/S0954394502143018

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge University Press.

Bybee, J., File-Muriel, R., & Napoleão de Souza, R. (2016). Special reduction: A usage-based approach. *Language and Cognition*, *1*, 1–26. https://doi.org/10.1017/langcog.2016.19

Bybee, J., & Torres, R. (2008). Phonological and Grammatical Variation in Exemplar Models. *Studies in Hispanic and Lusophone Linguistics*, *1*(2), 399–414. https://doi.org/10.1515/shll-2008-1026

Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics – the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory Phonology 10* (pp. 31–60). De Gruyter Mouton.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics*, *24*(2), 263–282. https://doi.org/10.1006/jpho.1996.0014

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873. https://doi.org/10.1207/s15516709cog2706_2

Cho, T. (2016). Prosodic Boundary Strengthening in the Phonetics–Prosody Interface. *Language and Linguistics Compass*, *10*(3), 120–141. https://doi.org/10.1111/lnc3.12178

Clopper, C. G. (2004). *Linguistic experience and the perceptual classification of dialect variation*. Indiana University.

Clopper, C. G., & Turnbull, R. (2018). 2. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking Reduction* (pp. 25–72). De Gruyter. https://doi.org/10.1515/9783110524178-002

Cole, J., & Shattuck-Hufnagel, S. (2016). New Methods for Prosodic Transcription: Capturing Variability as a Source of Information. *Laboratory Phonology*, *7*(1), Article 1. https://doi.org/10.5334/labphon.29

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review, 11*(6), 1084–1089. https://doi.org/10.3758/BF03196741

Crystal, T., & House, A. (1982). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America, 72*(3), 705-716.

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General, 152*(3), 864–889. https://doi.org/10.1037/xge0001301

Dilley, L. C., & Pitt, M. A. (2010). Altering Context Speech Rate Can Cause Words to Appear or Disappear. *Psychological Science, 21*(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Duez, D. (1995). On spontaneous French speech: Aspects of the reduction and contextual assimilation of voiced stops. *Journal of Phonetics, 23*(4), 407–427. https://doi.org/10.1006/jpho.1995.0031

Duez, D. (2002). Reduction and assimilatory processes in conversational French speech. Implication for speech synthesis. *Improvement in speech synthesis*, 228-236.

Durand, J., Slater, C., & Wise, H. (1987). *Observations on schwa in southern French. 25*(5), 983–1004. https://doi.org/10.1515/ling.1987.25.5.983

Durand, J., Laks, B., & Lyche, C. (2001). *La phonologie du français contemporain: Usages, variétés et structures* (p. 93). Gunter Narr. https://shs.hal.science/halshs-00166484

Elman, J. L., & Mcclelland, J. L. (1984). Speech Perception as a Cognitive Process: The Interactive Activation Model. In *Speech and Language* (Vol. 10, pp. 337–374). Elsevier. https://doi.org/10.1016/B978-0-12-608610-2.50013-3

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The Recognition of Reduced Word Forms. *Brain and Language*, *81*(1–3), 162–173. https://doi.org/10.1006/brln.2001.2514

Ernestus, M. T. C., & Baayen, R. H. (2007). The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration, and frequency of occurrence.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*(3), 253–260. https://doi.org/10.1016/S0095-4470(11)00055-6

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41. https://doi.org/10.1016/j.lingua.2012.12.006

Ernestus, M., Hanique, I., & Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics*, *48*, 60–75. https://doi.org/10.1016/j.wocn.2014.08.001

Ernestus, M., & Smith, R. (2018). 5. Qualitative and quantitative aspects of phonetic variation in Dutch eigenlijk. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler,

& M. Zellers (Eds.), *Rethinking Reduction* (pp. 129–163). De Gruyter. https://doi.org/10.1515/9783110524178-005

Farnetani, E., & Recasens, D. (1997). Coarticulation and connected speech processes. *The handbook of phonetic sciences, 371*, 404.

Farnetani, E., & Recasens, D. (2010). Coarticulation and Connected Speech Processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (1st ed., pp. 316–352). Wiley. https://doi.org/10.1002/9781444317251.ch9

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *112*(1), 259–271. https://doi.org/10.1121/1.1482078

Fikkert, P. (2006). Developing representations and the emergence of phonology: Evidence from perception and production *. *Laboratory Phonology*, *10*.

Fletcher, A. R., McAuliffe, M. J., Lansford, K. L., & Liss, J. M. (2017). Assessing Vowel Centralization in Dysarthria: A Comparison of Methods. *Journal of Speech, Language, and Hearing Research : JSLHR*, *60*(2), 341–354. https://doi.org/10.1044/2016_JSLHR-S-15-0355

Flipsen, P. (2003). Articulation Rate and Speech-Sound Normalization Failure. *Journal of Speech, Language, and Hearing Research*, *46*(3), 724–737. https://doi.org/10.1044/1092-4388(2003/058)

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic

domains. *The Journal of the Acoustical Society of America*, *101*(6), 3728–3740.

https://doi.org/10.1121/1.418332

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of*

Phonetics, *34*(4), 409–438. https://doi.org/10.1016/j.wocn.2005.08.002

Fourakis, M. (1991). Tempo, stress, and vowel reduction in American English. *The Journal*

of the Acoustical society of America, 90, 1816-1827.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in

speech and listeners' perception and use of the distinction. *Journal of Memory and*

Language, *26*(5), 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Fruehwald, J. (2013). *The phonological influence on phonetic change*. University of

Pennsylvania..

Fuchs, S., Winkler, R., & Perrier, P. (2008). *Do Speakers' Vocal Tract Geometries Shape their*

Articulatory Vowel Space?

Gahl, S., & Garnsey, S. M. (2004). Knowledge of Grammar, Knowledge of Usage: Syntactic

Probabilities Affect Pronunciation Variation. *Language*, *80*(4), 748–775.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density

and phonetic reduction in spontaneous speech. *Journal of Memory and Language*,

66(4), 789–806. https://doi.org/10.1016/j.jml.2011.11.006

Gahl, S., & Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language, 89*, 162–178. https://doi.org/10.1016/j.jml.2015.12.006

Garvey, C., & and Berninger, G. (1981). Timing and turn-taking in children's conversations 1. *Discourse Processes*, *4*(1), 27–57. https://doi.org/10.1080/01638538109544505

Gendrot, C., & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *Interspeech 2005*, 2453–2456. https://doi.org/10.21437/Interspeech.2005-753

Gendrot, C., Adda-Decker, M., & Schmid, C. (2012). Comparaison de parole journalistique et de parole spontanée: Analyses de séquences entre pauses (Comparison of journalistic and spontaneous speech: analysis of sequences between pauses) [in French]. In L. Besacier, B. Lecouteux, & G. Sérasset (Eds.), *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 1: JEP* (pp. 649–656). ATALA/AFCP. https://aclanthology.org/F12-1082/

Goffman, L. (1999). Prosodic Influences on Speech Production in Children With Specific Language Impairment and Speech Deficits. *Journal of Speech, Language, and Hearing Research, 42*(6), 1499–1517. https://doi.org/10.1044/jslhr.4206.1499

Goffman, L. (2004). Kinematic Differentiation of Prosodic Categories in Normal and Disordered Language Development. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1088–1102. https://doi.org/10.1044/1092-4388(2004/081)

Goldinger, S. D. (1996). Auditory Lexical Decision. *Language and Cognitive Processes*, *11*(6), 559–568. https://doi.org/10.1080/016909696386944

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, *105*(2), 251.

Goumri, D. E., Agrawal, A., Nikolaus, M., Vu, H. D. T., Bodur, K., Emmar, E., Armand, C., Mazzocconi, C., Gupta, S., Prévot, L., Favre, B., Becerra-Bonache, L., & Fourtassi, A. (2024). CHICA: A Developmental Corpus of Child-Caregiver`s Face-to-face vs. Video Call Conversations in Middle Childhood. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 3153–3164). ELRA and ICCL. https://aclanthology.org/2024.lrec-main.280/

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination. *Journal of Speech, Language, and Hearing Research : JSLHR*, *43*(1), 239–255. https://doi.org/10.1044/jslhr.4301.239

Greenberg, S. (1999). Speaking in shorthand -A syllable-centric perspective for understanding pronunciation variation. *Speech Communication, 29*, 159-176.

Guy, G. R. (1991a). Contextual conditioning in variable lexical phonology. *Language Variation and Change, 3*(2), 223–239. https://doi.org/10.1017/S0954394500000533

Guy, G. R. (1991b). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change, 3*(1), 1–22. https://doi.org/10.1017/S0954394500000429

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard, 4*(s2). https://doi.org/10.1515/lingvan-2017-0027

Hanique, I., Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e linguaggio, Rivista semestrale*, 2/2012, 147-164, https://doi.org/10.1418/38783

Haselager, G. J. T., Slis ,I. H., & Rietveld, A. C. M. (1991). An alternative method of studying the development of speech rate. *Clinical Linguistics & Phonetics, 5*(1), 53–63. https://doi.org/10.3109/02699209108985502

Hess, L. J., & Johnston, J. R. (1988). Acquisition of back channel listener responses to adequate messages. *Discourse Processes, 11*(3), 319–335. https://doi.org/10.1080/01638538809544706

Himmelmann, N. P., & Ladd, D. R. (2008). Prosodic Description: An Introduction for Fieldworkers. *Language Documentation, 2*(2).

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256. https://doi.org/10.1017/S0954394509990093

Jakobson, R., & Halle, M. (1956). Phonology and phonetics. *Fundamentals of Language*, 1-51.

Johnson, K. (2004). Massive reduction in conversational American English. *Proceedings Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, (pp. 29-54).

Jun, S.-A., & Fougeron, C. (1995). The Accentual Phrase and the Prosodic Structure of French. *Proceedings of the 13th International Congress of Phonetic Sciences*, 2, 722–725. https://hal.science/hal-02436320

Jun, S.-A., & Fougeron, C. (2000). A Phonological Model of French Intonation. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 209–242). Springer Netherlands. https://doi.org/10.1007/978-94-011-4317-2_10

Jurafsky, D., Bell, A., Gregory, M. & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Ed.), *Frequency and the Emergence of Linguistic Structure* (pp. 229-254). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.45.13jur

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*(1), 117–127. https://doi.org/10.1016/S0093-934X(03)00425-5

Kilbourn-Ceron, O., Clayards, M., & Wagner, M. (2020). Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations. *Laboratory Phonology*, *11*(1), Article 1. https://doi.org/10.5334/labphon.168

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Kohler, K. J. (1990). Segmental Reduction in Connected Speech in German: Phonological Facts and Phonetic Explanations. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 69–92). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_4

Kohler, K. J. (2009). The Phonetics/Phonology Issue in the Study of Articulatory Reduction. *Phonetica*, *48*(2–4), 180–192. https://doi.org/10.1159/000261883

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Labov, W. (1966). The linguistic variable as a structural unit.

Labov, W. (1986). THE SOCIAL ORIGINS OF SOUND CHANGE1. In H. B. Allen & M.

D. Linn (Eds.), *Dialect and Language Variation* (pp. 524–541). Academic Press.

https://doi.org/10.1016/B978-0-12-051130-3.50044-6

Ladd, D. R. (2006). "Distinctive phones" in surface representation. In L. Goldstein, D. H.

Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 3–26). Mouton de Gruyter.

https://doi.org/10.1515/9783110197211.1.3

Lau, P. (2008). The Lombard Effect as a Communicative Phenomenon. *UC Berkeley

Phonology Lab Annual Reports*, *4*. https://doi.org/10.5070/P719J8J0B6

Lieberman, P. (1963). Some Effects of Semantic and Grammatical Context on the

Production and Perception of Speech. *Language and Speech, 6(3), 172-187.*

https://doi.org/10.1177/002383096300600306

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the acoustical

society of America, 35*(11), 1773-1783. https://doi.org/10.1121/1.1918816

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech

Production and Speech Modeling, 55*, 403-439. Springer Netherlands.

https://doi.org/10.1007/978-94-009-2037-8_16

Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers.

The Journal of the Acoustical Society of America, *49*(2B), 606–608.

https://doi.org/10.1121/1.1912396

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36.

McCarthy, J. J. (2008). The gradual path to cluster simplification. *Phonology*, *25*(2), 271–319. https://doi.org/10.1017/S0952675708001486

McLeod, S., Doorn, J. V., & Reed, V. A. (2001). Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology*, *10*(2), 99–110. https://doi.org/10.1044/1058-0360(2001/011)

Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., & Le Besnerais, M. (2003). Production and perception of vowels: Does the density of the system play a role? *Proceedings of International Congress of Phonetic Sciences (ICPhS)*, 723–726. https://hal.science/hal-00142978

Michelas, A., & D'Imperio, M. (2010). Durational Cues and Prosodic Phrasing in French: Evidence for the Intermediate Phrase. *Proceedings of Speech Prosody*, 4. https://hal.science/hal-00463205

Mielke, J. (2008). *The Emergence of Distinctive Features*. Oxford University Press UK.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of Americe, 96*(1), 40-55. https://doi.org/10.1121/1.410492

Munson, B., & Solomon, N. P. (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1048–1058. https://doi.org/10.1044/1092-4388(2004/078)

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language*, 22(2), 171–184. https://doi.org/10.1016/j.csl.2007.07.003

Niebuhr, O., & Kohler, K. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics, 39*, 319-329. https://doi.org/10.1016/j.wocn.2010.12.003

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. *Papers in laboratory phonology II: Gesture, segment, prosody*, 261-280.

Ohala, J. J. (1993). Coarticulation and Phonology. *Language and Speech*, *36*(2–3), 155–170. https://doi.org/10.1177/002383099303600303

Pardo, J. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00559

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Parrell, B., & Narayanan, S. (2018). Explaining coronal reduction: Prosodic structure and articulatory posture. *Phonetica, 75*(2), 151-181.

Peterson, G. E., & Lehiste, I. (1960). Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America*, *32*(6), 693–703. https://doi.org/10.1121/1.1908183

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. https://doi.org/10.1073/pnas.1012551108

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking Clearly for the Hard of Hearing II. *Journal of Speech, Language, and Hearing Research*, *29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

Pierrehumbert, J. B. (2008a). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137–158). John Benjamins Publishing Company. https://www.degruyterbrill.com/document/doi/10.1075/tsl.45.08pie/html

Pierrehumbert, J. B. (2008b). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101–140). De Gruyter Mouton.

Pierrehumbert, J. B. (2016). Phonological Representation: Beyond Abstract Versus Episodic. *Annual Review of Linguistics*, *2*(Volume 2, 2016), 33–52. https://doi.org/10.1146/annurev-linguistics-030514-125050

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, *13*(1–2), 109–125. https://doi.org/10.1016/0167-6393(93)90063-q

Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, *39*(3), 304–311. https://doi.org/10.1016/j.wocn.2010.07.004

Pluymaekers, M., Ernestus, M., & Baayen, R. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America, 118*(4), 2561-2569. https://doi.org/10.1121/1.2011150

Polzin, T. S., & Waibel, A. (1998, May). Pronunciation variations in emotional speech. In *Proc Tutorial and Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition (MPV)* (pp. 103-108).

Prasad, R., Matsoukas, S., Kao, C.-L., Ma, J. Z., Xu, D.-X., Colthurst, T., Kimball, O., Schwartz, R., Gauvain, J.-L., Lamel, L., Schwenk, H., Adda, G., & Lefevre, F. (2005). The 2004 BBN/LIMSI 20xRT English conversational telephone speech recognition system. *Interspeech 2005*, 1645–1648. https://doi.org/10.21437/Interspeech.2005-539

Priva, U. C. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, *6*(2), 243–278. https://doi.org/10.1515/lp-2015-0008

Ranbom, L., & Connine, C. (2007). Lexical representation of phonological variation. *Journal of Memory and Language*, *57*, 273–298. https://doi.org/10.1016/j.jml.2007.04.001

Ranbom, L., Connine, C., & Yudman, E. (2009). Is Phonological Context Always Used to Recognize Variant Forms in Spoken Word Recognition? The Role of Variant Frequency and Context Distribution. *Journal of Experimental Psychology. Human Perception and Performance*, *35*, 1205–1220. https://doi.org/10.1037/a0015022

Redford, M. (2014). The perceived clarity of children's speech varies as a function of their default articulation rate. *The Journal of the Acoustical Society of America*, *135*, 2952. https://doi.org/10.1121/1.4869820

Redford, M. A. (2015a). The Acquisition of Temporal Patterns. In M. A. Redford (Ed.), *The Handbook of Speech Production* (1st ed., pp. 379–403). Wiley. https://doi.org/10.1002/9781118584156.ch17

Redford, M. A. (2015b). *The Handbook of Speech Production*. John Wiley & Sons.

Redford, M. A. (2019a). Speech Production From a Developmental Perspective. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 2946–2962. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0130

Redford, M. A. (2019b). Speech Production From a Developmental Perspective. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 2946–2962. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0130

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant Centralization Ratio: A Proposal for a New Acoustic Measure of Dysarthric Speech. *Journal of Speech, Language, and Hearing Research*, *53*(1), 114–125. https://doi.org/10.1044/1092-4388(2009/08-0184)

Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics*, *41*(6), 491–508. https://doi.org/10.1016/j.wocn.2013.09.004

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2008). Preparing a corpus of dutch spontaneous dialogues for automatic phonetic analysis. *Interspeech 2008*, 1638–1641. https://doi.org/10.21437/Interspeech.2008-455

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*(1), 96–109. https://doi.org/10.1016/j.wocn.2010.11.006

Schuppler, B., Van Dommelen, W. A., Koreman, J., & Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies

at the phonemic and sub-phonemic level. *Journal of Phonetics*, *40*(4), 595–607. https://doi.org/10.1016/j.wocn.2012.05.004

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*(3 Pt 1), 1677–1688. https://doi.org/10.1121/1.2000788

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Language and Linguistics Compass*, *3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Snoeren, N. D., Hallé, P. A., & Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics*, *34*(2), 241–268. https://doi.org/10.1016/j.wocn.2005.06.001

Stemberger, J. P. (2004). Neighbourhood effects on error rates in speech production. *Brain and Language*, *90*(1), 413–422. https://doi.org/10.1016/S0093-934X(03)00452-8

Strik, H., & Cucchiarini, C. (1999). Modeling pronunciation variation for ASR: A survey of the literature. *Speech Communication, 29*, 225-246. https://doi.org/10.1016/S0167-6393(99)00038-2

Strik, H., Elffers, A., Bavcar, D., & Cucchiarini, C. (2006). Half a word is enough for listeners, but problematic for ASE. *ISCA workshop on speech.* Toulouse, France: ISCA.

Strycharczuk, P., Ćavar, M., & Coretat, S. (2021). Distance vs time. Acoustic and articulatory consequences of reduced vowel duration in Polish. *The Journal of the Acoustical Society of America, 150*, 592-607.

Sumner, M. (2013). A phonetic explanation of pronunciation variant effects. *The Journal of the Acoustical Society of America*, *134*(1), EL26–EL32. https://doi.org/10.1121/1.4807432

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(2). https://doi.org/10.1515/labphon.2011.012

Torreira, F., & Ernestus, M. (2012). Weakening of Intervocalic /s/ in the Nijmegen Corpus of Casual Spanish. *Phonetica*, *69*(3), 124–148. https://doi.org/10.1159/000343635

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, *39*(3), 312–318. https://doi.org/10.1016/j.wocn.2010.12.001

Tucker, B. V., & Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon*, *11*(3), 375–400. https://doi.org/10.1075/ml.11.3.03tuc

Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Phonetics, and Speech Motor Control*. Oxford University Press.

Van Bergem, D. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication, 12*(1), 1-23.

Van Bergem, D. (1995). Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress. *Speech communication, 16*, 329-358.

Van de Ven, M., Ernestus, M., & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology, 3*(2). https://doi.org/10.1515/lp-2012-0020

Van Son, R. J. J. H., & Pols, L. C. W. (1995). What does consonant reduction look like, if it exists? *4th European Conference on Speech Communication and Technology (Eurospeech 1995)*, 1909–1912. https://doi.org/10.21437/Eurospeech.1995-466

Van Son, R., & Pols, L. (1999). An acoustic description of consonant reduction. *Speech Communication, 28*(2), 125-140. https://doi.org/10.1016/S0167-6393(99)00009-6

Van Son, R., Middag, C., & Demuynck, K. (2018). Vowel Space as a Tool to Evaluate Articulation Problems. *Interspeech 2018*, 357–361. https://doi.org/10.21437/Interspeech.2018-68

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America, 84*(3), 917–928.

Vasilescu, I., Adda-Decker, M., Lamel, L., & Hallé, P. (2009). A perceptual investigation of speech transcription errors involving frequent near-homophones in French and

american English. *Interspeech 2009*, 144–147. https://doi.org/10.21437/Interspeech.2009-53

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From Babbling to Speech: A Re-Assessment of the Continuity Issue. *Language*, *61*(2), 397–445. https://doi.org/10.2307/414151

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945. https://doi.org/10.1080/01690961003589492

Warner, N., & Tucker, B. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America, 130*, 1606-1617. https://doi.org/10.1121/1.3621306

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, *91*(3), 1707–1717. https://doi.org/10.1121/1.402450

Wu, Y., & Adda-Decker, M. (2020). Réduction temporelle en français spontané : où se cache-t-elle ? Une étude des segments, des mots et séquences de mots fréquemment réduits (). In *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL,*

22e édition). Volume 1 : Journées d'Études sur la Parole, pages 627–635, Nancy, France. ATALA et AFCP.

Wu, Y., & Adda-Decker, M. (2021). Réduction des segments en français spontané: apports des grands corpus et du traitement automatique de la parole. *Corpus*(22).

Yao, Y., & Meunier, C. (2014). Effects of phonological neighborhood density on phonetic variation: The curious case of French. *Conference on Laboratory Phonology (LABPHON) 14*, non paginé. https://hal.science/hal-01507637

Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, *47*, 18–35. https://doi.org/10.1016/j.wocn.2014.09.002

ANNEX

The words detected through the bottom-up approach are listed. For each word, we report the number of times it was detected in the reduced sequences, as well as its total frequency in the corpus (CID), resulting in its corresponding reduction ratio.

Word	Times in reduced sequences	Total occurrence in corpus	Ratio
accord	1	1	1
appartenaient	1	1	1
appréciable	1	1	1
arti-	1	1	1
avoi-	1	1	1
boirais	1	1	1
bouquinais	1	1	1
boursier	1	1	1
bracelet	1	1	1
catastrophiste	1	1	1
catégorie	1	1	1
Chéryl	1	1	1
cinoche	1	1	1
construisait	1	1	1
contrôleur	1	1	1
critères	1	1	1
débranché	1	1	1
décharge	1	1	1
démerdé	1	1	1
descriptions	1	1	1
développer	1	1	1
dialogue	1	1	1
dir-	1	1	1
donnais	1	1	1
écrivent	1	1	1
engueulé	1	1	1
entraînement	1	1	1
expliques	3	3	1
factures	3	3	1
fourrure	1	1	1

fout-	1	1	1
fouteuse	1	1	1
galeries	1	1	1
gardait	1	1	1
gestionnaire	1	1	1
impressio-	1	1	1
incidence	1	1	1
inconvénient	1	1	1
inscrivez	1	1	1
introduit	1	1	1
kilomètres	1	1	1
Lafayette	1	1	1
lectures	1	1	1
levait	1	1	1
manuel	1	1	1
Marie-Laure	1	1	1
orthophoniste	1	1	1
perçois	1	1	1
plais-	1	1	1
poss-	1	1	1
prendra	1	1	1
préparation	1	1	1
reçus	1	1	1
regardée	1	1	1
renvoies	1	1	1
repar-	1	1	1
résumer	1	1	1
ro-	1	1	1
science-fiction	1	1	1
signalisations	1	1	1
SNCF	1	1	1
souviendrai	1	1	1
spécialiste	1	1	1
store	2	2	1
surveilles	1	1	1
traînaient	1	1	1
venues	1	1	1
voies	1	1	1
utilisateurs	3	4	0,75
enregistrements	2	3	0,666666667
croyais	7	11	0,636363636
électricité	5	8	0,625
adhésif	1	2	0,5
allocataire	1	2	0,5

animation	1	2	0,5
Arabie	1	2	0,5
assisté	1	2	0,5
bouffait	1	2	0,5
casseroles	1	2	0,5
changement	1	2	0,5
chataîgnes	1	2	0,5
chihuahua	1	2	0,5
considéré	1	2	0,5
construction	1	2	0,5
construisent	2	4	0,5
croyait	1	2	0,5
débrouillait	1	2	0,5
déclarés	1	2	0,5
dépressive	1	2	0,5
dernières	2	4	0,5
descriptive	1	2	0,5
devraient	1	2	0,5
dirigiste	1	2	0,5
écrivait	1	2	0,5
électriques	1	2	0,5
espagnol	1	2	0,5
essaies	2	4	0,5
éteindre	1	2	0,5
expressions	1	2	0,5
extérieur	1	2	0,5
extraterrestre	1	2	0,5
extraterrestres	1	2	0,5
frustrations	1	2	0,5
gigantesque	1	2	0,5
graduate	1	2	0,5
insulte	1	2	0,5
inutile	1	2	0,5
journaliste	2	4	0,5
la_plupart	1	2	0,5
La_Rochelle	1	2	0,5
Las	1	2	0,5
lignes	1	2	0,5
lucarne	1	2	0,5
marchais	1	2	0,5
marseillais	1	2	0,5
mémoire	1	2	0,5
méthodes	1	2	0,5
minotte	1	2	0,5

moyenne	1	2	0,5
néanderthal	1	2	0,5
nouvelles	2	4	0,5
ouvrent	1	2	0,5
parcelles	1	2	0,5
participait	1	2	0,5
particulièrement	1	2	0,5
personnalité	1	2	0,5
pleuré	1	2	0,5
plupart	1	2	0,5
prétexte	1	2	0,5
provençale	1	2	0,5
randonnée	1	2	0,5
reçoit	1	2	0,5
réfugié	1	2	0,5
ren-	1	2	0,5
reparti	1	2	0,5
répondu	1	2	0,5
résiliations	1	2	0,5
restés	1	2	0,5
retrouvait	1	2	0,5
reviennent	1	2	0,5
ruban	1	2	0,5
sèche	1	2	0,5
sélectionné	1	2	0,5
siestes	1	2	0,5
signaux	1	2	0,5
signent	1	2	0,5
surprise	1	2	0,5
trajet	1	2	0,5
trouvés	1	2	0,5
tuyau	1	2	0,5
unité	2	4	0,5
construire	3	7	0,428571429
animateurs	2	5	0,4
devient	2	5	0,4
linguistique	2	5	0,4
Nassima	2	5	0,4
obligés	2	5	0,4
travaillait	2	5	0,4
doivent	3	8	0,375
dois	5	14	0,357142857
accouchement	2	6	0,333333333
administration	1	3	0,333333333

aimé	1	3	0,333333333
aimerais	2	6	0,333333333
boutique	1	3	0,333333333
cités	1	3	0,333333333
comprendre	2	6	0,333333333
construit	1	3	0,333333333
CPE	1	3	0,333333333
débrancher	1	3	0,333333333
devrais	1	3	0,333333333
différentes	1	3	0,333333333
échappé	1	3	0,333333333
engrainé	1	3	0,333333333
enregistres	1	3	0,333333333
envoyaient	2	6	0,333333333
euh-	2	6	0,333333333
expression	1	3	0,333333333
falloir	3	9	0,333333333
généralement	2	6	0,333333333
glisser	1	3	0,333333333
grands-pères	1	3	0,333333333
hiérarchiques	1	3	0,333333333
influence	1	3	0,333333333
interprétation	1	3	0,333333333
loué	1	3	0,333333333
matinée	1	3	0,333333333
occasion	1	3	0,333333333
oublié	2	6	0,333333333
post_graduate	1	3	0,333333333
propriétaire	1	3	0,333333333
psychologue	1	3	0,333333333
rattachement	1	3	0,333333333
recruté	1	3	0,333333333
rectorat	1	3	0,333333333
regardaient	1	3	0,333333333
Rémi	2	6	0,333333333
rentrait	1	3	0,333333333
réveillé	1	3	0,333333333
réveillée	1	3	0,333333333
rouler	1	3	0,333333333
Roxane	1	3	0,333333333
scotch	1	3	0,333333333
sucre	2	6	0,333333333
terminé	1	3	0,333333333
théoriquement	1	3	0,333333333

traité	1	3	0,333333333
utiliser	1	3	0,333333333
voient	1	3	0,333333333
voyez	2	6	0,333333333
avoir	36	112	0,321428571
droit	7	22	0,318181818
plusieurs	5	16	0,3125
appelait	3	10	0,3
aujourd'hui	3	10	0,3
demandait	3	10	0,3
obligé	7	24	0,291666667
première	7	24	0,291666667
enregistrer	2	7	0,285714286
es-	2	7	0,285714286
premières	2	7	0,285714286
régler	2	7	0,285714286
retrouvée	2	7	0,285714286
trouvait	2	7	0,285714286
pouvoir	5	18	0,277777778
crois	32	117	0,273504274
normalement	6	22	0,272727273
allumer	1	4	0,25
balade	1	4	0,25
celui-là	2	8	0,25
césarienne	1	4	0,25
commençais	1	4	0,25
créé	1	4	0,25
cuillère	1	4	0,25
discuté	1	4	0,25
essayait	2	8	0,25
exo-planètes	1	4	0,25
fur_a_mesure	1	4	0,25
habitude	1	4	0,25
in-	1	4	0,25
inscription	1	4	0,25
instruments	1	4	0,25
limite	7	28	0,25
logique	1	4	0,25
maire	1	4	0,25
maîtrise	1	4	0,25
manière	1	4	0,25
maximum	1	4	0,25
oeuvre	1	4	0,25
pare-brise	1	4	0,25

Patrick	1	4	0,25
patronne	1	4	0,25
plombes	1	4	0,25
quarts	1	4	0,25
racheté	1	4	0,25
rapports	1	4	0,25
reçu	1	4	0,25
relativement	1	4	0,25
responsable	1	4	0,25
sieste	1	4	0,25
téléphoné	1	4	0,25
troisième	3	13	0,230769231
logiciel	2	9	0,222222222
personnellement	2	9	0,222222222
pourrait	3	14	0,214285714
pourquoi	11	53	0,20754717
accouché	1	5	0,2
allaient	1	5	0,2
amené	1	5	0,2
argent	1	5	0,2
balancé	1	5	0,2
commençait	1	5	0,2
convoqué	1	5	0,2
écrire	1	5	0,2
envoyer	1	5	0,2
êt-	1	5	0,2
étrangers	1	5	0,2
expériences	1	5	0,2
lycée	1	5	0,2
malin	1	5	0,2
mauvaise	1	5	0,2
nettoyer	1	5	0,2
obligée	1	5	0,2
parlais	2	10	0,2
parlent	2	10	0,2
prénoms	1	5	0,2
profs	1	5	0,2
protéger	1	5	0,2
recherche	2	10	0,2
regardais	1	5	0,2
reviens	1	5	0,2
rigole	1	5	0,2
semblait	1	5	0,2
Sian	1	5	0,2

suis	20	102	0,196078431
doit	13	67	0,194029851
c'est-à -dire	3	16	0,1875
trouvais	3	16	0,1875
foyer	2	11	0,181818182
puisqu'	2	11	0,181818182
raconté	2	11	0,181818182
travaille	2	11	0,181818182
allée	5	30	0,166666667
branché	1	6	0,166666667
combien	3	18	0,166666667
doigts	2	12	0,166666667
donne	2	12	0,166666667
écoutez	2	12	0,166666667
excuse	1	6	0,166666667
films	1	6	0,166666667
marchait	1	6	0,166666667
mercredi	1	6	0,166666667
monter	2	12	0,166666667
par_exemple	10	60	0,166666667
période	1	6	0,166666667
prochaine	1	6	0,166666667
regardé	1	6	0,166666667
rendu	1	6	0,166666667
retraite	1	6	0,166666667
voiture	3	18	0,166666667
vouloir	1	6	0,166666667
lui	53	320	0,165625
reste	5	31	0,161290323
dernière	2	13	0,153846154
directrice	2	13	0,153846154
leurs	2	13	0,153846154
portable	2	13	0,153846154
aurais	4	27	0,148148148
accent	1	7	0,142857143
balader	1	7	0,142857143
boire	1	7	0,142857143
bonnes	1	7	0,142857143
certainement	1	7	0,142857143
collégue	1	7	0,142857143
en_train	5	35	0,142857143
enlevé	1	7	0,142857143
impressionnant	1	7	0,142857143
Maé	1	7	0,142857143

montrer	1	7	0,142857143
parles	2	14	0,142857143
prévu	1	7	0,142857143
remettre	1	7	0,142857143
rentré	1	7	0,142857143
tout_de_suite	1	7	0,142857143
voulez	1	7	0,142857143
met	3	22	0,136363636
voir	15	110	0,136363636
essayé	2	15	0,133333333
mêmes	2	15	0,133333333
acheter	1	8	0,125
appartements	1	8	0,125
cassé	1	8	0,125
commencer	1	8	0,125
connaît	1	8	0,125
essayer	2	16	0,125
étranger	1	8	0,125
intérieur	1	8	0,125
mètres	1	8	0,125
oublier	1	8	0,125
posent	1	8	0,125
prenait	1	8	0,125
programmes	1	8	0,125
rang	1	8	0,125
restaurant	1	8	0,125
aurait	4	33	0,121212121
arrivée	3	25	0,12
acheté	2	17	0,117647059
directeur	2	17	0,117647059
laisse	2	17	0,117647059
viennent	2	17	0,117647059
devait	4	35	0,114285714
adresse	1	9	0,111111111
bien_sûr	2	18	0,111111111
bourses	1	9	0,111111111
conflictuelles	1	9	0,111111111
connaissait	1	9	0,111111111
coupé	1	9	0,111111111
dangereux	1	9	0,111111111
difficile	1	9	0,111111111
donné	2	18	0,111111111
intéresse	1	9	0,111111111
mets	2	18	0,111111111

Myriam	1	9	0,111111111
particulier	1	9	0,111111111
vivre	1	9	0,111111111
voyais	1	9	0,111111111
en_général	2	19	0,105263158
parlait	2	19	0,105263158
aller	9	86	0,104651163
impression	3	29	0,103448276
article	1	10	0,1
existe	1	10	0,1
marqué	1	10	0,1
mort	1	10	0,1
observer	1	10	0,1
perdu	1	10	0,1
premier	4	40	0,1
questions	1	10	0,1
resté	1	10	0,1
table	1	10	0,1
aime	2	21	0,095238095
travailler	2	21	0,095238095
trouve	4	42	0,095238095
complétement	8	87	0,091954023
demandé	1	11	0,090909091
drôle	1	11	0,090909091
fils	1	11	0,090909091
finalement	3	33	0,090909091
guillemets	1	11	0,090909091
madame	2	22	0,090909091
mari	1	11	0,090909091
paraît	1	11	0,090909091
parlé	2	22	0,090909091
personne	4	44	0,090909091
relations	1	11	0,090909091
réunion	1	11	0,090909091
trouver	2	22	0,090909091
voulaient	1	11	0,090909091
pouvait	3	34	0,088235294
absolument	2	23	0,086956522
vais	8	93	0,086021505
etc	6	72	0,083333333
fenêtre	1	12	0,083333333
image	1	12	0,083333333
prises	1	12	0,083333333
rappelles	1	12	0,083333333

repas	1	12	0,083333333
souviens	1	12	0,083333333
vendredi	1	12	0,083333333
alors	32	404	0,079207921
vois	47	606	0,077557756
avec	26	338	0,076923077
d'ailleurs	2	26	0,076923077
étudiants	1	13	0,076923077
histoires	1	13	0,076923077
intérêt	1	13	0,076923077
passait	1	13	0,076923077
situations	1	13	0,076923077
rappelle	4	53	0,075471698
bizarre	2	27	0,074074074
évidemment	2	27	0,074074074
manger	2	27	0,074074074
problèmes	2	27	0,074074074
celui	1	14	0,071428571
demander	1	14	0,071428571
là-dessus	1	14	0,071428571
parfois	1	14	0,071428571
soeur	1	14	0,071428571
pareil	3	43	0,069767442
veut	3	43	0,069767442
trouvé	2	29	0,068965517
soit	6	88	0,068181818
bref	1	15	0,066666667
derrière	2	30	0,066666667
dessous	1	15	0,066666667
expérience	1	15	0,066666667
regarder	1	15	0,066666667
remarque	1	15	0,066666667
terrain	1	15	0,066666667
toute	5	75	0,066666667
voit	2	31	0,064516129
gens	9	140	0,064285714
avaient	6	94	0,063829787
moment	6	95	0,063157895
avait	38	603	0,063018242
moi	31	493	0,062880325
arrivait	1	16	0,0625
français	1	16	0,0625
marche	1	16	0,0625
point	1	16	0,0625

être	10	164	0,06097561
elle	43	726	0,05922865
appelle	1	17	0,058823529
Grenoble	2	34	0,058823529
long	1	17	0,058823529
milieu	1	17	0,058823529
prise	1	17	0,058823529
toujours	7	120	0,058333333
je	101	1745	0,057879656
faisait	6	104	0,057692308
mec	5	87	0,057471264
mieux	2	35	0,057142857
normal	2	35	0,057142857
problème	3	53	0,056603774
année	2	36	0,055555556
façon	4	72	0,055555556
vaut	1	18	0,055555556
chaque	2	37	0,054054054
avais	14	261	0,053639847
apparemment	1	19	0,052631579
donner	1	19	0,052631579
grande	1	19	0,052631579
personnes	1	19	0,052631579
place	1	19	0,052631579
sept	1	19	0,052631579
venue	1	19	0,052631579
que	59	1140	0,051754386
pour	31	607	0,05107084
vrai	8	157	0,050955414
monde	3	59	0,050847458
dit	21	414	0,050724638
comment	6	119	0,050420168
conneries	1	20	0,05
ville	1	20	0,05
été	7	141	0,04964539
la	70	1415	0,049469965
histoire	2	41	0,048780488
soir	2	41	0,048780488
carte	1	21	0,047619048
demande	1	21	0,047619048
sait	1	21	0,047619048
voulait	2	42	0,047619048
voilà	14	295	0,047457627
ma	6	127	0,047244094

fois	9	192	0,046875
me	22	471	0,04670913
truc	11	238	0,046218487
parce_que	21	457	0,04595186
contre	1	22	0,045454545
cuisine	1	22	0,045454545
instit	1	23	0,043478261
parle	1	23	0,043478261
question	1	23	0,043478261
sinon	2	46	0,043478261
travail	1	23	0,043478261
viens	1	23	0,043478261
l'	33	768	0,04296875
parce_qu'	5	118	0,042372881
il	72	1721	0,041836142
arriver	1	24	0,041666667
savoir	1	24	0,041666667
semaine	1	24	0,041666667
vous	10	242	0,041322314
arrive	3	74	0,040540541
tiens	1	25	0,04
leur	5	127	0,039370079
allez	1	26	0,038461538
pu	1	26	0,038461538
vient	1	26	0,038461538
y	39	1021	0,038197845
une	30	788	0,038071066
va	8	215	0,037209302
de	83	2236	0,037119857
exactement	1	27	0,037037037
grand-mère	1	27	0,037037037
sortir	1	27	0,037037037
thèse	1	27	0,037037037
sur	12	326	0,036809816
bien	11	302	0,036423841
air	1	28	0,035714286
elles	4	112	0,035714286
venir	1	28	0,035714286
qu'	23	659	0,034901366
tu	82	2351	0,034878775
allé	1	29	0,034482759
trucs	6	175	0,034285714
comme	16	478	0,033472803
clair	1	30	0,033333333

par_contre	1	30	0,033333333
puis	14	420	0,033333333
se	12	362	0,033149171
les	49	1481	0,033085753
d'	14	426	0,03286385
dire	7	213	0,03286385
mère	3	92	0,032608696
était	36	1109	0,032461677
conflits	2	62	0,032258065
disais	1	31	0,032258065
dur	1	31	0,032258065
parti	1	31	0,032258065
seul	1	31	0,032258065
pris	2	63	0,031746032
ceux	1	32	0,03125
maintenant	2	64	0,03125
vont	1	32	0,03125
ils	27	878	0,030751708
matin	1	33	0,03030303
partir	1	33	0,03030303
des	37	1223	0,030253475
le	56	1885	0,029708223
ce	14	474	0,029535865
avez	1	34	0,029411765
niveau	1	34	0,029411765
passer	1	34	0,029411765
là-bas	1	35	0,028571429
après	6	213	0,028169014
début	1	36	0,027777778
justement	1	36	0,027777778
pense	2	72	0,027777778
rentrer	1	36	0,027777778
trois	3	108	0,027777778
mois	2	73	0,02739726
allait	1	37	0,027027027
partie	1	37	0,027027027
prendre	1	37	0,027027027
quand_même	6	222	0,027027027
nous	7	261	0,026819923
ai	17	645	0,026356589
bout	1	38	0,026315789
fallait	1	38	0,026315789
fini	1	38	0,026315789
tête	1	38	0,026315789

peux	3	115	0,026086957
vas	2	77	0,025974026
faire	8	315	0,025396825
genre	2	79	0,025316456
même	5	199	0,025125628
étais	5	200	0,025
sans	1	40	0,025
compte	1	41	0,024390244
qui	30	1258	0,023847377
souvent	1	42	0,023809524
passé	1	43	0,023255814
qu-	1	43	0,023255814
effectivement	1	44	0,022727273
ta	1	44	0,022727273
dans	17	749	0,022696929
veux	4	178	0,02247191
parler	1	45	0,022222222
trop	3	136	0,022058824
a	39	1788	0,021812081
disait	1	46	0,02173913
sont	5	233	0,021459227
donc	16	746	0,021447721
m'	6	281	0,021352313
f-	1	47	0,021276596
plus	9	425	0,021176471
peut-être	2	95	0,021052632
surtout	1	48	0,020833333
faut	4	196	0,020408163
mettre	1	49	0,020408163
d-	2	100	0,02
v-	1	50	0,02
déjà	2	101	0,01980198
envie	1	51	0,019607843
ses	1	51	0,019607843
vu	2	104	0,019230769
on	26	1356	0,019174041
n'	2	106	0,018867925
s'	5	266	0,018796992
dis	3	160	0,01875
autre	2	108	0,018518519
choses	1	54	0,018518519
mes	1	54	0,018518519
temps	1	54	0,018518519
j'	20	1090	0,018348624

en_fait	7	382	0,018324607
autres	1	55	0,018181818
passe	1	57	0,01754386
un_peu	5	288	0,017361111
vraiment	3	174	0,017241379
chez	1	59	0,016949153
maison	1	60	0,016666667
mis	1	60	0,016666667
à	37	2237	0,016540009
où	5	303	0,01650165
sais	7	432	0,016203704
exemple	1	62	0,016129032
ont	4	255	0,015686275
a-	1	64	0,015625
mon	2	129	0,015503876
ne	1	65	0,015384615
sa	1	65	0,015384615
ça	32	2083	0,015362458
peu	1	66	0,015151515
étaient	2	134	0,014925373
beaucoup	1	69	0,014492754
es	3	207	0,014492754
ton	1	69	0,014492754
est	49	3392	0,014445755
là	10	698	0,014326648
mais	23	1617	0,014223871
en	15	1060	0,014150943
as	5	355	0,014084507
plein	1	71	0,014084507
en_plus	2	144	0,013888889
font	1	72	0,013888889
pas	30	2171	0,013818517
un	18	1319	0,013646702
entre	1	74	0,013513514
te	4	303	0,01320132
attends	1	78	0,012820513
heure	1	79	0,012658228
non	9	714	0,012605042
j-	1	80	0,0125
ben	5	402	0,012437811
si	6	489	0,012269939
c'	40	3296	0,012135922
heures	1	84	0,011904762
m-	1	87	0,011494253

quoi	14	1293	0,010827533
rien	2	187	0,010695187
tout	6	568	0,01056338
fais	1	98	0,010204082
eu	1	100	0,01
mal	1	102	0,009803922
bon	7	740	0,009459459
i-	1	106	0,009433962
toi	1	111	0,009009009
petit	1	120	0,008333333
du	3	389	0,007712082
t'	2	282	0,007092199
ou	3	428	0,007009346
fait	3	485	0,006185567
au	2	325	0,006153846
quand	2	360	0,005555556
aussi	1	195	0,005128205
deux	1	207	0,004830918
très	1	213	0,004694836
putain	1	227	0,004405286
enfin	3	728	0,004120879
et	8	2658	0,003009782
ouais	9	3191	0,002820432
oui	1	561	0,001782531
ah	1	875	0,001142857