



**HAL**  
open science

# Flexibility of speech comprehension processes: The use of acoustic-phonetic and semantic cues in French speech segmentation

Maria del Mar Cordero Rull

► **To cite this version:**

Maria del Mar Cordero Rull. Flexibility of speech comprehension processes: The use of acoustic-phonetic and semantic cues in French speech segmentation. Linguistics. Université Côte d'Azur, 2023. English. NNT: 2023COAZ2047 . tel-04541919

**HAL Id: tel-04541919**

**<https://theses.hal.science/tel-04541919>**

Submitted on 11 Apr 2024

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

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

# THÈSE DE DOCTORAT

## Flexibilité des processus de compréhension de la parole

L'utilisation d'indices acoustico-phonétiques et sémantiques dans la  
segmentation de la parole en français

**Maria del Mar CORDERO RULL**

Bases, Corpus, Langage (BCL)

**Présentée en vue de l'obtention  
du grade de docteur en Sciences du  
Langage  
d'Université Côte d'Azur**

**Dirigée par :** Fanny Meunier

**Soutenu le :** 12 décembre 2023

**Devant le jury, composé de :**

Clara Martin, Professor, BCBL (Basque  
Center on Cognition, Brain and Language),  
San Sebastián (Spain)

Olivier Crouzet, Assistant Professor,  
Université de Nantes (France)

Amanda Edmonds, Full professor,  
Université Côte d'Azur (France)

Frédéric Isel, Professor, Université Paris  
Nanterre – Paris Lumières (France)

# Flexibilité des processus de compréhension de la parole

L'utilisation d'indices acoustico-phonétiques et  
sémantiques dans la segmentation de la parole en français

## Flexibility of speech comprehension processes

The use of acoustic-phonetic and semantic cues in French  
speech segmentation

---

**Jury :**

Présidente du jury :

Amanda Edmonds, Full professor, Université Côte d'Azur (France)

Rapporteurs :

Clara Martin, Professor, BCBL (Basque Center on Cognition, Brain and  
Language), San Sebastián (Spain)

Frédéric Isel, Professor, Université Paris Nanterre – Paris Lumières (France)

Examineurs :

Amanda Edmonds, Full professor, Université Côte d'Azur (France)

Olivier Crouzet, Assistant Professor, Université de Nantes (France)

## En français

**Titre :** Flexibilité des processus de compréhension de la parole : L'utilisation d'indices acoustico-phonétiques et sémantiques dans la segmentation de la parole en français

---

### Résumé

La segmentation de la parole continue est un processus cognitif quotidien. Malgré l'absence de limites explicites entre mots, la compréhension de la parole semble aisée. Les paires de syntagmes déterminants (SD) phonémiquement identiques en élision, comme *l'affiche-la fiche* (prononcés /lafif/), peuvent poser un défi. Dans ces cas, la segmentation précise devient cruciale pour la compréhension. Des études antérieures montrent que les auditeurs français dépassent les niveaux de hasard en discriminant et identifiant ces SD. Les analyses acoustiques révèlent des différences en durée et fréquence fondamentale (F0). Bien que ces indices ne soient généralement pas utilisés en segmentation du français, ces résultats suggèrent leur exploitation.

La première partie de cette thèse comprend deux études explorant l'interaction des indices de bas (acoustiques) et haut niveau (sémantiques) dans le traitement de SD homophoniques (/lafif/ ou *l'amie-la mie* (/lami/)). Les mesures comportementales ne révèlent pas directement la segmentation de la parole, mais éclairent probablement les processus post-segmentation. Deux expériences EEG saisissent les corrélats neuronaux de l'exploitation des deux types d'indices lors de la perception de ces SD. Une tâche de jugement de phrases semi-passive (Expérience 1) et d'identification de mots (Expérience 2) testaient 51 francophones en manipulant l'attention vers le sens de la phrase (Exp. 1) ou la forme lexicale (Exp. 2). Les indices acoustico-phonétiques en contextes défavorables étaient attendus à générer des composantes ERP P3 et N400 par rapport aux contextes favorables, reflétant la détection acoustique inattendue et la difficulté de récupération lexicale. Des amplitudes P3 plus larges étaient observées en contextes favorables dans les deux expériences. L'effet inverse était observé significatif uniquement dans l'Expérience 2, avec des amplitudes N400 plus larges pour les contextes défavorables. Nos résultats suggèrent que les locuteurs français détectent les indices acoustico-phonétiques des SD homophoniques en élision. Cependant, l'accès lexical semble plus difficile lorsque l'attention est sur l'information lexicale plutôt que sémantique. L'information de haut niveau semble pertinente lors du traitement des

phrases, tandis que celle de bas niveau est suffisamment saillante pour être détectée et traitée.

La seconde partie examine comment deux populations gèrent les défis de segmentation sans contexte : adultes atteints de dyslexie développementale (DD) et apprenants français. L'Expérience 3 testait 49 adultes avec DD et 49 lecteurs typiques via une tâche de répétition de mots. Le rôle de deux paramètres F0 (moyenne et pente) était évalué, explorant l'impact des déficits phonologiques de la DD sur la segmentation. L'augmentation de la moyenne de F0 sur /a/ devait favoriser réponses aux items avec voyelle initiale pour les lecteurs typiques, mais prévoyait une sensibilité réduite chez les individus avec DD. Les résultats montraient plus de choix de voyelles initiales basée sur les propriétés de F0. L'absence d'interaction conditions×group suggère que les deux groupes adoptent des stratégies de segmentation similaires. Nos résultats soulignent le rôle crucial de la moyenne et la pente de F0 dans la segmentation du français.

Enfin, l'Expérience 4 testait 31 locuteurs français et 22 apprenants (11 anglophones, 11 hispanophones) via une tâche de discrimination AX. Leur sensibilité aux indices acoustico-phonétiques de ces SD était examinée. Les anglophones et hispanophones utilisent des indices différents de ceux des locuteurs français, en se basant sur la durée, la F0 et l'intensité, et manquent une forme élidée d'article défini. Les résultats révélaient que les apprenants surpassent les natifs dans la discrimination des SD. Les apprenants bénéficient de leurs stratégies de L1, suggérant que les indices dans l'élision guident les locuteurs natifs et non natifs vers la segmentation des mots.

---

**Mots clés :** segmentation de la parole, perception de la parole, indices acoustiques, français, EEG, dyslexie, L2

## English

**Title:** Flexibility of speech comprehension processes: The use of acoustic-phonetic and semantic cues in French speech segmentation

---

### Abstract

Continuous speech segmentation is a cognitive process that humans engage in daily. Despite the lack of explicit word boundaries, speech comprehension seems effortless. A challenge may arise when encountering pairs of phonemically identical determiner phrases (DPs) in elision situations, like *l'affiche* (“the poster”) – *la fiche* (“the file”), both produced as /lafif/. Accurate segmentation becomes crucial for comprehension in these cases. Previous research shows that French listeners surpass chance levels in discriminating and identifying such DPs. Acoustic analyses unveil differences in parameters like duration and fundamental frequency (F0). While these cues are not typically used in French speech segmentation, these findings suggest their potential exploitation for accurate segmentation.

The first part of this dissertation includes two experiments examining the interplay of acoustic and semantic cues in processing French homophonic DPs, like *l'affiche–la fiche* or *l'amie–la mie* (/lami/). We aimed to ascertain the extent to which these cues are used when acoustic and contextual information is available. Behavioural measures may not directly elucidate speech segmentation, but they likely provide insights into post-segmentation processes. To address cue modulation during the perception of homophonic DPs, we conducted two EEG experiments to capture neural correlates by exploiting low-level (acoustic) and high-level (semantic) cues. A semi-passive sentence judgement task (Experiment 1) and a word identification task (Experiment 2) were administered to 51 French-speaking participants, manipulating attention towards sentence meaning (Experiment 1) or lexical form (Experiment 2). Acoustic-phonetic cues in disfavoured contexts were expected to yield P3 and N400 ERP components compared to favouring ones, reflecting unexpected acoustic detection and lexical retrieval difficulty. Larger P3 amplitudes were found in favouring contexts for homophonic DPs in both experiments. The reverse effect was found significant only in Experiment 2, with larger N400 amplitudes for disfavoured contexts. Our findings suggest that French speakers detect acoustic-phonetic information in homophonic DPs in elision. However, lexical access seems more challenging when attention was on lexical than sentential information. High-

level information seems relevant during sentence processing, while low-level information is salient enough to be detected and processed.

The second part explores how two populations navigate segmentation challenges in the absence of context: adults with Developmental Dyslexia (DD) and French learners. Experiment 3 tested 49 DD adults and 49 typical readers through a word repetition task. We assessed the role of two F0 parameters (slope and mean), exploring the impact of phonological deficits in DD on speech segmentation. Increasing the mean F0 of /a/ was expected to enhance vowel-initial item segmentation for typical readers, but individuals with DD were expected to show reduced sensitivity. Results showed increased vowel choices based on F0 properties, regardless of group. No interaction between conditions and groups suggests that both groups display similar segmentation strategies. Our findings underscore the crucial role of both F0 mean and slope in accurate French segmentation.

Finally, Experiment 4 tested 31 native French speakers and 22 French learners (11 English, 11 Spanish speakers) through an AX discrimination task. We examined the sensitivity to acoustic-phonetic information of French homophonic DPs. English and Spanish speakers use different cues than French, relying on duration, F0, and intensity, and lack an elided form of the definite article. Results revealed that learners outperformed native listeners in discriminating homophonic DPs. Learners benefited from their L1 strategies in French segmentation, suggesting that cues embedded in elision guide both native and non-native listeners towards word segmentation.

Flexibility of speech comprehension processes: The use of acoustic-phonetic and semantic cues in French speech segmentation

---

**Keywords:** speech segmentation, speech perception, acoustic cues, French, EEG, dyslexia, L2





## Table of contents

Résumé .....	2
Abstract.....	4
Acknowledgements .....	15
Abbreviations .....	17
<b>CHAPTER 1. Speech segmentation and word recognition .....</b>	<b>20</b>
1.1. The challenge of speech segmentation .....	20
1.2. Understanding speech perception from models of word recognition.....	23
1.2.1. Cohort model. ....	23
1.2.2. TRACE. ....	25
1.2.3. Shortlist.....	26
1.2.4. NAM.....	27
1.3. The Use of Multiple Cues in Speech Segmentation .....	28
1.4. The role of acoustic cues during segmentation .....	30
1.5. Fundamental frequency as a cue to continuous speech segmentation.....	33
1.6. Summary and conclusion .....	34
<b>CHAPTER 2. Speech segmentation in French .....</b>	<b>36</b>
2.1. Cues to speech segmentation in French.....	36
2.2. Segmenting ambiguity in French: The case of resyllabification .....	38
2.3. Cues to elision in French .....	41
2.4. Summary and conclusion .....	44
<b>CHAPTER 3. Speech segmentation in adults with developmental dyslexia .....</b>	<b>46</b>
3.1. The challenge of segmentation .....	46
3.2. Phonological awareness.....	50
3.3. Deficits in rise time and F0 perception in dyslexic adults.....	52
3.4. Auditory processing theories of developmental dyslexia.....	54
3.4.1. Anchoring and categorical déficits. ....	54
3.4.2. Phonology-based theories.....	55
3.4.3. Temporal sampling theory and AST. ....	56
3.5. Summary and conclusion .....	58
<b>CHAPTER 4: Segmenting Non-Native Speech.....</b>	<b>60</b>
4.1. The challenge of segmenting non-native speech.....	60
4.2. Word recognition in an L2.....	63
4.3. Psycholinguistic models of L2 speech perception.....	64
4.3.1. PAM. ....	64
4.3.2. SLM.....	65
4.3.3. DMAP.....	66
4.3.4. The Cue-weighting Theory.....	67
4.4. The use of acoustic cues during L2 processing for English and Spanish speakers .	69
4.4.1. The exploitation of acoustic cues for L1 English. ....	69
4.4.2. The exploitation of acoustic cues for L1 Spanish. ....	71
4.5. Processing French resyllabification as an L2 .....	72
4.6. Summary and conclusion .....	74
<b>CHAPTER 5. Acoustic cues of the material in this dissertation.....</b>	<b>77</b>
5.1. DPs in Experiments 1 and 2 .....	78

5.1.1. Materials.....	78
5.1.2. Measurements and analyses.....	79
5.2. DPs in Experiments 3 and 4.....	82
5.2.1. Materials.....	82
5.2.2. Measurements and analyses.....	82
5.3. Analyses of the F0 trajectory of DPs in Experiment 3.....	85
5.3.1. Materials.....	85
5.3.2. Measurements and analyses.....	86
5.4. Summary and conclusion.....	90

**CHAPTER 6. The interplay of acoustic and semantic cues during the processing of homophonic DPs in French..... 91**

6.1. Introduction.....	91
6.2. Goals and hypotheses.....	94
6.3. Experiment 1: Electrophysiological evidence to online segmentation during auditory perception with a judgement task (focus on semantics).....	97
6.3.1. Method.....	97
6.3.1.1. Participants.....	97
6.3.1.2. Materials.....	98
6.3.1.3. Procedure.....	100
6.3.2. EEG recording and preprocessing.....	102
6.3.3. EEG data analysis.....	104
6.3.4. ERP results.....	105
6.3.5. Discussion.....	107
6.4. Experiment 2: Behavioural and electrophysiological evidence to online segmentation during auditory perception with an identification task (focus on word form).....	108
6.4.1. Method.....	108
6.4.1.1. Participants.....	108
6.4.1.2. Materials.....	108
6.4.1.3. Procedure.....	109
6.4.2. EEG recording and preprocessing.....	110
6.4.3. Data analysis.....	110
6.4.3.1. Behavioural measures.....	110
6.4.3.2. EEG data.....	112
6.4.4. Results.....	112
6.4.4.1. Behavioural measures.....	112
6.4.4.2. ERP results.....	116
6.4.5. Discussion.....	118
6.5. Tasks comparison.....	119
6.5.1. Analysis.....	120
6.5.2. Results.....	120
6.5.3. Discussion.....	121
6.6. General discussion and conclusion.....	122

**CHAPTER 7. The role of F0 trajectories during the segmentation of French in skilled readers vs. adults with dyslexia..... 127**

7.1. Introduction.....	127
7.2. Hypotheses.....	130
7.3. Method.....	131

7.3.1. Participants.....	131
7.3.2. Cognitive and phonological assessment tests.....	132
7.3.3. Stimuli.....	136
7.3.4. Procedure.....	137
7.4. Data analysis.....	138
7.5. Results.....	139
7.5.1. Cognitive and phonological tests.....	139
7.5.2. Response to vowel targets.....	139
7.5.3. Correlations with phonological UP.....	141
7.6. Discussion.....	142

**CHAPTER 8. The interplay of L1 and L2 segmentation strategies in French speech segmentation ..... 149**

8.1. Introduction.....	149
8.2. The present study.....	153
8.3. Aims and hypotheses.....	154
8.4. Method.....	155
8.4.1. Participants.....	155
8.4.2. Language Measures.....	157
8.4.3. Materials.....	158
8.4.4. Acoustic analyses.....	158
8.4.5. Procedure.....	159
8.5. Data analysis.....	161
8.6. Results.....	162
8.6.1. Language measures.....	162
8.6.2. Sensitivity measures to discrimination.....	163
8.6.3. Correlation analyses: battery of language tasks and language proficiency.....	165
8.7. Discussion and conclusion.....	165

**CHAPTER 9. General discussion and conclusion ..... 171**

9.1. Recapitulation of main findings.....	171
9.2. Neural correlates of acoustic-phonetic detection.....	172
9.3. Integrating low- and high-level cues during elision.....	174
9.4. The role of F0 as a cue to speech segmentation in French.....	175
9.5. The exploitation of acoustic-phonetic cues to elision within spoken word recognition models.....	177
9.6. The implications of the phonological deficits in developmental dyslexia to the detection of acoustic-phonetic information.....	178
9.7. The role of L1 strategies during the segmentation of French speech as an L2.....	179
9.8. Main differences between liaison and elision: The differing roles of consonants and vowels.....	182
9.9. Conclusion.....	183
9.10. Limitations.....	186
9.11. Future research perspectives.....	187

References.....	189
-----------------	-----

Annexes.....	224
--------------	-----

## Table of figures

- Figure 1.** Signal waveform and spectrogram of the sentence “Le théâtre national a choisi l’affiche du spectacle”, one of the sentences used in Experiments 1 and 2 (Chapter 6) in this dissertation. The green line illustrates the F0 modulations of the sentence. The blue line shows the pitch values throughout the sentence. .... 34
- Figure 2.** Fundamental frequency rise in the phonemically identical phrases C’est la mie (left panel) and C’est la mie (right panel), both /selami/. The spectrogram shows the F0 modulations along the phrases. F0 rise is marked by a white arrow, marking the beginning of the content words (amie and mie, respectively). Y-axis shows F0 Hz, x-axis shows the duration of the segment in seconds. .... 41
- Figure 3.** Illustration of the acoustic manipulations performed to generate our stimuli: A) cross-splicing, B) identity-splicing. Green arrows show the generation of the congruent condition, red arrows show the generation of the incongruent condition. Panel (1) illustrates the two original productions of la mie, panel (2) illustrates those of l’amie. 78
- Figure 4.** Illustration of the F0 manipulations of the first two phonemes (/la/) for the pair “la mie/l’amie”. Conditions are represented in separate lines: Vow-nat (*continuous red*), Co-nat (*discontinuous red*), Co-Slope (*yellow*), Co-Shift (*blue*), and Co-Slope+Shift (*green*). .... 86
- Figure 5.** Illustration of the F0 mean trajectory of the five conditions in Experiment 3, including four measures: onset, midpoint, maximal peak, and offset. Panel A shows the mean F0 measures of the natural vowel-initial items (Vow-nat condition); panel B shows the mean F0 measures of the natural consonant-initial items (Co-nat condition); panel C shows the mean F0 measures of the slope-resynthesised items (Co-slope condition); panel D shows the mean F0 measures of the mean-resynthesised items (Co-Shift condition); panel E shows the mean F0 measures of the mean-and-slope-resynthesised items (Co-Slope+Shift condition). Error bars show SDs. .... 89
- Figure 6.** Illustration of H1 (blue line) and H2 (orange line) hypotheses for the P3 amplitudes across conditions (control, incongruent and congruent). .... 95
- Figure 7.** Illustration of H1 (blue line) and H2 (orange line) hypotheses for the N400 amplitudes across conditions (control, incongruent and congruent). .... 96
- Figure 8.** Illustration of the acoustic manipulations performed to generate our stimuli: A) cross-splicing, B) identity-splicing. Green arrows show the generation of the congruent condition, and red arrows show the generation of the incongruent condition. Panel (1) illustrates the two original productions of la mie, panel (2) illustrates those of l’amie. .... 100
- Figure 9.** Example of the sequence of the presentation of stimuli for Experiment 1. . 102
- Figure 10.** Scalp-maps of the ERP waveforms in Experiment 1 covering a time window from 0 to 1100 ms. Panel (A) shows the signal amplitude (in  $\mu\text{V}$ ) for control, incongruent, and congruent conditions. Panel (B) shows the amplitude differences resulting from

permutation tests: control-incongruent, control-congruent, and incongruent-congruent. The blue dots indicate the electrodes showing a statistically significant effect, corresponding to channels in which p-values reached significance at  $\alpha \leq .05$ . ..... 105

**Figure 11.** Mean ERP amplitudes (and SDs) of frontal midline electrode Fz and central midline electrode Cz in Experiment 1. Black lines show ERP amplitudes for the control condition, red lines for incongruent, and green lines for congruent. Scalp maps displaying the effects of interest are shown on the top (P3a on the left, N400 on the right). ..... 106

**Figure 12.** Percentage of errors (and SDs) made during the identification task. Left bars show errors when identifying non-homophonic words (Other word), right bars show errors to homophone targets (Target word). Green bars show results in the congruent condition, red bars show results in the incongruent condition. .... 112

**Figure 13.** Mean reaction times in milliseconds (and SDs) to correct responses are shown to non-homophonic words (Other word, left bars) and to homophone targets (Target word, right bars). Green bars show results in the congruent condition, red bars show results in the incongruent condition. .... 114

**Figure 14.** Percentage of errors (and SDs) made during the identification task. The left panel shows errors in identifying non-homophonic words (Other word), right panel shows errors in homophone targets (Target word). Red bars show results for elided DPs (apostrophe), blue bars show results for non-elided DPs (article). This is shown for both congruent and incongruent conditions. .... 115

**Figure 15.** RT means in ms (and SDs) to correct responses in the identification task. The left panel shows RTs when identifying non-homophonic words (Other word), right panel shows RTs to homophone targets (Target word). Red bars show results for elided DPs (apostrophe), blue bars show results for non-elided DPs (article). This is shown for both congruent and incongruent conditions. .... 116

**Figure 16.** Scalp maps of the ERP waveforms covering a time window from 0 to 1100 ms. Panel A) shows the signal amplitude (in  $\mu\text{V}$ ) for control, incongruent, and congruent conditions. Panel B) shows the amplitude differences resulting from permutation tests: control-incongruent, control-congruent, and incongruent-congruent. The blue dots indicate the electrodes showing a statistically significant effect, corresponding to channels in which p-values reached significance at  $\alpha \leq .05$ . .... 117

**Figure 17.** Mean ERP amplitudes (and SDs) of frontal midline electrode Fz, central midline electrode Cz, and parietal midline electrode Pz. Black lines show amplitudes for the control condition, red lines for incongruent, and green lines for congruent. Scalp topography maps displaying the effects of interest are shown on the top (P3b on the left, N400 on the right). .... 118

**Figure 18.** Mean ERP amplitudes of both components of interest (P3, N400) for each task (1, 2). Black boxplots show effects for the control condition, red boxplots for the incongruent condition, and green boxplots for the congruent condition. Error bars show SDs. .... 120

<b>Figure 19.</b> Illustration of the task design. ....	137
<b>Figure 20.</b> Results from DD adults (red bars) and SR groups (blue bars) are shown for percentages of responses to vowel items for all 5 conditions by adults with dyslexia (red bars) and by skilled readers (blue bars). Stars show the significance level (p values: * p <.05; ** p <.01; *** p <.001) of contrasts revealed by a GLMM.....	139
<b>Figure 21.</b> Illustration of the experimental design.....	160
<b>Figure 22.</b> Mean d' scores for consonant–vowel and vowel–consonant pairs in the discrimination task for each group: French (red), English (green), and Spanish (blue). Left boxplots show scores to consonant–vowel pairs, right boxplots to vowel–consonant pairs. Error bars indicate SDs.....	164

### Table of tables

<b>Table 1.</b> Illustration of the levels of phonology included in the auditory cortical oscillatory hierarchy (extracted from Goswami, 2015).....	57
<b>Table 2.</b> Mean duration and F0 values (and SDs) for [l] and [a] phonemes in vowel-initial and consonant-initial items. ....	79
<b>Table 3.</b> Mean duration and F0 values (and SDs) for [l] and [a] phonemes in vowel-initial and consonant-initial items. ....	84
<b>Table 4.</b> Mean time points (in s) and frequency (in Hz) values (and SDs) for the four acoustic measures taken on the F0 (onset, midpoint, maximum peak, and offset) across conditions (Vow-nat, Co-nat, Co-Slope, Co-Shift, Co-Slope+Shift).....	87
<b>Table 5.</b> Mean (M), Standard deviation (SD), and range values across conditions (congruent, incongruent, control) for sentence total length, duration of target DPs, and target DPs' relative position within sentences.....	99
<b>Table 6.</b> Illustration of the distribution of sentences along the 4 experimental blocks (example pair <i>l'amie-la mie</i> ).....	101
<b>Table 7.</b> Design illustration of auditory primes and written word associations (example pair <i>l'amie-la mie</i> ).....	108
<b>Table 8.</b> Error contrasts performed in post hoc tests for Context and Written word...	113
<b>Table 9.</b> RT contrasts performed in post hoc tests for Context and Written word. ....	114
<b>Table 10.</b> Means (Standard Deviations) and t statistic of DD and SR groups for Chronological age, Reading level, and raw scores on Raven's Matrices. ....	135

<b>Table 11.</b> Means (Standard Deviations), <i>t</i> statistic, and effect size (Cohen's <i>d</i> ) of DD and SR groups for performance on Word reading, Phonological skills, and Vocabulary breadth. ....	135
<b>Table 12.</b> Estimated coefficients from GLMM in function of Group and Condition for vowel responses. ....	140
<b>Table 13.</b> Prosodic strategies used in the L1 groups chosen (French, English, and Spanish). ....	153
<b>Table 14.</b> Biographic and language background information of the two groups of learners of French (Spanish, English, and the mean of both). ....	156
<b>Table 15.</b> Estimated coefficients from LMM for mean <i>d'</i> scores across groups. ....	162
<b>Table 16.</b> LLAMA tests mean percent scores, vocabulary skills, and self-assessment of French proficiency for the two groups of French learners (Spanish, English, and the mean of both). ....	162





## Acknowledgements

I would first like to express my heartfelt appreciation to my dissertation director Fanny Meunier for her unwavering support and motivation during the three and a half years of my doctoral journey. Her encouragement and provision of resources have played a pivotal role in successfully reaching the finish line.

I would like to express my gratitude to Mireille Besson for the invaluable suggestions at various stages of this dissertation and for dedicating the time to help with the project. The ongoing support of Mireille and genuine concern for my well-being has been instrumental in this endeavour.

I would like to give special thanks to the members of the jury Amanda Edmonds, Clara Martin, Frédéric Isel, and Olivier Crouzet for honouring me by judging my work.

I would like to extend my gratitude to all the participants who took part in my studies. Their contributions have made this dissertation and its research findings possible. I am also thankful for the assistance provided by my research assistants at different stages of the thesis. Special thanks to Ahmed Ladhari for his dedication and support during the final recruitment process, as well as his unwavering moral support.

I would like to acknowledge and appreciate the valuable input and discussions from Karolina Broś, Irene de la Cruz Pavia, and Minerva Rojas throughout this process. Their patience, insightful conversations, and support have been greatly valued and cherished.

This research would also not have been possible without the support of the Université Côte d'Azur and the CNRS. I would particularly like to thank the administration members of the BCL (Department of Bases, Corpus, and Language) for their assistance in facilitating various bureaucratic and organizational aspects throughout the dissertation process. I am also grateful to the teams Langage et Cognition (L&C) and Acquisition, Apprentissage, Adaptation (AAA) within BCL for their seminars, meetings, and generous support, including funding for conference attendance.

I would also like to thank the CoCoLab platform at the Université Côte d'Azur for providing the necessary equipment and service during my research projects. I would like to extend a special thank you to Ambre Denis-Noël for her exceptional moral support and invaluable assistance at different stages of this project.

Last but not least, I would like to express my deepest appreciation to my dear friends and colleagues, Anaïs, Maëliiss, Morgane, Clémentine, Chang, Nicolas, and Cristobal for their immense support, and continuous encouragement throughout this process. I am immensely thankful for the support of my parents and my friends Bea, Sara, Emily, Cinta, Júlia, Cristina, Irene, Cristina, Oscar, Laia, Olga, and David, for their inspiring conversations, companionship, and understanding, which made this journey more meaningful. Lastly, I want to acknowledge my partner, Marc, for being a constant presence and source of support on this path.

## **Abbreviations**

- AM:** amplitude modulations (of the speech envelope).
- AP:** accentual phrase.
- DD:** developmental dyslexia.
- DP:** determiner phrase (definite article + noun).
- EEG:** electroencephalography.
- ERP:** event related potential.
- F0:** fundamental frequency.
- F1:** first formant.
- F2:** second formant.
- L:** low.
- L1:** first language or native language. Both terms are consistently employed in this dissertation to denote the language that an individual acquires from birth and is initially exposed to in their familiar context.
- L2:** second language. In this dissertation, the L2 term is used in a broad sense, including any languages other than the L1.
- H:** high.
- MMN:** mismatch negativity.
- PA:** phonological awareness.
- WM:** working memory.



# THEORETICAL PART

## **CHAPTER 1. Speech segmentation and word recognition**

### **1.1. The challenge of speech segmentation**

Spoken language serves as a medium for communication and embodies two fundamental attributes: continuity and variability. Unlike written language, no silences or gaps naturally demarcate word boundaries within stretches of speech. Instead, the continuous stream suffers from abrupt pronunciation changes across words that emerge from the talker variations, such as coarticulation or stress patterns (Johnsrude and Buchsbaum, 2017). Additionally, the production of speech sounds involves a degree of overlap between the articulations of consonants and vowels. Consequently, the task at hand involves decoding the variable speech signal into a discrete form that facilitates the recognition of individual words.

Despite the apparent complexity of spoken speech, listeners are able to achieve comprehension. Listeners do not perceive the signal stream in a continuous way; instead, they are able to form distinct categories for speech sounds (Pisoni, 1979). These categories span various linguistic dimensions, including features, phonemes, morphemes, syllables, and words. However, the relevance of speech categories differs across languages. For example, while the syllable seems to play an important role in languages like French (Dumay et al., 1998; Mehler et al., 1981), Spanish (Peperkamp et al., 2010; Pons & Bosch, 2010), and Dutch (Cutler et al., 2007), this may not be the case for English (Cutler et al., 2007). Thus, the perception of complex sounds (which builds into syllables, phrases and, ultimately, utterances) cannot be reduced to the way a simple sound is discriminated, as discussed by Lotto & Holt (2016: p. 186).

Speech sounds possess acoustic properties that can be altered by adjacent sounds. In other words, the variability of sounds can be observed as a result of the interplay or coarticulation between a given sound and its surrounding (i.e. preceding and subsequent) sounds (e.g., Ainsworth, 1975; Joos, 1948; Lindblom, 1963; Nearey, 1989). For instance, Lindblom (1963) evidenced this by comparing the frequency of the second formant (F2) between vowels and consonants. The F2 exhibited higher values in the articulation of the vowels [i] and [u] in monosyllables like [did] and [dud], as compared to their counterparts

in [bib] and [bub]. This research shows that the auditory system of listeners is capable of detecting changes in speech, which are contingent upon neighbouring sounds. This aligns with the general notion that our sensory systems are responsive to changes that deviate from what is predictable or constant (e.g., Garner & Hake, 1951; Miller, 1956).

Given the intricate nature of speech variability, conventional acoustic measures such as frequency or intensity may not suffice to fully account for the variability within sounds, as discussed by Lotto and Holt (2016). For instance, parameters like airflow and nasal flow are shown to correlate with the production of plosive and nasal sounds, respectively. These attributes are hypothesised to regulate the prosodic strength of phonemes, thereby influencing and modulating language-specific segmentation processes (e.g., Cho et al., 2002; Fougeron, 2001; see Cho & Mücke, 2021 for a comprehensive review). Furthermore, other factors beyond acoustic properties come into play, such as context, attention, or memory, which significantly contribute to the variability of speech sounds (Kidd et al., 2011; Näätänen et al., 1978; Snyder & Weintraub, 2013; Stilp, 2020; Sussman, 2017). There is, thus, growing evidence that the auditory system of listeners is attuned to the multidimensional variability of speech, reflecting its inherent flexibility, which ultimately contributes to speech comprehension.

The fact that native speakers of a language possess the ability to perceive speech as a series of discrete and meaningful units, raises questions about the nature of linguistic representations in the brain and the mechanisms involved in mapping the continuous auditory signal to such representations. For example, both the written form “cat” and the spoken word [kæt] are linked to the same linguistic representation in the brain, associated with the representation of a specific animal, a cat. Importantly, whether the word [kæt] is pronounced by Katy or Sam, the listener’s auditory system processes the input similarly, independently of their variable pronunciations. Experience and familiarity with speech shape perception, enabling the establishment of correspondences between the signal and linguistic representations. This occurs by highlighting and enhancing significant patterns of variation while minimising the salience of less meaningful variability (see Davis & Johnsruide, 2007, and Diehl et al., 2004 for reviews).

Various linguistic theories propose that languages have a tendency to maximise the distinctiveness of sounds in order to minimise the articulatory effort required (Quantal Theory –Diehl et al., 2003; Stevens, 1989–, Hyper- and Hypo-articulation (H&H) Theory –Lindblom, 1990–, and the Auditory Enhancement Theory –Diehl et al., 1991; Diehl &

Kluender, 1989–). The Theory of Auditory Enhancement, in particular, highlights the use of combined articulations to enhance the distinctiveness of sounds. This phenomenon was observed in behavioural studies revealing that acoustic properties such as frequency influenced the perception of speech stretches. For instance, empirical evidence has shown that low frequent sounds bias the perception of subsequent sounds as high frequent (and vice versa) (e.g., Mann & Repp, 1981; Stilp, 2018), and sounds are perceived as shorter when preceded by long sounds (e.g., Denes, 1955; Kluender et al., 1988; Wade & Holt, 2005). In CV syllables like /da/ or /ga/, a higher frequency of F2 transitions in /s/ led to enhanced identifications of /k/, having lower-frequency formant transition onsets (Mann & Repp, 1981). Conversely, lower-frequency /j/ (such as the consonant in “she”) promoted more /t/ identifications, which has higher-frequency formant transition onsets. This effect has been accounted for as the “exaggeration sampling method” of contrasts of the auditory system (Lotto & Holt, 2016, p.188).

The enhanced distinctiveness of sounds is also conveyed in studies examining infants’ acquisition of the native language (L1). Young infants are sensitive to acoustic variations occurring at the phonetic boundaries between categories (Eimas, 1975; Eimas et al., 1971). The acoustic differences on which phonetic units depend are very small. As an illustration, one of the differences between /p/ and /b/ is timing, namely, only 10 ms marks the difference between one phoneme and the other (example extracted from Kuhl, 2007). These findings lead to the idea that, from birth, the listener’s auditory system is able to detect and discriminate subtle differences, an ability that allows infants to acquire language.

Experience greatly affects both the mental representation of sounds and their perception in adults, as demonstrated by studies testing speech perception and categorisation (e.g., Dupoux et al., 1997; Flege et al., 1996; Fodor & Bever, 1965; Goto, 1971; Lukyanchenko et al., 2011; Peperkamp et al., 2010). An illustrative case involves clicks, which are plosive consonant sounds that are found in languages in Southern and East Africa. Fodor and Bever (1965) asked English speakers to locate artificial click sounds embedded in English utterances. Speakers exhibited poor performance in identifying these clicks. The authors argued that clicks were perceived as non-linguistic sounds by English speakers, evidencing a bias of the speakers’ acoustic system towards their L1 knowledge. Hence, even though speech variability is present across languages, not all sound features are perceived equally across speakers.



The way listeners attune to variabilities in the signal to achieve word recognition and understand the transmitted information has been the focus of several psycholinguistic models. Psycholinguistic models of word recognition have diverse approaches to predicting the processes from speech perception to final word selection. Central to this is speech segmentation, where listeners associate specific cues in the speech signal to either the beginning or end of a word (Cutler, 1996). This is an important process as this information is then used to accurately identify the intended word (e.g., *ham*) while distinguishing it from competing words that share similar segments but have different initial or final boundaries (e.g., *hamster*, *hamstring*, *Hamlet*) (example extracted from Salverda et al., 2003). However, psycholinguistic models do not explicitly address how listeners segment the continuous signal in order to achieve recognition of individual words. The following section addresses how listeners overcome the process of word recognition and a proposition for the process of segmentation.

## **1.2. Understanding speech perception from models of word recognition**

Subtle acoustic differences present in the speech signal exert an influence on speech perception, thereby affecting the process of recognising words. The accurate selection of appropriate lexical items depends on how these differences are processed and integrated by the auditory system. Psycholinguistic models of word recognition have varied approaches to predicting activation-inhibition processes of target candidates and the temporal trajectory of word recognition. In the following, we provide an overview of the main psycholinguistic models that consider low-level (acoustic) information during the processes of speech segmentation and word recognition.

### **1.2.1. Cohort model.**

The Cohort model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980) is one of the pioneering and highly influential models of spoken word recognition. It encompasses three different stages: access, selection, and integration. During access, acoustic-phonetic components present in the speech signal are mapped onto corresponding words stored in the lexicon. The matching words are simultaneously activated and form a “cohort”, comprising a group of words sharing similar initial phonetic characteristics. This model considers that after the initial 150–200 ms (roughly corresponding with the first two phonemes), words starting with those same phonemes will be activated. Since the model considers initial similarities, the recognition of words does not necessarily require the processing of all phonemes involved. In the final stage,

for a word to be integrated, a process of semantic and syntactic retrieval is performed, allowing compatibility checks with higher-level constraints of activated words. Words are excluded from the cohort if a mismatch exists with the contextual constraints at this level.

The activation of candidates in the Cohort model is, thus, affected by increasing constraints as a word unfolds. Initially, candidate words receive bottom-up activation based on acoustic cues, but this activation subsequently transitions to a top-down processing approach influenced by contextual and semantic information, which filters and refines the set of candidate words. Therefore, the model considers that words in context can be recognised sooner than in isolation.

Several studies have challenged this model providing support for a continuous and dynamic process of spoken word recognition. For instance, in an eye-tracking study by Allopenna et al. (1998), participants engaged in an auditory word recognition task with a visual-word paradigm. Overall, listeners fixated on pictures depicting those words that were being heard. However, the authors observed that a proportion of fixations were also directed towards phonological competitors, such as *candy* (cohort competitor) and *sandal* (rhyme competitor) when *candle* was displayed, as opposed to the unrelated *beaker*. Particularly noteworthy is the observation that fixations on rhyme competitors tended to occur later than fixations on cohort competitors. As argued by the authors, this temporal pattern suggests the existence of a continuum in the processing of acoustic cues within a word. These findings provide direct evidence supporting the existence of phonological competition during the course of spoken word recognition.

To tackle these and other limitations, a subsequent version of the model, Cohort II (Marslen-Wilson, 1987, 1990), was developed. The Cohort II model adopts a connectionist approach. The connectionist approach is grounded on models of neuronal organization in the brain and proposes that rules and behaviour emerge through the repeated activation and subsequent reinforcement of neural connections. This approach suggests that learning is facilitated by the gradual accumulation and strengthening of input-output connections within these neural connections. Cohort II introduced a bottom-up approach that extended the scope of lexical activation to encompass additional phonetic and phonological cues beyond initial phonemes. This change of approach resulted in a more refined cohort of potential word candidates. Yet, a notable challenge of the Cohort model (original and Cohort II) is the lack of a mechanism to identify word

boundaries. The model assumes that segmenting the continuous speech stream lies in the recognition of individual words, with the offset of a word marking the beginning of a new word. However, the model falls short in explaining how word offsets are specifically identified in contrast to word onsets.

### 1.2.2. TRACE.

The TRACE model (McClelland & Elman, 1986), which, like Cohort II, adopts a connectionist approach. TRACE also holds significant influence and comprises three distinct layers that represent various linguistic units: feature, phoneme, and word. It takes multidimensional features as input, which represent words as phonemic strings. This design enables the model to activate multiple-word candidates that match different segments of the speech input, which proves advantageous in scenarios where speech is ambiguous or distorted.

Unlike the Cohort model, the competition among word candidates in TRACE does not adhere to an incremental hierarchy as the word unfolds. Instead, it takes place interactively among nodes<sup>1</sup>. For instance, the word “gift” activates the phoneme node for /g/, which is mutually excitatory for the feature nodes that match the acoustic features of /g/. Moreover, TRACE does not involve inhibition between layers, and the activation of words does not decrease in the presence of input mismatches. As a result, word candidates that share onset with the speech input are likely to experience earlier activation compared to word candidates sharing rhyme overlap. Moreover, a subsequent version of the model (Dahan et al., 2001) included lexical frequency effects during the activation of lexical candidates. This was done to account for the empirical evidence demonstrating the influence of lexical frequency on word recognition, with research showing enhanced activation for high-frequency words compared to low-frequency ones (e.g., Münte et al., 2001; Taft & Hambly, 1986; Van Petten & Kutas, 1990).

TRACE’s interactive activation framework allows for both bottom-up and top-down effects throughout the process of word recognition. Hence, the interaction between layers enables the lexical knowledge to exert an impact on perception. Words with the highest activation inhibit candidate words with lower activation, leading to the recognition of the candidate word that best matches the input. Consequently, akin to the Cohort model, the

---

<sup>1</sup> One node represents one representational unit. For example, the phoneme /k/ has a node in the phoneme layer that represents it.

location of word boundaries in TRACE follows a similar pattern, becoming more distinct as words are activated and recognised. In other words, stronger word activation leads to greater confidence regarding word boundaries, thereby informing the listener about the onset of the subsequent word.

### **1.2.3. Shortlist.**

The Shortlist model, also based on a connectionist approach, was developed to address the limitations of TRACE. Shortlist separates the activation of lexical candidates and the competition between them in two separate stages. It considers that the listeners' auditory system processes strings of phonemes as input to word recognition. The first stage comprises the activation of lexical candidates, which occurs serially, i.e. candidate words that are consistent with the input are activated at any moment in time as speech unfolds. A shortlist of candidate words is selected in this stage and is constantly updated as more phonemic information comes in. In the second stage, word candidates are integrated into an interactive activation network. This means that all activated words compete for recognition. Within this network, words sharing characteristics within the same input fragment are connected through inhibitory links, leading to competition between words. The degree of activation of lexical candidates is contingent upon their compatibility with the input, with decreasing activation when there is a mismatch between input and candidate. The candidate word displaying the highest activation will then exert inhibitory influence on words with lower activation, ultimately resulting in the recognition of the candidate word that most closely matches the input.

In Shortlist, longer words receive higher activation levels compared to shorter words. Simulations of Shortlist using continuous, degraded speech have shown that longer words are less affected by acoustic degradation than shorter ones. Within this model, similar to TRACE, segmentation arises as a byproduct of the lexical competition among candidates. The identification of word boundaries depends on recognising the offset of the preceding word.

Shortlist, like Cohort, implements a feed-forward flow of information. In contrast, Shortlist incorporates support for stress information by strengthening the "strong onset" of syllables, thereby amplifying the activation of intended lexical candidates. Nevertheless, the serial activation and competition processes in this model imply that the subphonemic variation in the signal is lost in early encoding stages.

This shortcoming is not included in a subsequent version of the model, Shortlist B, which is based on Bayesian principles. In Shortlist B, word candidates no longer possess word activations but are associated with word probabilities. The input in the new model are sequences of phoneme probabilities distributed across “three time slices per segment”, which has been considered problematic (Norris et al., 2018; Smits et al., 2003). Furthermore, the model incorporates word frequencies as prior probabilities and can accommodate mismatches in the input.

#### **1.2.4. NAM.**

The Neighborhood Activation Model (NAM) exhibits notable similarities to the previously mentioned models of spoken word recognition and integrates numerous concepts that have been established by previous models. NAM assumes that the input elicits the activation of a collection of words, which are stored as acoustic-phonetic patterns (Luce & Pisoni, 1998). These words differ from the input by a maximum of one phoneme, which takes into account the potential similarity with neighbours. For example, the word [kæt] (“cat”) would activate [pæt] (“pat”), [kIt] (“kit”), and [kæn] (“can”). This model thus provides an account for the influence of the number of similar words and their lexical frequency on spoken-word recognition (e.g., Goldinger, 1998; Luce, 1986).

The level of activation for each word is determined by its degree of match with the input. NAM computes a frequency-weighted neighbourhood probability for each word. Subsequently, acoustic-phonetic patterns activate word decision units. The activation of word decision units is determined by the activation of the corresponding acoustic-phonetic patterns and higher-level lexical information, such as word frequency. This process enables to emphasise bottom-up information, akin to the principles of the Cohort model. NAM generates several predictions regarding the effects of the number of neighbours and their lexical frequency on word recognition. A word is then recognised if its decision value surpasses a specific threshold. Like the previously mentioned models, word recognition occurs within the context of existing words stored in memory. The NAM model predicts that the process of word recognition is affected by the acoustic-phonetic similarity among the activated candidates and their lexical frequency. Thus, compared to low-frequency words, high-frequency words tend to have higher activation levels and may be recognised faster. However, high-frequency neighbours will slow down the competition among candidates compared to low-frequency neighbours.

PARSYN (Luce et al., 2000) was modelled to address several shortcomings of the NAM model and was built under a connectionist framework. It comprises three hierarchical levels: an input level consisting of position-specific allophones, an allophone level, and a word level. Word activation within PARSYN propagates in a bottom-up manner across these levels. Competition among words is achieved through inhibitory connections that are established between words at the word level.

NAM or PARSYN models do not address directly the process of speech segmentation. In both, syllable and word boundaries are explicitly indicated in the input of the model. Therefore, unlike other models such as Cohort, TRACE, and Shortlist, NAM and PARSYN are limited to recognising words in isolation and are unable to handle continuous speech.

The psycholinguistic models presented in this section show different approaches to word recognition, however, they do not explicitly consider the segmentation of speech. For these models, segmentation is a byproduct of the preceding word, i.e. it is not a process. In the following section, we present a psycholinguistic model based on a series of studies that consider the dynamics of cues from both lower and higher levels of processing during speech perception.

### **1.3. The Use of Multiple Cues in Speech Segmentation**

The psycholinguistic models of word recognition have paid limited attention to the mechanisms involved in segmenting continuous speech and accurately identifying word boundaries. Research conducted by Mattys and colleagues (Mattys, 2004; Mattys et al., 2005, 2007; Mattys & Melhorn, 2007; White et al., 2012) has focused on how listeners use information in the speech signal to segment it and achieve comprehension.

There is empirical evidence that shows that the listener's perceptual system simultaneously integrates information from multiple levels as the stream of speech unfolds. To locate word boundaries, listeners rely on a variety of cues including sentence-level information such as syntactic structure or semantic content (Dahan & Brent, 1999; Sanders & Neville, 2000), suprasegmental features such as allophonic variation or stress patterns (Quené, 1992; Sanders & Neville, 2003a), and/or sublexical information such as phonotactics (Dal Ben et al., 2021; McQueen, 1998; Mehler et al., 1981). Research from Mattys and colleagues explores the weighting of multiple segmentation strategies, aiming

to develop a hierarchical framework that assigns relative importance to each distinct cue in the process of speech processing.

Using a word monitoring task, Mattys et al. (2005) evaluated the processing of sublexical and lexical cues in English listeners. Real words (e.g., *male*) followed either a word (*calculus male*) or a nonword (*baltuluf male*) in intact and truncated contexts. Participants were asked to identify when a specific target word was present or absent in a subsequent utterance. The authors found that under normal listening conditions (intact contexts) when all cues were optimally available, listeners relied on lexical information to segment speech, with semantic and syntactic information contributing to the segmentation of lexical content.

In a series of studies, the authors highlight the substantial effect of noisy environments in the detection of and reliance on cues (Mattys, 2004; Mattys et al., 2012; Mattys & Melhorn, 2007; White et al., 2012). The main findings point to a predominant role of stress and acoustic-phonetic information over lexical and contextual cues when the signal is degraded. However, this cue reliance is reversed under intact listening conditions. The authors propose a framework of hierarchy of cues that establishes the relative dominance of lower and higher levels of cues (e.g., stress vs. acoustic-phonetics vs. semantics) in interaction with listening environments (optimal vs. noisy).

While this model proposes a rather secondary role of subphonemic cues during segmentation when other (higher-level) cues are available, the authors found that, under intact listening conditions, low-level information plays a relevant role when lexical and contextual information is limited. This was demonstrated in a follow-up experiment using a lexical decision task (Mattys et al., 2005). Words like *creMATE* were embedded in contexts favouring the full word (here, *cremate* and not *mate*), such as in *An alternative to traditional burial is to creMATE the dead*. Words like *cremate* were presented as primes. Results revealed that contextual information outweighed stress. However, when the phonotactic information was manipulated to favour *mate* instead of the full word *cremate*, no effects were found.

In summary, these studies reveal that in degraded speech signals, suprasegmental and segmental cues, such as stress and acoustic-phonetic information, have a more pronounced impact than lexical and contextual cues. These findings are accounted for by the hierarchical approach to speech segmentation, which holds that the relative

dominance of lower (stress, acoustic-phonetics) and higher (semantics) level cues, in conjunction with the listening environment (optimal vs. noisy), determines their influence on speech perception.

This series of studies tested native speakers of American English. However, as mentioned earlier in this chapter, listeners tune to a variety of cues depending on their L1. Therefore, in the following section, we will turn our attention to the primary cues that have been evidenced to be used by listeners during speech processing.

#### **1.4. The role of acoustic cues during segmentation**

As mentioned earlier in this chapter, listeners tune into the variability of speech sounds to successfully recognise words and comprehend speech. However, not all speech sounds are perceived equally across speakers. Consider the French words *bain* [bɛ̃] and *pain* [pɛ̃]. Although only the voicing of the first phoneme differs –[p] being a voiceless bilabial plosive and [b] a voiced bilabial plosive–, these words have different meanings: “bath” and “bread”, respectively. These sounds are thus contrastive in French. Contrastive features in a particular language arise solely from feature differences within speech sounds that result in distinct meanings (Anderson, 2018). If, for example, a voiced bilabial fricative [β] was produced instead of a voiced bilabial plosive [b] for 'bain' ([bɛ̃]), it would not change the meaning. Therefore, [b] and [β] are not contrastive in French, and they are considered allophones of the same phoneme /b/.

Allophones and non-contrastive variations in phoneme realisation (e.g., biologically or socio-demographically related) may not affect word recognition. However, the continuity of speech may give rise to ambiguous speech stretches that may confuse, at least, some listeners. For instance, the French phrase /dɛʁ.njɛ.kɔ̃.jɔ̃/ corresponds to both *dernier rognon* (“last kidney”) and *dernier oignon* (“last onion”) (extracted from Spinelli et al., 2003), and the English phrase /plʌmpaɪ/ can be segmented as either “plum pie” or “plump eye” (extracted from Mattys & Melhorn, 2007). The variations in speech fragments raise questions about how listeners deal with transient ambiguities that emerge from phonologically identical speech stretches. To resolve such ambiguities and disambiguate meaning, the listeners’ acoustic system must be attuned not only to contrastive features but also to non-contrastive differences present in the signal.

Several cues have been identified to guide listeners during speech segmentation, particularly those related to word boundaries. An increasing body of research has shown



that specific acoustic cues in the speech signal can serve as markers for word boundaries. Among these cues, the metric structure, also known as prosodic structure, and intonational patterns of the language play crucial roles in speech segmentation. Intonational patterns, including stress, are considered major indicators of word boundaries (e.g., Chrabaszcz et al., 2014; Zhang & Francis, 2010). The importance of prosody is emphasised in most of the literature on L1 segmentation, as it is the first auditory component of language that infants learn (Koponen, 2001). For instance, English-speaking infants as young as 6 to 9 months old display a preference for L1 stress patterns (Jusczyk et al., 1993)."

Research in adulthood demonstrates continued reliance on stress patterns as a cue for L1 segmentation in stressed languages like English, German, or Dutch. For instance, strong syllables are often linked to the onset of content words in English. In a word spotting task, Cutler & Norris (1988) used monosyllabic words like *mint* embedded in nonsense monosyllables. They showed that the detection of words was faster when words had different stress patterns. Words were easier to detect in strong-weak sequences (*MINTes*, capital letter signaling where the primary stress is placed) than in strong-strong sequences (*MINTESH*). The relevance of stress patterns was also shown in Dutch. Using a series of cross-modal priming tasks, van Donselaar et al. (2005) found that Dutch listeners were more accurate at recognising the word *Oktober* ("October") when it was presented after a priming fragment with matching stress pattern (okTO-, from *oktober*) compared to when it was presented after a mismatching fragment as a prime (OCTo-, from *octopus*). Findings from van Donselaar et al. (2005) were further replicated in an eye-tracking study, that corroborated that intonational information is efficiently used in early stages of speech processing in Dutch, modulating word recognition (Reinisch et al., 2010). These are some empirical findings that demonstrate the relevance of intonational patterns in various languages for word segmentation and lexical processing, influencing the activation of competing lexical candidates.

A well-studied prosodic feature that has been found to be a robust cue to word limits in several languages is duration. Precisely, the duration of speech segments has been shown to vary depending on their position relative to word boundaries (Klatt, 1976; Quené, 1992). Quené (1992) demonstrated through a forced-choice task that Dutch speakers can accurately achieve correct segmentation of ambiguous utterances that only varied in duration, such as in *die pin* vs. *diep in*. Namely, the duration of the pivotal consonant was

shorter in word-final positions than in word-initial positions (e.g., /p/ was shorter in *diep in* than in *die pin*). In a follow-up experiment, the duration of the pivotal consonant was manipulated to test its influence on perception. The results showed that increasing its duration (here, of /p/) led to more consonant-initial responses (*die pin*), whereas decreasing its duration resulted in more vowel-initial responses (*diep in*), which was expected from the acoustic properties of the natural productions. This study demonstrated that Dutch listeners can segment ambiguous utterances based on duration. Similar empirical data have revealed the relevance of duration in other languages such as English, French, and Korean, showing that word-initial consonants (e.g., /s/ in *I scream*) are longer than word-final consonants (e.g., /s/ in *ice cream*) (Cho & Keating, 2001; Ito & Strange, 2009; Shatzman & McQueen, 2006; Spinelli et al., 2003).

Another prosodic cue that has been shown to guide listeners during segmentation are variations in the fundamental frequency (F0) (Face, 2005; Haggard et al., 1981; Lehiste, 1970; Llisterri et al., 2002; Peperkamp et al., 2010; Warner et al., 2010; Wu, 2019). For instance, Warner et al. (2010) investigated the relationship between initial F0 rises and word onsets using a corpus of spontaneous Japanese speech from Tokyo. The authors observed that a large proportion of Japanese words were accompanied by an F0 rise (up to 68%). Furthermore, using a word spotting task, the researchers demonstrated that the initial F0 rise was used as a cue during speech segmentation. Listeners identified words such as *kazari* ("decoration") more rapidly and accurately in meaningless sequences (e.g., *rekerenikazari*) when the accentual phrase<sup>2</sup> began at the onset of the word ([ka]) rather than at the beginning of the sequence ([re]).

This section has presented some of the growing evidence that demonstrates how listeners rely on prosodic patterns in a language to segment speech. The findings also suggest that even sublexical information and subtle acoustic cues such as duration and F0 rises can indicate word boundaries in many languages. While F0 has been established as a cue to speech segmentation in several languages, its quantification varies across studies. In the following section, we present various approaches to studying the perception of F0 in speech and we discuss the main findings regarding how F0 influences speech perception.

---

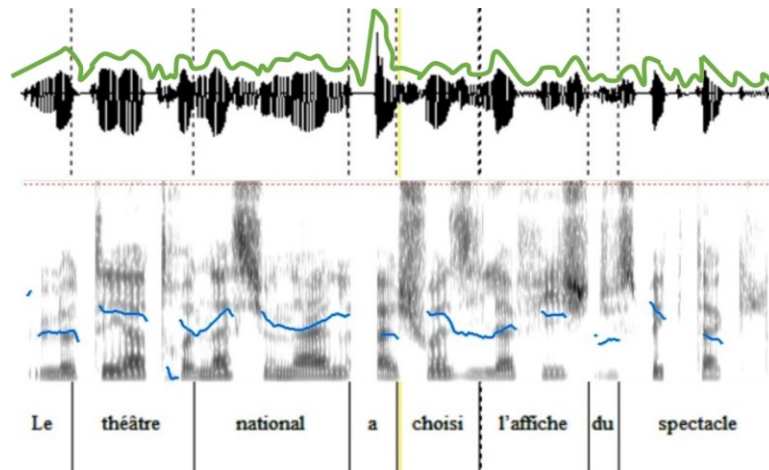
<sup>2</sup> The accentual phrase is a unit larger than a word that has been considered to be the basic unit of intonation in languages such as Japanese (e.g., Pierrehumbert & Beckman, 1989; Poser, 2000) and French (e.g., Di Cristo, 2000; Jun & Fougeron, 2002; Post, 2002).

### **1.5. Fundamental frequency as a cue to continuous speech segmentation**

The F0 is an acoustic property of language associated with the lowest frequency within a harmonic series (Ives & Patterson, 2008). F0 represents the frequency of the vibrations of the auditory signal that is transmitted through the medium (typically air). Nonetheless, F0 is sometimes referred to as “pitch”, which is the perceived frequency, i.e. the transmitted signal that is partly analysed in the ear. F0 is a relevant cue to speech, encompassing modulations that serve a variety of functions depending on the language. These functions include marking syllables, words, or tones, conveying information about speech acts like statements or questions, and expressing emotions. Figure 1 illustrates these speech properties.

The acoustic features of F0 as a segmentation cue have been the focus of numerous studies centered on speech intelligibility. Research has traditionally contrasted natural speech with speech in which the F0 contours are artificially flattened (Assmann, 1999; Binns & Culling, 2007; Laures & Weismer, 1999). Overall, the flattening of F0 resulted in the detriment of sentence intelligibility. In the study by Laures & Weismer (1999), low-predictability sentences were resynthesised and the F0 contour was flattened to the mean F0 value of the whole sentence. Participants listened to both natural and monotonous sentences presented in a background white noise. Listeners found monotonous sentences less intelligible and made more errors identifying words compared to natural sentences. The authors concluded that flattening F0 contours removes linguistic cues that drive attention to content words and guide listeners during speech comprehension (see also Cutler & Foss, 1977, and Binns & Culling, 2007).

Most studies examining the role of F0 in speech perception and segmentation predominantly focus on its mean value. However, natural utterances exhibit patterns of gliding, in which F0 varies rather continuously, as depicted in Figure 1. In fact, several studies have characterised the F0 pattern modulations observed in the data as displaying "elbows" or rising and/or falling features (e.g., d'Imperio et al., 2007; Haggard et al., 1981; Padeloup, 1990; Post, 2002; Welby, 2007). As a consequence, it becomes interesting to consider the F0 segmentation cue within the context of its temporal dynamics, i.e. the value of its slope.



**Figure 1.** Signal waveform and spectrogram of the sentence “Le théâtre national a choisi l’affiche du spectacle”, one of the sentences used in Experiments 1 and 2 (Chapter 6) in this dissertation. The green line illustrates the F0 modulations of the sentence. The blue line shows the pitch values throughout the sentence.

There is empirical evidence showing the role of F0 variations during continuous speech segmentation in several languages (e.g., Haggard et al., 1981, for British English; Laures & Weismer, 1999, for American English; Tremblay et al., 2019, for Korean; Weber et al., 2006, for German; Wu, 2019, for Mandarin). For instance, there is one recent study suggesting the role of the F0 slope value in speech segmentation in Korean. In Korean, the expected intonational pattern is L(HL)H tones (with L=Low and H=high). Thus, a word final H tone followed by an initial L tone would signal a word boundary. In an artificial language learning study, Tremblay et al. (2019) varied both the timing of the peak in the final H tones (early vs. late in the last syllable) and the value of the initial L tones (High, Mid, Low). Their results showed that Korean listeners mainly used the difference in pitch value (scaling) between final H and initial L tones (signalling a boundary) while segmenting the speech stream. Raising the value of the initial L tone disrupted segmentation performance. Moreover, in the specific cases where listeners could not rely on this scaling information (i.e. when initial L tones were raised), they used the slope between final H and initial L tones to achieve correct segmentation. When the slope between final H tones and initial L tones of the following words was steeper (as provided by the late peak in the last syllable), segmentation performance was enhanced, suggesting that listeners used the slope values to locate word boundaries. These findings suggest that variations in F0 slope values guide listeners during speech segmentation.

## **1.6. Summary and conclusion**

The flexibility of the listeners' auditory system allows them to navigate through the variability and continuity of the signal to achieve successful segmentation and comprehension. Substantial evidence has demonstrated the reliance on sublexical information during speech segmentation. However, the presence of higher-level information leads to a recalibration of cues. These aspects are acknowledged in most of the psycholinguistic models of word recognition, which propose that as the sounds within a word unfold and higher-level cues become available, lexical competition is reduced. However, these models scarcely address the specific process of speech segmentation.

The hierarchical framework proposed by Mattys et al. (2005) captures the observation that while each cue may independently influence the activation of lexical candidates, certain cues take precedence over others when multiple cues are present. The weighting assigned to these cues varies based on the saliency or accessibility of other cues within the speech signal at any given moment. As Shoemaker (2009) discusses, integrating this framework with existing studies on processing and segmentation cues reveals that no single segmentation cue can be considered necessary or sufficient for the comprehensive processing of natural speech.

In the following chapter, we will introduce the main findings on the cues used to segment French, which is the language of study of the present dissertation.

*Speech segmentation is an elusive construct insofar as it is often little more than a by-product of word recognition.*  
(Sven L. Mattys and Heather Bortfeld, 2017: 69)

## **CHAPTER 2. Speech segmentation in French**

In order to elucidate the mechanisms underlying word recognition, psycholinguistic models of word recognition propose units that work as a bridge between the acoustic signal and the processing system (the brain). The unit that most models consider as input is the phoneme (Cohort, NAM, PARSYN) or a string of phonemes (TRACE, Shortlist). However, individual languages exhibit varied rhythm structures, thereby providing different cues to speech. For instance, stress serves as an indicator of word boundaries in English (Cutler & Norris, 1988a), while listeners in French typically rely on syllables (Banel & Bacri, 1997). Therefore, in this chapter, we will present relevant cues evidenced to segment French speech.

### **2.1. Cues to speech segmentation in French**

Research has shown that the syllable is a crucial perceptual unit in French and it has a relevant role during segmentation (Content et al., 2001; Dumay et al., 2002; Mehler et al., 1981). For instance, Mehler et al. (1981) exposed native French speakers to pairs of words such as *balcon* and *balance* during a syllable-monitoring task. The bisyllabic items differed in terms of the syllable boundary's position relative to the first three segments. In *bal.con* the boundary occurred between the consonants /l/ and /c/, whereas in *ba.lance* the boundary occurred between the /a/ and the /l/. The results revealed that listeners exhibited faster response times when monitoring for syllable sequences like /ba/ or /bal/ when the target fully matched the first syllable of the word. In other words, responses to combinations such as /ba/—"balance" and /bal/—"balcon" were faster than to inverse combinations such as /bal/—"balance" and /ba/—"balcon". Dumay et al. (1998) showed that French speakers were significantly faster identifying *lac* when syllable onsets were aligned (such as in the non-word ZUN.LAC –/nl/ is not a possible onset in French–) compared to cases where *lac* was not aligned to the syllable onset (such as in ZU.GLAC –/gl/ is a possible onset in French, which may misalign syllable and word boundaries–). These studies evidence that French listeners rely on the syllable as a cue for inferring word boundaries.

Nonetheless, other acoustic cues have been found to be exploited by French listeners. A growing body of evidence suggests that French listeners are sensitive to the prosodic patterns of their L1, which are used during word segmentation. Even though French has no lexical stress, it exhibits phrasal prominence. The prosodic structure in French is marked by the accentual phrase (AP), a unit of phrasing below the utterance level (Jun & Fougeron, 2000, 2002; Welby, 2003, 2006). An AP in French encompasses one or more content words (such as nouns and verbs) as well as any preceding function words (such as determiners and conjunctions), and its final syllable is lengthened. The final non-reduced syllable of the last word within the AP is characterised by an F0 rise and it is lengthened, but this occurs exclusively in positions that are not at the end of the utterance. This late (final) rise serves as an indicator of the so-called *primary stress* or *primary accent*. An early F0 rise is occasionally observed close to the beginning of the AP. This early rise is a marker of *secondary stress* or *secondary accent*, though it does not typically change meaning. Thus, the intonational pattern in French follows a two-rise configuration characterised by a Low-High-Low-High (LHLH) sequence.

Empirical support for the use of intonational information during the segmentation of French is evident in a study conducted by Banel and Bacri (1994). Their research demonstrated that French listeners are able to exploit the rhythmic patterns of their L1 for effective word segmentation. In instances involving ambiguous sequences like /ba.gaz/ (*bagage* “luggage” or *bas gage* “low pledge”), listeners were more likely to perceive a single word (*bagage*) when the second syllable was lengthened and two words (*bas gage*) when the first syllable was lengthened. This observation aligns with the assumption that a syllable positioned at the end of a phrase experiences lengthening and that phrase boundaries are not expected to occur in the middle of a word. Moreover, Rietveld (1980) found phonetic differences in F0, intensity, and duration between minimal pairs such as *le comtat saccagé* (“the devastated county”) and *le comte a saccagé* (“the count has laid waste”). These differences corresponded to a primary accent and a late rise on the final syllable /ta/ of *le comtat* (/lə#kɔ̃.ta/) and a primary accent and late rise on the /kɔ̃/ of *le comte a* (/lə#kɔ̃.t#a/). Moreover, they found a stretching of the final vowel of the noun. The study highlighted the capacity of French listeners to discern the intended meanings of such minimal pairs, with durational differences being the most reliable cues.

Research by Welby (2003, 2007) and Spinelli et al. (2010) has demonstrated that French speakers are sensitive to simpler variations of the two-rise L1H1L2H2 pattern during

word segmentation. It was proven that the first early rise (L1H1) and the simple F0 inflection unaccompanied by a subsequent rise serve as cues to content-word onsets. For instance, listeners interpreted meaningless sequences as a single non-word (*mélamondine*) when the F0 rise began at the first syllable ([me]), and as two words (*mes#lamondines*, “my lamondines”) when the F0 rise started at the second syllable ([la]) (Welby, 2003, 2007). The author suggested that the F0 rise could mark the onset of content words, a proposition that finds support in previous empirical evidence (Di Cristo, 2000; Vaissière, 1997). Spinelli et al. (2010) exposed French listeners to ambiguous phrases containing a content word preceded by a function word, such as *l’affiche* (“the poster”) versus *la fiche* (“the sheet”), both realised as [la.fiʃ]. An increase in the F0 of the vowel /a/ within function words led to the enhanced perception of vowel-initial content words (such as “affiche”), thereby also enhancing the activation of lexical targets with vowel-initial forms within the lexicon.

Research on the use of intonational information in French is part of the growing evidence showing a “tonal alignment” between the timing of the F0 peaks and valleys with respect to segmental markers such as consonants, vowels, and syllables, with variations dependent on the specific language (e.g., Ladd & Schepman, 2003, and Pierrehumbert, 1980, for English; Pierrehumbert & Beckman, 1989, for Japanese; Prieto et al., 1995, for Spanish). Nevertheless, French possesses phonological phenomena that misalign both syllable and word boundaries, like *liaison* or *elision*. As a consequence, potential ambiguity is introduced into the auditory signal, which challenges the notion that the syllable consistently functions as a reliable cue for speech segmentation in French. Furthermore, this intricacy creates a challenging scenario for models of word recognition to address and accommodate.

## 2.2. Segmenting ambiguity in French: The case of resyllabification

In Chapter 1, we have discussed how listeners manage to navigate through the continuous and variable nature of speech to attain comprehension. However, French presents certain phenomena that can pose challenges to the process of speech segmentation. Notably, French possesses three phonological phenomena that occur at the junction between two words: *liaison*, *enchaînement*, and *elision*. These are referred to as sandhi<sup>3</sup> phenomena

---

<sup>3</sup> A term used for a wide variety of sound changes that occur at morpheme or word boundaries (Andersen, 1986).



and they frequently introduce ambiguity in the location of both syllable and word boundaries.

French sandhi phenomena occur between two adjacent words, leading to adjustments in adjacent sounds (Sampson, 2016). The particular case of liaison occurs when a word-final consonant is only pronounced if followed by a vowel. For example, phrases like *un air* (“a tune”) and *un nerf* (“a nerve”) become phonemically identical due to the liaison between the consonant /n/ and the following vowel /a/ in “un air”. Liaison disrupts the alignment of the function word “un” (determiner) and the content word “air” (noun). Consequently, both phrases are realised as [œ̃.nɛʁ] (note that the dot delimits syllable boundaries) (example extracted from Spinelli et al., 2003).

In French, word-final consonants can be syllabified with the vowel of the following vowel-initial word, which is called *enchainment* (Andersen, 1986). Consider the phrase *petit chou* [pə.ti#.ʃu] (“little cabbage”) (note that the # delimits word boundaries) versus *petit ami* [pə.ti.t#a.mi] (“boyfriend”) (example extracted from Durand & Lyche, 2008). *Petit ami* is resyllabified by *enchainment* as [pə.ti.t#a.mi] instead of [pə.tit#a.mi], where word boundaries between adjective and noun would be respected.

The third resyllabification phenomenon is elision, which occurs when a function word with a vowel in its final position is followed by a vowel-initial (or [h]-initial) word (Andersen, 1986). This phenomenon is orthographically represented by an apostrophe (‘). Elision can be observed in the phrase *l’éléphant* (“the elephant”), where the final vowel of the function word “le” is elided before the vowel-initial word “éléphant”. Consequently, the pronunciation becomes /l#e.le.fã/ instead of /le#e.le.fã/.

These resyllabification phenomena may hinder speech segmentation for French speakers. However, several studies testing the perception of liaison suggest that resyllabification does not seem to compromise the listeners’ perceptual abilities during French processing. As an example, using a set of phrases within liaison environments, Spinelli et al. (2002) found better performance in identifying vowel-initial words in liaison (such as *agneau* in *petit agneau* [pə. ti. t#a. no] ‘little lamb’) than in illegal liaison contexts (such as *agneau* in *demi t agneau* [də. mi. #t#a. no] ‘half \*t lamb’, where ‘t’ cannot occur). These results were argued as a facilitative aspect of the lexical knowledge concerning liaison, supporting target recognition in adequate contexts while impoverishing its identification when the context does not allow liaison (illegal contexts). In a different study using two

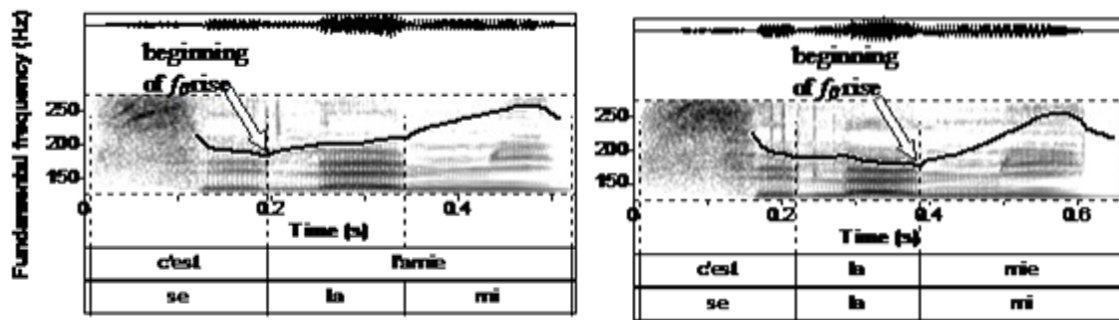
priming experiments, Gaskell et al. (2002) demonstrated comparable levels of lexical activation during the perception of both aligned and misaligned syllables in phrases such as *un généreux italien* (“a generous Italian”, /œ̃#ʒe.ne.ʁø.z#i.ta.lj'ɛ/) and *un chapeau italien* (“an italian hat”, /œ̃#ʃa.po#i.ta.lj'ɛ/). Furthermore, a word monitoring task revealed a clear advantage for the misaligned conditions in recognising the subsequent word. As posited by the authors, these findings do not support the idea that resyllabification hampers or delays the recognition of the following word. Instead, they suggest a potential advantage of resyllabification. The authors proposed that the decrease in the duration of consonants associated with resyllabification, along with other subtle differences, may have been exploited by the perceptual system of listeners. Additionally, in a cross-modal priming study, Spinelli et al. (2003) observed the facilitatory effects of the liaison consonant [ʁ] in phrases such as *c'est le dernier oignon* (“it's the last onion”) versus the non-liaison consonant [ʁ] in *c'est le dernier rognon* (“it's the last kidney”), both sentences syllabified and produced as [se.lə.dɛʁ.nje.ʁɔ̃.ɲɔ̃].

The findings presented in this section evidence that correct target recognition in liaison contexts seems to occur through the use of several acoustic-phonetic cues. Prior research observed differences between liaison consonants and their consonant counterparts. Delattre (1940) initially observed that consonants involved in liaison seem weaker compared to the same segments in word-initial positions. Subsequent research has validated this observation for several consonants (t, n, r, z, g, p) and revealed durational differences between these consonants in both liaison environments and word onsets (M. G. Gaskell et al., 2002; Spinelli et al., 2003; Wauquier-Gravelines & Braud, 2005). The study by Shoemaker (2014) provided further support for such differences and showed that French listeners exploit duration as a source for the localisation of word boundaries. Overall, analyses of the materials used across studies revealed that word-initial consonants are on average 10 ms longer (difference range: 6–12 ms) than consonants in liaison environments.

Several studies have examined the liaison phenomenon, with its investigation dating back to as early as 1940. As summarised earlier in this chapter, the existing literature on liaison has primarily observed durational and F0 differences. However, relatively limited research focuses on the resyllabification phenomenon of elision. Are there any prosodic properties associated with elision? Does this resyllabification phenomenon entail cues to guide listeners during segmentation? We will discuss this in the following section.

### 2.3. Cues to elision in French

Previous studies have found differences in acoustic details to elision contexts in French. In a study by Spinelli et al. (2007), they used a set of phonemically identical phrases constructed by a noun preceded by a definite article, such as *l'ami-la mie* [la.mi], *l'affiche-la fiche* [la.fi]. The authors examined the variability of the acoustic measures in two tokens of each item and found acoustic differences in their material. They found small duration differences of 7 ms on average between the pivotal vowel in items. Particularly, analyses revealed that the [a] was shorter in content words (e.g., “l’ami”) than in the function word (“la”) across items. Formant values were also different in elision vowels. F2 were lower when [a] was the onset of content words compared to the [a] in function words. The authors associated higher F2 with a more frontal articulation of [a] for the article. Additionally, [a] had lower F0 values in function words than in content words, with a difference of 16 Hz. The onset of [la] in *l'ami* exhibited a higher F0 value, while in *la mie*, the higher F0 occurred later at [mi], coinciding with the onset of the content word. See an illustration in Figure 2. This finding is consistent with results from other investigations that explored elision in French and found a higher F0 across the first syllable, which was interpreted as a marker of content-word onsets (Spinelli et al., 2010; Welby, 2003, 2007).



**Figure 2.** Fundamental frequency rise in the phonemically identical phrases *C'est la mie* (left panel) and *C'est la mie* (right panel), both /selami/. The spectrogram shows the F0 modulations along the phrases. F0 rise is marked by a white arrow, marking the beginning of the content words (*amie* and *mie*, respectively). Y-axis shows F0 Hz, x-axis shows the duration of the segment in seconds.

Besides variations observed in the pivotal vowel in elision, Spinelli et al. (2007) also found acoustic dissimilarities in adjacent phonemes. The preceding phoneme [l] was longer in elided phrases (e.g., *l'ami*) compared to non-elided ones (e.g., *la mie*), with an average difference of 5 ms. The first syllable [la] was longer for elided phrases (*l'ami*) than for non-elided phrases (*la mie*). Additionally, the articulatory position of second

vowels in elided content words differed from that in function words. For example, back vowels were produced with a more posterior tongue position in consonant-initial phrases like *la location* (“the rental”) than in vowel-initial ones (*l’allocation*, “the grant”).

Research using ambiguous phrases in elision contexts has provided evidence indicating that the acoustic information in such items is exploited by French listeners (Do Carmo-Blanco et al., 2019; Spinelli et al., 2007, 2010). Using an ABX discrimination task, Spinelli et al. (2007) showed that when presented with isolated homophonic phrases like *l’ami–la mie* (both [lami]), French speakers exhibited an average correct discrimination rate of 66.31% when distinguishing between the items. In a follow-up experiment, participants were presented with only one of the items (e.g., *l’affiche*) and were asked to identify it within the two possible phrases (*affiche–fiche*). Listeners correctly identified the target with an average accuracy of 75.48%. Findings from these studies indicate that fine-grained acoustic detail is sufficiently salient and plays a role in guiding listeners during the segmentation of phrases in elision environments when presented in isolation. Moreover, through a forced-choice identification task and a lexical decision task, Spinelli et al. (2010) showed that the manipulation of the F0 rise modulates the perception of vowel-initial and consonant-initial items in elision environments. The authors showed that raising the F0 mean value in the pivotal [a] in consonant-initial productions like *la fiche* resulted in the perception of more vowel-initial items (*l’affiche*). Remarkably, this rise also led to an increased activation of the lexical representation of vowel-initial targets (in this case, *affiche*) compared to consonant-initial ones (*fiche*).

Recent research using electrophysiological measures (EEG) has provided insights into the neural correlates of the perception of fine-grained acoustic details in elision. In a study employing the mismatch negativity (MMN) paradigm, Do Carmo-Blanco et al. (2019) investigated the relevance of non-contrastive features found in elision environments during speech segmentation in French. The MMN paradigm is commonly used in EEG studies to assess the deviation from sensory predictions and discriminate relevant and irrelevant prosodic and phonemic cues in the signal. In a first experiment, participants were presented with the initial syllables of the homophonic phrases *l’allocution* and *la locution* (i.e. l#a and la#). A subsequent experiment included the complete phrases. Results revealed that French listeners are sensitive to differences between elided and non-elided items, in both isolated syllables and phrases. These effects were found significant despite variations in productions across speakers. The authors argued that the

representations of the speech units in memory are stored with their subphonemic information. However, results did not exhibit differences in the topographical distribution of items, which was interpreted as an absence of activation of competing candidates during lexical access.

Overall, the empirical evidence presented in this section suggests that elision affects not only the pivotal vowel ([a]) but also neighbouring phonemes, thereby providing cues that guide listeners in the segmentation of intended targets within elision. In this chapter, we have shown evidence that is consistent with existing evidence highlighting the perceptual system's sensitivity to subphonemic information (see Chapter 1 for a concise overview). We have shown this also for the particular cases of items that are affected by resyllabification. However, while both elision and liaison are phonological phenomena that lead to misalignments in syllable and word boundaries, a noteworthy distinction between the two lies in the pivotal phoneme involved: consonants in liaison contexts and vowels in elision contexts.

Empirical evidence has revealed that consonant and vowel information fulfil different roles during lexical activation and word recognition processes. Studies using a word reconstruction paradigm (Cutler et al., 2000; van Ooijen, 1996) have provided insights into the differing impact of vowel and consonant information on speech perception. Specifically, these studies demonstrate that vowel information has less influence on listeners compared to consonant information. This effect has been associated with the idea that consonants carry more weight than vowels during the processes of lexical retrieval. For instance, when presented with the nonword “kebra”, listeners are more likely to choose *cobra* than *zebra* in order to transform it into a real word (Cutler et al., 2000). Differences in the functional roles of liaison and elision are further discussed in the General Discussion.

The findings related to resyllabified phenomena in French sparked the interest of this dissertation in examining the perceptual sensitivity to cues involved in processing the common phenomenon of elision in French. The present dissertation aims to provide further insight into the role of fine-grained acoustic details during the segmentation of speech in the context of elision in French.

## **2.4. Summary and conclusion**

Taken together, several studies have found distinctive acoustic characteristics in resyllabified sounds compared to their non-resyllabified counterparts. Remarkably, in both liaison and elision, there appears to be an intonational pattern characterised by an F0 rise that seems to indicate the beginning of content words. This effect aligns with previous work from Di Cristo (2000), Padeloup (1990), and Vaissière (1983, 1997), who proposed that variations in F0 as an intonational rise could serve as an indicator of word onset and facilitate lexical access in French. Additionally, in elision, there have been observed variations in duration and formant characteristics in surrounding phonemes around the pivotal vowel. Particularly, it seems that durational differences observed in elision contexts are on average smaller than those in liaison contexts.

It is worth noting that, while liaison environments involve consonant phonemes, elision environments involve vocalic phonemes. The variety of acoustic properties associated with each of these phonemes, as well as their distinct functional roles, suggests that the perception of liaison and elision extends beyond cues for segmentation and misalignment of syllable and word boundaries. Specifically, the pivotal phonemes involved in liaison may also exhibit differences that are linked to lexical processing, whereas elision may entail variations in structural information.

Despite the potential challenges that French resyllabification phenomena may present for syllable-based word segmentation strategies, a growing body of research indicates that there are no processing costs associated with the perception of words that are not aligned with syllable boundaries in French. Overall, findings from the perception of resyllabification phenomena in French suggest that listeners rely on the prosodic information related to word boundaries, such as acoustic-phonetic cues, to compensate for syllable misalignments during speech segmentation.

Chapters 1 and 2 of this dissertation have presented the roles of various cues in speech processing, highlighting their importance in segmenting speech and facilitating word recognition. Chapter 2, in particular, has explored a potential challenge in French spoken language: resyllabification phenomena. We have examined several studies that demonstrate the significance of acoustic-phonetic cues in resolving ambiguity within liaison contexts. Moreover, we have reviewed the scarce studies that examine the perception of ambiguous phrases in elision environments in isolation. The main goal of the current thesis is to contribute to the existing evidence on the perception of elision in French and provide insight into the role of these cues when other information is available.

In Chapter 6, we aimed to study these cues in interaction with higher-level cues (contextual) to assess the limits (or *flexibility*) of the auditory system in French speakers. There has been limited research examining to what extent listeners rely on intonational cues during word segmentation when they have access to sentential information. Chapter 7 introduces a paradigm that allowed us to explore the relative importance of F0 features during elision.

Additionally, we wanted to extend the understanding of the processing of elision in populations that might face difficulties in processing spoken language. To achieve this, we examined the sensitivity to F0 features in adults with developmental dyslexia, who were matched in age to control participants (Chapter 7). Finally, we evaluated the discrimination abilities of adult French learners in Chapter 8. Consequently, Chapters 3 and 4 offer a concise overview of the main challenges related to speech processing in individuals with dyslexia and language learners, respectively.

In Chapter 5, we conduct acoustic analyses of the material used across our experiments, employing analogous acoustic metrics as in previous studies, including F0, duration, and formant values. We perform additional analyses to that consider adjacent consonants. This aspect has not been extensively explored and has the potential to reveal differences between elided and non-elided items.

*Those who have spent many hours debugging computer programs or repairing broken machines know that the main difficulty in accomplishing these tasks consists in figuring out what the machine does to accomplish a task. To have any hope of success, one must try to picture the state in which it is stuck, to understand how it interprets the incoming signals and to identify which intervention will bring it back to the desired state.*  
(Dehaene, 2009: 232)

## **CHAPTER 3. Speech segmentation in adults with developmental dyslexia**

### **3.1. The challenge of segmentation**

Developmental dyslexia (hereafter, dyslexia) is a learning disorder with a neurobiological origin that has been widely researched and affects 1-7% of the population (e.g., Lindgren et al., 1985; Noordenbos, 2013; Saksida et al., 2016). It does not only affect the early stages of development but persists into adulthood. It is primarily characterised by impaired reading and spelling abilities, although the difficulties are not confined to reading. Dyslexia is also associated with phonological processing deficits, which involve difficulties in accessing, processing, and manipulating speech sounds (e.g., Blomert, 2011; Ramus et al., 2003; Ramus, 2014a; Ramus & Szenkovits, 2008; Snowling, 1995; Szenkovits & Ramus, 2005; Vellutino et al., 2004). This has been observed through the use of a variety of tasks such as sound segment manipulation, non-word reading, rapid picture and digit naming, short-term memory tasks, and sound categorization and discrimination (Bogliotti et al., 2008; de Carvalho et al., 2014; Godfrey et al., 1981; Goswami, 2011; Hämäläinen et al., 2012; Noordenbos & Serniclaes, 2015; Schraeyen et al., 2019; Tallal, 1980; Virtala et al., 2020).

The hallmarks associated with dyslexia have been documented across various languages (e.g., English, Spanish, and Chinese: Goswami et al., 2011; English, German: Landerl & Wimmer, 2000, Ziegler et al., 2003; French: Muneaux et al., 2004) and cannot be accounted for by low IQ level, poor education, or sensory or neurological deficits. However, the impact of these difficulties may vary over the lifespan of individuals due to compensatory mechanisms (Cavalli et al., 2017; Kershner & Micallef, 1992; McNulty, 2003; Moojen et al., 2020; Shaywitz et al., 2003).

Research has demonstrated that some dyslexic adults can compensate for their impairments and minimise the impact on their reading skills. Compensated dyslexics seem to achieve successful word reading by employing various top-down strategies, such



as using contextual cues (Nation & Snowling, 2008) or semantic knowledge (Snowling, 2001), harnessing visual memory (Campbell & Butterworth, 1985), or tapping into morphological knowledge (Elbro & Arnbak, 1996). These strategies allow them to circumvent the difficulties experienced in their phonological skills. Nevertheless, although adults with dyslexia may attain functional reading skills for every day reading situations, this does not necessarily extend to their phonological abilities. Rather, they consistently exhibit poor performance on phonological processing tasks (e.g., Elbro et al., 1994; Gruber, 2003).

The nature and causes of the phonological deficit in dyslexia are still under debate. Some authors propose that phonological deficits in dyslexia can be attributed to a limited capacity for segmenting phonological representations of language; that is, the encoding of abstract representations of speech sounds in the brain. These difficulties seem to encompass deficits in the identification, discrimination, and categorisation of speech sounds (Blomert & Mitterer, 2004; Chiappe et al., 2001; Paul et al., 2006; Serniclaes et al., 2001; Vandermosten et al., 2010; Ziegler et al., 2009). As an illustration of these difficulties, dyslexics have been found to exhibit poorer phonemic discrimination compared to typical readers. For example, dyslexics make more errors than typical readers when tasked with discriminating between pairs of spoken syllables differing by a single phoneme (e.g., /ba/ and /da/) (Adlard & Hazan, 1998; Bogliotti et al., 2002; Schulte-Körne et al., 2001). On the other hand, in the case of certain allophones, individuals with dyslexia seem to display better discrimination abilities than typical readers, with their perception being more allophonic than categorical. For instance, dyslexic children showed better performance distinguishing acoustic variations among stimuli within the same phonemic category compared to their typically developing peers (e.g., Serniclaes et al., 2001, 2004). This evidence indicates that dyslexic individuals tend to exhibit less categorical speech perception, as they show no difficulties in recognising differences within a given category.

The phonological deficit in dyslexia encompasses more than just specific difficulties in phonological processing tasks. These also include perceptual difficulties during both speech and non-speech processing (e.g., Baldeweg et al., 1999; Farmer & Klein, 1995; Kujala et al., 2000; Kujala & Näätänen, 2001; McGivern et al., 1991; Reed, 1989; Tallal, 1980; Temple et al., 2000; Wright et al., 2000). It has been proposed that a major difficulty in speech perception in dyslexia is the perception of short and rapidly occurring elements

(Farmer & Klein, 1995; McArthur & Bishop, 2001; Tallal, 1980; Temple et al., 2000). For example, in a study conducted by Tallal (1980), children with dyslexia and the control group did not exhibit significant differences in discriminating tones presented at slow rates (i.e. tones consistently displayed with a 428-ms time interval). However, dyslexic children demonstrated poorer performance in discriminating rapidly presented tones (i.e. intervals randomly decreased to 8, 15, 30, 60, 150, or 305 ms). These tones with varying interval durations were presented along three tasks. In the same-different discrimination test, participants were required to differentiate between identical and different tone sequences, while in the sequencing and rapid perception tests participants had to determine the order of the tone sequences after they were presented. Results revealed similar performance between dyslexics and controls when the interval durations between tones were longer (428 ms). However, when the interval durations were reduced, dyslexics made significantly more errors than controls in both tone sequencing and tone discrimination tasks. Moreover, the author found correlations between the number of errors and the reading ability of dyslexic children, particularly with reading nonsense words –a task that requires the use of the knowledge of phonetic rules–. These findings indicated that dyslexic children exhibit a deficient auditory processing compared to controls. The author suggested that the impaired perception of temporal patterns in dyslexic children might stem from a more fundamental perceptual deficit affecting their ability to process perceptual information at an appropriate rate. However, other studies did not find such effects (Adlard & Hazan, 1998; Rosen, 2003; Rosen & Manganari, 2001).

While the perception of some acoustic attributes such as frequency (Corbera et al., 2006) and pitch (Baldeweg et al., 1999) has been shown to be impaired in dyslexia, difficulties in the perception of other features such as duration are less straightforward. The discrimination of short-sound durations (100 vs. 33 ms) seems to be deficient in dyslexic individuals (Corbera et al., 2006); however, discriminating longer sound durations (200 vs. 160, 120, 80, and 40 ms) does not seem to be impaired in dyslexia (Baldeweg et al., 1999). Remarkably, when sounds are embedded in complex tone patterns, the discrimination of sounds based on their duration was poorer in individuals with dyslexia compared to controls (Kujala et al., 2000, 2003; Kujala et al., 2006; Schulte-Körne et al., 1999). Therefore, it seems that whereas the discrimination of some sound features (e.g., stimulus frequency or duration of brief sounds) is impaired at the automatic level of

information processing (e.g., Baldeweg et al., 1999; Kujala et al., 2003; Renvall & Hari, 2003), the processing of some other acoustic features (e.g., duration of long stimuli) becomes compromised when the sounds are embedded in longer sound patterns. These findings reveal the complexity of the perception difficulties present in dyslexia.

Research has focused on other aspects of low-level auditory processing, such as the perception of prosody and rhythm. One of the most examined sensory abilities in dyslexia is auditory frequency discrimination, which has served as an index of sensory processing. Electrophysiological evidence has indicated that pitch perception probably entails the contribution of a group of regions both within and outside the auditory cortex (Kumar & Schönwiesner, 2012). It has been suggested that since pitch perception is distributed over different brain areas, the auditory deficits in dyslexia may also arise from higher levels of processing other than lower auditory processing. As suggested by Witton et al. (2020), close interactions between the processing of sensory information and relevant cognitive information may modulate associations between frequency discrimination and reading.

Evidence from neuroimaging studies has linked the deficits in dyslexia to major abnormalities in cerebral connectivity, as well as in cortical structure, particularly in the left hemisphere language network (e.g., Hampson et al., 2006; Xia et al., 2016). In an EEG study by Breznitz & Meyler (2003), dyslexic adults showed a slow recognition of letters and a deficient sequencing of both letters and sounds compared to controls. Both linguistic and non-linguistic auditory stimuli were presented to participants. While stimuli processing started in the left hemisphere for normal readers, its processing began in the right hemisphere for dyslexics. The authors suggested a tendency of dyslexics to engage the right hemisphere, which is not specialised in language processing, possibly due to deficits in the left hemisphere. This reliance on the right hemisphere may result in less efficient processing, leading to slower processing times for individuals with dyslexia. The authors attributed these difficulties to impaired temporal processing, rather than solely related to decoding deficits.

Evidence from various approaches highlights the intricate nature of the phonological deficit in dyslexia. The perceptual impairment in dyslexia has been shown to impact language-related processing tasks, such as phoneme categorisation, but also the discrimination of rapidly presented non-speech sounds. Furthermore, neuroimaging studies indicate that these difficulties may stem from an underlying cognitive processing deficit.

As suggested by Kujala et al. (2006), phonological impairments in dyslexia specifically target the capacity to effectively establish connections between the written language's code (orthography) and the language's sound structure (phonology). According to this perspective, some research proposes that deficiencies in phonological processing stem from a broader challenge in perceiving acoustic information accurately (Baldeweg et al., 1999; Farmer & Klein, 1995; Kujala et al., 2000; Kujala & Näätänen, 2001; McGivern et al., 1991; Reed, 1989; Tallal, 1980; Wright et al., 2000). This perspective suggests that a more general deficit in acoustic processing hampers the precise perception of essential acoustic components within spoken language, thereby disrupting the formation of an accurate phonological code. Therefore, it is important to address the extent to which phonological awareness (PA, hereafter), the ability to recognise and manipulate sound units in words, impacts dyslexia.

### **3.2. Phonological awareness**

In spoken language, individuals possess the ability to generate an unlimited number of words by combining and rearranging a limited set of phonological segments, specifically consonants, and vowels. These phonological segments serve as the fundamental building blocks of the biological system for language. For instance, the French words “classe” ([klas]) and “glace” ([glas]) have plosive velar consonants as onset (/k/ and /g/, respectively). However, while /k/ is voiceless, /g/ is voiced. Therefore, to properly recognise the intended words, the phonological representations of all the sounds must be robust (i.e. must be contained within the same phonemic category) and distinct (i.e. /k/ must be distinguished from /g/) (Ladefoged, 2001).

Through the adoption of alphabetic writing systems, such linguistic potential is extended to individuals who engage in reading, albeit contingent upon their ability to associate the arbitrary characters (letters) with their corresponding phonological segments. Establishing this association requires an understanding that all words can be deconstructed into phonological segments (Shaywitz et al., 1998; Sprenger-Charolles & Colé, 2013). Hence, it is this understanding (i.e. PA) that enables readers to establish a connection between the letter sequences (orthography) and the corresponding units of speech (phonological constituents) that they represent. Hence, it has been proposed that PA may be a result of recognising individual words rather than a prerequisite for this process (Castles & Friedmann, 2014). Moreover, the speed and efficiency of representations at the phonemic level are proposed to vary with the degree of orthographic

transparency of the language, i.e. how transparent this grapheme-phoneme relationship is to the learner (e.g, Borleffs et al., 2019; Landerl et al., 2013; Ziegler et al., 2003a, 2003b; Ziegler & Ferrand, 1998; Ziegler & Goswami, 2005).

Extensive research indicates a notable deficiency in PA among individuals with dyslexia, both in children and adults, which in turn affects text comprehension (e.g., Bradley & Bryant, 1983; Stein, 2018). For instance, in a functional magnetic resonance imaging (fMRI) study, dyslexic adults and age-matched controls were presented with combinations of letters and speech sounds that were either congruent or incongruent (Blau et al., 2009). Unlike the findings observed in the control group, no enhanced activation in the superior temporal gyrus was detected among dyslexic participants in response to letter-speech sound congruent pairs as compared to incongruent pairs. The authors interpreted these results as indicative of a deficient automatic detection of letter-speech congruencies, suggesting a reduced integration of letter-speech sound information in individuals with dyslexia.

Moreover, tasks assessing PA in spoken language have also shown poor performance for dyslexics compared to controls. For instance, dyslexic adults experience difficulties in tasks assessing PA such as phoneme deletion (Cavalli et al., 2016a; Martin et al., 2010), verbal short-term memory (Majerus & Cowan, 2016; Paulesu et al., 2001; Vasic et al., 2008), and speech sound discrimination (Berent et al., 2012).

Awareness of phonemes in speech has traditionally been considered to be generally restricted to users of alphabetic written languages (Charles-Luce & Luce, 1990; Marslen-Wilson & Warren, 1994). However, recent accounts propose that PA is an “emergent property of acoustic structure” and suggest that speech rhythm is a primary cue to encode language during language development (Goswami, 2011; Leong & Goswami, 2014a). Moreover, difficulties in dyslexia cannot be solely attributed to deficient PA, as several studies have revealed difficulties among dyslexic individuals of non-alphabetic languages, such as Chinese, in reading (e.g., Daniels & Share, 2018; Siok et al., 2008) and speech processing (e.g., Cheung et al., 2009; Goswami et al., 2011; Wang et al., 2012) tasks.

Difficulties in dyslexia extend to processing levels where PA may not be directly involved such as in the perception of variations in pitch and rhythm. As previously discussed in this dissertation, F0 serves as an important indicator of prosodic patterns in speech.

However, there is limited research focusing on the sensitivity to F0 modulations in dyslexia. The following section provides an overview of the main findings from studies that have explored the perception of intonational patterns in individuals with dyslexia.

### **3.3. Deficits in rise time and F0 perception in dyslexic adults**

The phonological deficit observed in individuals with dyslexia has been extensively attributed to deficiencies in basic auditory processing. In recent years, there has been a growing body of research focusing on the sensitivity to speech prosody and rhythm (Goswami et al., 2010; Goswami et al., 2013; Holliman et al., 2010, 2012; Leong et al., 2011; Leong & Goswami, 2014a; Mundy & Carroll, 2012). These studies have consistently demonstrated that individuals with dyslexia, both children and adults, exhibit reduced sensitivity to rhythm and prosody in speech, specifically in relation to syllable stress patterns. Even in the domain of speech production, dyslexia has been associated with deficits in stress processing. For instance, students between 10 and 16 years old were asked to repeat patterns of stressed and unstressed syllables, such as "PA pa pa" or "pa pa PA pa". Compared to age-matched controls, participants with dyslexia showed a higher frequency of errors in assigning stress and exhibited significantly lower amplitude increments for stressed syllables compared to unstressed syllables (Wolff, 2002).

Difficulties with rhythm and prosody in speech have been associated with a deficient detection of frequency variations that correspond to the syllable level (i.e. 2–10 Hz, e.g., Talcott et al., 2000; Witton et al., 2020). These difficulties can arise from localising syllable boundaries or processing temporal information within spoken language inaccurately. As mentioned in Leong & Goswami (2014a), such disruptions in syllable timing may have a “cascading effect on spoken word representations” during language development in individuals, which would lead to “atypical neural specification of phonology” at all levels of phonological processing.

In the last 20 years, research focused on the processing of temporal information has shown that the perception of the amplitude envelope onset (i.e. the change in intensity of the signal or rise time) is consistently found to be impaired for individuals with dyslexia (e.g., see Goswami et al., 2002, Pasquini et al., 2007 for English; Goswami et al., 2011 for a comparison between English, Spanish, and Chinese; Hämäläinen et al., 2009 for Finnish; Lorenzi et al., 2000, Muneaux et al., 2004 for French; see Hämäläinen et al., 2012, and Goswami & Leong, 2013, for comprehensive reviews). Particularly, a study conducted

by Leong et al. (2011) explored the relationship between the processing of rise time and the perception of syllable stress in English speakers with dyslexia. The authors assessed the rise time discrimination of adult participants in a tone task that involved three tones with varying rising times. Through a same-different judgment task, the authors tested the perception of words that had either the same or different syllabic stress patterns. For example, *Difficulty* was presented after either *Difficulty* (same stress pattern) or *diFFiculty* (different stress pattern), and were displayed as identical item combinations (*difficulty–difficulty*) or different item combinations (*difficulty–voluntary*). Results showed that dyslexic adults exhibited poorer rise time discrimination compared to the control group. Moreover, dyslexics had poorer discrimination of syllable stress, regardless of item combinations. Interestingly, individual differences in rise time discrimination predicted performance in stress discrimination. These results suggest that dyslexics exhibit a reduced sensitivity to the rhythmic structure of speech. The authors argued that this may have an impact on the development of the “high-quality phonological representations of spoken words necessary for the acquisition of literacy” (Leong et al., 2011, p. 12). Therefore, individuals with a reduced sensitivity to prosodic patterns may struggle with a prosodic-based speech segmentation strategy.

Rhythmic differences in the material used in this dissertation are captured by F0 variations that differ between homophonic pairs such *l’amie* vs. *la mie* (both /lami/) (see Chapter 5 for an extensive analysis of the acoustic features of the items). It is thus unclear whether differences between these homophonic pairs would be correctly detected by individuals with dyslexia since listeners would have to rely on the acoustic properties of the items, which are located in the onset segment of the phrases.

To our knowledge, no study has directly focused on the exploitation of F0 in dyslexia. The closest prosodic feature that has been explored are speech contour amplitude modulations (AMs, i.e. peaks and valleys present in the speech signal), which have been assessed using artificial sinusoids or speech-like material. For example, to bring stimuli closer to real speech, being transferable to its natural characteristics, Leong & Goswami (2014b) used processed tone-vocoded sentences to investigate difficulties in processing amplitude variations in speech using different frequency bands (stress-only AM, syllable-only AM, and distinct Hz-rate AMs). The authors showed that dyslexics perform similarly to controls when exposed to AMs with unique rates. However, when exposed to combined

stimuli, the performance of dyslexic participants was poorer, suggesting a processing difficulty when auditory stimuli are complex.

As mentioned by Goswami et al. (2002), the phonological deficits in dyslexia may go beyond phonological awareness and phonological representation of words, and underly processes of extraction of suprasegmental information. However, it is still unclear to what extent the phonological processing deficits in adults with dyslexia affect the processing of certain properties of the speech signal.

The specific nature of the phonological deficit in dyslexia is still under debate (e.g., Ramus, 2014b; Saksida et al., 2016; Szenkovits et al., 2016). In the previous sections, we have described the main findings that have given rise to hypotheses about the specific difficulties of phonological deficits, as well as possible causes. In the following section, we will create the link between this evidence and the main state-of-the-art theories.

### **3.4. Auditory processing theories of developmental dyslexia**

#### **3.4.1. Anchoring and categorical deficits.**

Dyslexia is nowadays understood as a complex disorder influenced by multiple factors and its behavioural symptoms cannot be attributed to a single cognitive deficit. However, one of the difficulties that has been observed in numerous studies is the *categorical perception*. Dyslexics have been found to struggle with categorising syllables such as /ba/ and /da/ along a continuum (Reed, 1989), particularly when the syllables differ in the consonants' place of articulation (/ba/-/da/) or voicing (/ba/-/pa/) (Manis et al., 1997). This observation gave rise to the hypothesis that dyslexics exhibit a deficit in the "categorical perception" of phonemes. Categorical perception refers to the constrained ability to discriminate speech sounds based on phonemic labels, where only acoustic differences between phonemic categories can be discerned, while within-category differences cannot be perceived (Adlard & Hazan, 1998; Bogliotti et al., 2002; Liberman et al., 1957; Schulte-Körne et al., 2001). However, recent research has revealed that the enhanced ability of individuals with dyslexia to discriminate acoustic differences within phoneme categories is attributed to their greater sensitivity to allophonic features. These findings indicate that individuals with dyslexia exhibit an "altered" perception of allophonic features, leading to enhanced discrimination abilities within phoneme categories (Serniclaes et al., 2004; see also Serniclaes & Sprenger-Charolles, 2015, for a review).



The debate about whether these deficits in dyslexia are specific to language processing or extend to non-verbal processing gave rise to the *anchoring deficit hypothesis* (Ahissar, 2007; Ahissar et al., 2006). This hypothesis assumes that individuals with dyslexia face difficulties in the automatic extraction of regular patterns from auditory or visual stimuli, which challenge an effective store or "anchor" of perceptual information. Anchoring occurs through repeated exposure to a reference token (the *anchor*). (Ahissar, 2007; Ahissar et al., 2006; Lieder et al., 2019; Oganian & Ahissar, 2012). See also Shulver & Badcock (2021) for a review and meta-analysis of the anchoring deficit in dyslexia.

The theoretical implications were first supported by a study by Ahissar et al., (2006), which employed a two-tone frequency discrimination task. The inclusion of a tone as a reference condition (that is, the first tone remained consistent across trials) benefited typical readers but not those with dyslexia. Furthermore, the performance of the dyslexic group in the reference condition demonstrated a strong correlation with their phonological working memory, indicating a shared impairment between these two processes (Ahissar et al., 2006). Building upon this finding and its implications for reading, the authors proposed that repeated exposure to specific combinations of letters or words and their corresponding sounds plays a pivotal role in the development of reading expertise.

### **3.4.2. Phonology-based theories.**

For many years, research suggested that the underlying cause of the phonological deficits in dyslexia might be either persistent difficulties in translating phonetic features into stable phonological representations (Degraded Phonological Representations Hypothesis, Goswami, 2000; Leong et al., 2011; Ziegler & Goswami, 2005) or laborious access and retrieval of phonological representations (Deficient Phonological Access Hypothesis, Ramus & Szenkovits, 2008).

The Degraded Phonological Representations Hypothesis posits that throughout the process of language development, individuals with dyslexia encounter difficulties in establishing representations of phonological units that are sufficiently robust and distinct to effectively recognise and produce words. The specific characteristics of the presumed degradation of representations vary among authors and encompass elements such as increased noise, phonetic under-specification, reduced temporal or spectral resolution, and/or less categorical representations (Adlard & Hazan, 1998; Elbro, 1998; Snowling,

2000; Tallal, 1980). However, this theory has been challenged in several studies (Boets et al., 2013; Mengisidou & Marshall, 2019; Ramus, 2014a; Ramus & Ahissar, 2012).

On the other hand, the Deficient Phonological Access Hypothesis assumes that tasks involving verbal short-term memory, phonological awareness, and rapid naming are problematic for dyslexic individuals and result in very low performances compared to typical readers. Ramus & Szenkovits (2008) argued that while the phonological representations of dyslexics are intact, the deficit in the mentioned tasks might stem from difficulties at later stages, during phonological access. This hypothesis has been supported by several empirical studies (Boets et al., 2013; Mengisidou & Marshall, 2019; Ramus et al., 2013; Soroli et al., 2010; Szenkovits et al., 2016). It assumes that certain cognitive skills involved in these tasks may be deficient, such as short-term and working memory or speeded access (Ramus et al., 2013).

### **3.4.3. Temporal sampling theory and AST.**

An alternative sensory/neural approach is the Temporal sampling theory (Goswami, 2011), which emerged from empirical data revealing the difficulties of dyslexics in perceiving amplitude envelope rise times across languages (Goswami, 2018; Goswami et al., 2002, 2011, 2014; Goswami & Leong, 2013; Hämäläinen et al., 2012; Huss et al., 2011). This has led to a theoretical framework suggesting that deficits in dyslexia occur during the discrimination of amplitude envelope rise times at slower temporal rates, affecting the detection of rhythm and prosody in speech (Goswami, 2011).

The theory operates on the premise that the phonological processing of speech relies on the "sampling" of the speech stream by neural oscillations at varying time scales or frequencies (Giraud & Poeppel, 2012). It proposes that rise times serve as auditory "edges" that synchronise ongoing neuronal oscillations with AM peaks. Deviations in oscillatory alignment would consequently impact the perceptual organization of AM, resulting in a poor encoding of stressed syllables, syllables, and divisions between onsets and rimes. The neural encoding of speech is provided in an auditory cortical oscillatory hierarchy (see Table 1 for an illustration).

**Table 1.** Illustration of the levels of phonology included in the auditory cortical oscillatory hierarchy (extracted from Goswami, 2015).

<b>Phonological level</b>	<b>Oscillatory frequency (EEG band)</b>	<b>Example(s)</b>	<b>Age at which reflective awareness develops</b>
Intonational phrase	~1Hz and lower	Who's a pretty boy then?	Not yet ascertained
Stressed syllable	~2Hz (delta)	PE-ter PI-per PICKED a PECK of PICK-led PEPP-ers	Not yet ascertained
Syllable	~5Hz ( $\theta$ )	<ul style="list-style-type: none"> <li>• an-i-mal</li> <li>• wig-wam</li> </ul>	2–3 years
Onset-rime	~cued by rising $\theta$ -slope	<ul style="list-style-type: none"> <li>• c-at</li> <li>• str-eam</li> <li>• cl-amp</li> </ul>	3–4 years
Phoneme	~35Hz ( $\gamma$ )	c-l-a-m-p	With alphabetic tuition

On this basis, the Temporal sampling theory proposes that the phonological impairments observed in dyslexia may be partially attributed to the atypical "temporal sampling" of the speech signal by neural oscillations. Specifically, an atypical temporal sampling was observed at slower rates associated with syllable- and stress-related frequencies (i.e. theta and delta frequency bands, respectively) which are below 10 Hz (Goswami, 2011). This proposal has received support from behavioural studies demonstrating diminished sensitivity to syllable stress patterns in dyslexia (Goswami et al., 2010, 2013; Holliman et al., 2010, 2012; Kitzen, 2001; Leong et al., 2011). These findings suggest that difficulties in the perception of AM would have implications for phonological skills across different languages.

The deficient perception of AM is not confined to speech. These findings have been extended to non-speech stimuli (Hämäläinen et al., 2012) and music (Goswami, Huss, et al., 2013; Huss et al., 2011). For instance, adults with dyslexia exhibited significantly reduced phase locking within the delta range (stress-related range, 2 Hz) when exposed to white noise AM (Hämäläinen et al., 2012). This was interpreted as indicative of perceptual deficits in processing slower AM in individuals with dyslexia compared to controls. Moreover, dyslexic children exhibited significantly poorer abilities than controls

in detecting deviations from musical metrical structures in tone sequences (Goswami, Huss, et al., 2013; Huss et al., 2011), extending the findings and supporting the theory in the domain of music.

The Asymmetric Sampling Theory (AST) is a neural framework for speech perception that emerged parallelly to the Temporal sampling theory. It posits that the brain engages in temporal sampling of the speech signal across “multiple time scales” in order to capture phonological elements of different levels (Giraud & Poeppel, 2012; Poeppel, 2003). According to the AST, temporal sampling is achieved through the intrinsic oscillatory activity in the auditory cortex, which synchronises (via phase-locking) with the spectro-temporal modulation patterns associated with various phonological elements in speech. This oscillatory activity emerges from fluctuations in the local field potential of neural populations and is predominantly observed within specific frequency ranges (Buzsáki & Draguhn, 2004). Particularly, this theory highlights the relevance of neural oscillations in the delta (1–3 Hz), theta (4–7 Hz), and gamma (25–80 Hz) frequency bands, which are implicated in the temporal sampling of prosodic, syllabic, and phonemic speech information, respectively (illustrated in Table 1; Ghitza, 2011; Ghitza & Greenberg, 2009; Giraud & Poeppel, 2012).

Neuroimaging studies observed heightened neural responses in dyslexic individuals at higher frequencies (around 40 Hz), suggesting that auditory sampling may occur at a faster rate compared to typical readers (Lehongre et al., 2011, 2013). For instance, in a study by Lehongre et al. (2011), dyslexic participants did not exhibit the lateralised amplification of acoustic modulations in the low-gamma range observed in control participants. The authors concluded that the left auditory cortex of dyslexics might display reduced responsiveness to modulations occurring at specific frequencies crucial for phonemic analysis (30 Hz). The auditory sampling at this rate is claimed to capture rapid transitions in speech (Rosen, 1992), which is assumed to enhance phonemic parsing by Lehongre et al. (2011).

### **3.5. Summary and conclusion**

A growing body of research is starting to unravel the specific impairments that affect individuals with dyslexia. Several studies using different methodological approaches reveal that dyslexia is associated with deficits in acoustic processing and phonological skills that vary from sound categorisation to rate, rhythm, and rise-time detection.

Nonetheless, some individuals demonstrate the ability to overcome these challenges through compensatory mechanisms.

The diverse nature of these deficits makes it challenging to attribute them solely to PA difficulties. Theoretical frameworks have adopted different approaches to account for the challenges experienced by individuals with dyslexia. Theories based on phonological impairments propose that deficits in the representation, storage, or retrieval of sounds can hinder the acquisition of grapheme-phoneme correspondences, resulting in reading difficulties. However, more recent theoretical frameworks broaden these impairments to include perceptual challenges in signal processing. Both the Temporal sampling theory and the AST incorporate neural mechanisms of auditory signal processing to account for deficiencies of the dyslexic brain in capturing modulations in the signal, which can be found in speech, non-speech, or even music.

*Every day I learn new words, new expressions ... The problem is that the signifier has become severed from the signifieds. The words I learn now don't stand for things in the same unquestioned way they did in my native tongue. 'River' in Polish was a vital sound, energized with the essence of riverhood, of my rivers, of my being immersed in rivers. 'River' in English is cold—a word without an aura.*

*When I see a river now, it is not shaped, assimilated by the word that accommodates it to the psyche—a word that makes a body of water a river rather than an uncontained element. The river before me remains a thing, absolutely other, absolutely unbending to the grasp of my mind.*  
(Eva Hoffman, 1991: 106)

## **CHAPTER 4: Segmenting Non-Native Speech**

### **4.1. The challenge of segmenting non-native speech**

As mentioned at the beginning of the introduction of this dissertation, continuous speech is characterised by a lack of clear boundary markers, unlike written language. Hence, the segmentation process involves the identification of sounds in a continuous and highly variable speech signal. This process requires listeners to employ various perceptual strategies to identify word limits. Yet, the segmentation cues used to locate word boundaries depend on the listeners' L1 (Altenberg, 2005; Ito & Strange, 2009; Quené, 1992; Shatzman & McQueen, 2007; Zora et al., 2015, see Chapter 1 for some examples).

The process of speech segmentation appears to be less efficient and less effortless for non-natives than for natives (Carroll, 2001; Shoemaker, 2010). There are a variety of factors contributing to this phenomenon. Particularly, research has highlighted that non-native speech (L2<sup>4</sup>, hereafter) processing may be impacted by the listeners' L1 and the use of varying cues during speech segmentation. For example, when individuals process an unfamiliar or artificial language, they use the patterns of their L1 and apply them to unfamiliar stimuli in order to achieve segmentation (Cutler et al., 1986; Sanders & Neville, 2000; Vroomen et al., 1998). Consequently, as proposed by Cutler & Norris (1988) and Tyler & Cutler (2009), performance in L2 speech processing may be degraded due to the different weighting of cues in the target L2 compared to the listener's L1.

---

<sup>4</sup> Note that the use of the term “second language” or “L2” will be used in this dissertation to refer to any non-native language learned after the onset of the first or native language (L1). Due to the scope of this dissertation, we did not research further into the different implications that have learning an L3, L4 or Ln.

There exists a large (albeit finite) number of speech sounds that occur in the different languages in the world but only a subset of speech sounds is used phonemically (i.e. to convey different meanings) in any particular language. Research has demonstrated that the ability to discriminate L2 contrasts that do not exist in the listener's L1 decreases after the first year of life, while discrimination of L1 sounds improves (e.g., Kuhl et al., 2006; Werker & Tees, 1984).

L2 sounds that are not contrastive in the listener's L1 are difficult to assimilate and discriminate (Best et al., 2001; Best & Tyler, 2007; Goto, 1971; Harnsberger, 2001; Iverson et al., 2003). Numerous studies have shown the difficulty in discriminating between the English /ɪ/-/i/ contrast for Japanese listeners (Flege et al., 1996; Lively et al., 1993). Another example is the case of the reduced sensitivity of American English speakers to the Polish [ɕ] and [ɟ] fricatives, which are not contrastive in English. These speakers show poorer performance than Polish speakers in discriminating between such sounds due to their reduced perception of the distinctive features of both phonemes (both categorised as /j/ in English) (G. L. McGuire, 2007). These are some of the findings that reveal the impact of L1 on the use of cues, as well as reduced discrimination abilities among L2 listeners, as also discussed by Harnsberger (2001).

Sensitivity to low-level cues has been the focus of several studies in the context of L2 speech processing. Particularly, a sublexical cue that has been well studied in L2 segmentation is prosody (e.g., Dupoux et al., 1997, 2001, 2010 for the use of stress in French and Spanish; Kim & Tremblay, 2021 for Gyeongsang Korean and Seoul Korean processing stress in English; Tremblay, Kim, et al., 2021 for the use of prosodic patterns of English and French learners of Korean; Peperkamp et al., 2010 for the use of stress in Standard and South-eastern French, Finnish, Hungarian and Polish). For instance, studies by Bahler et al. (2011) and Coughlin & Tremblay (2011) reveal that English learners of French are not able to use the L2 prosodic patterns to correctly identify targets. The authors measured the sensitivity to F0 rises during the identification of words in French. Learners identified fewer words such as *chalet* when the F0 rise was word-final (the indicator of word boundaries in French) compared to when the F0 rise was word-initial (the word boundary indicator in English). These findings reveal the predominance of L1 prosodic patterns and their influence on L2 speech processing.

As mentioned previously in this dissertation, F0 encompasses variations that are relevant for speech segmentation and word recognition across many languages. As a consequence,

sensitivity to new F0 patterns can be challenging for learners (So & Best, 2010, 2014), as F0 cues can have different informativeness and functional load in the relative languages (L1 and L2, Holt & Lotto, 2006). For instance, while F0 variations in Mandarin Chinese carry lexical information in the form of four discrete tones (Qin et al., 2017; Wiener & Goss, 2019), F0 in English is an important cue to locate stress but it is highly variable and is only informative of less than 1% of words (e.g., Shibata & Shibata, 1990).

A growing body of research has been focusing on the impact of L1 routines in L2 processing. Some evidence indicates that languages with similar acoustic-phonetic patterns will facilitate the transfer of segmentation strategies to apply those of the L1 in L2 processing (e.g., Kim & Tremblay, 2022; Kim et al., 2008; Murty et al., 2007). Under these circumstances, L1 cues seem to benefit L2 speech segmentation. However, if intonational patterns differ between languages, the use of prosodic cues will be affected by possible interferences between them. In this case, L2 perception can be limited by L1 segmentation routines, such as phonotactics (e.g., Weber, 2001) and prosody (e.g., Cutler et al., 1989; Dupoux et al., 1997). A series of studies by Dupoux and colleagues (1997, 2001, 2008) reveal that L1 prosodic patterns may interfere with and impede the correct processing of L2 speech (see section 4.4.2 for a more detailed explanation of the studies). However, other studies reveal that in the absence of the primary L1 cue to segmentation, listeners are able to resort to secondary cues to segment L2 speech (Ivanova et al., 2023; Rojczyk, 2013). In the study conducted by Rojczyk (2013), synthetic variations of the word *record* were presented to Polish advanced learners of English, with manipulations of duration and vowel quality while F0 was held constant. Duration and vowel quality are cues exploited in English, but F0 is the main cue used during speech segmentation in Polish. When Polish participants were asked to identify the word *record* as either the noun form (RE.cord) or the verb form (re.CORD), they were able to rely on both duration and vowel quality differences to identify the targets. Specifically, longer final vowels were identified as the stressed syllables, which was strengthened when the vowel in the initial syllable had a lower F1 and higher F2 compared to the other items (e.g., /rɪkɔ:d/ compared to /rəkɔ:d/ and /rekɔ:d/). Despite the differing outcomes, it seems important to take into account the respective hierarchies of non-lexical cues in L1 and L2 to achieve adequate segmentation in an L2 (Chrabaszcz & Gor, 2014; Mattys & Bortfeld, 2017; Zeng et al., 2022; Zhang et al., 2023; Zhang & Francis, 2010).



Becoming a proficient L2 listener depends not only on the control of segmentation strategies (e.g., Altenberg, 2005; Dupoux et al., 2001; Peperkamp et al., 2010; Tremblay et al., 2012; Tremblay & Spinelli, 2014) but also on the availability of lexical representations in the L2 (e.g., Ivanova et al., 2023; Mora, 2005; Sanders & Neville, 2003; Tremblay, 2011). In order to compare the mechanisms by which words compete and are chosen in L1 and L2, in the next section we will delve into the process of word recognition in the context of L2 processing.

#### **4.2. Word recognition in an L2**

When speech segmentation takes place, hearing meaningful units and recognising words results from the activation of adequate lexical representation. The process of recognising utterances, whether it is the L1 or an L2, seems to be based on the same process: the activation of multiple candidate words (Marslen-Wilson & Welsh, 1978; Zwitserlood, 1989) and competition between them (McQueen et al., 1994), which will lead to the recognition of the target. Therefore, the architecture of the spoken-word recognition system is fundamentally not language-dependent.

Due to the similarity between the listener's task of recognising speech in their L1 and L2, certain skills that are already developed for listening to the L1 can be exploited when processing an L2. However, the availability of these skills may not always be advantageous for recognising words in an L2. As mentioned in Lecumberri et al. (2010), the availability of candidate words and the inhibition of potential candidates in speech recognition are heavily influenced by the crucial role of phonemic distinctions. Accurate phonemic perception is essential for this process. However, when there are differences in the phoneme repertoires between the L1 and the L2, phonemic perception tends to be notoriously inaccurate. For instance, Broersma & Cutler (2011) conducted a series of experiments in which Dutch learners of English exhibited enhanced activation of lexical candidates for near-words such as *lemp* or *daf*. In contrast, this enhanced activation was not observed in native English speakers. These near-words had a vocalic phonemic realisation of [ɛ] instead of the [æ] present in the real words *lamp* or *deaf*. This vowel contrast has been reported difficult for Dutch speakers, regardless of their proficiency level in English. These findings provide evidence that, regardless of their level of proficiency in the L2, Dutch learners face challenges in achieving a perception level of certain phonemic contrasts in English equivalent to that of native speakers, especially when specific sounds remain confusable for them. The authors concluded that difficulties

in effective speech perception may lead to input misinterpretations that may result in the activation of erroneous word candidates.

In addition to these difficulties in speech perception, the L2 vocabulary may be limited compared to that of L1 speakers (e.g., Cooper et al., 2002; Mora, 2005; Sanders & Neville; 2003), which presents an additional layer of complexity. Moreover, for L2 listeners, their set of competitors may include words from their L1. As a result of these factors, the correct candidate word may not always be available, or the competition among candidates may be skewed. The influence of the L1 vocabulary is particularly notable. Even in the case of advanced L2 learners, several studies have shown that L1 vocabulary is activated alongside the L2 vocabulary during L2 processing. For instance, several eye-tracking studies have shown that when listeners are instructed to fixate on an object in their L2 (e.g., *desk* in English), they tend to initially direct their gaze towards an object whose name in their L1 is easily confused with the name of the target (e.g., look at a *lid*, which is called “deksel” in Dutch) (Cutler et al., 2006; Spivey & Marian, 1999; Weber & Cutler, 2004). This evidence underscores the heightened complexity of lexical competition for L2 speakers compared to L1 speakers.

Evidence from this and previous sections reflects the importance of the perception of relevant (and adequate) sounds in an L2 in order to achieve successful word recognition and speech comprehension. In the forthcoming section, we will explain the main assumptions of several influential psycholinguistic models regarding the perception and processing of L2 speech sounds across different languages. In particular, the role of low-level cues during L2 speech segmentation is accounted for in the psycholinguistic models in subsections 4.3.3 and 4.3.4.

### **4.3. Psycholinguistic models of L2 speech perception**

#### **4.3.1. PAM.**

The perceptual assimilation model (PAM, Best, 1993, 1994, 1995; PAM-L2, Best & Tyler, 2007) proposes that the articulatory properties of the L2 sounds determine the degree to which they will be assimilated into the phonetic categories of the L1 system. This, in turn, influences the learner's ability to perceive contrasts involving that specific sound in the L2.

According to the PAM model, L2 phonemes are classified as instances of pre-existing L1 categories. This means that L2 learners map the L2 sounds onto the nearest phonetic categories already present in their L1 phonological system. This is based on the idea that during the acquisition of L1 in infancy, listeners develop distinct categories for the sounds they encounter and produce. Hence, when encountering L2 sounds, listeners would find it easier to handle them when they can be assimilated into the established categories of L1 sounds.

Even though the model proposes that L2 sounds are perceived as distinct and can be reliably discriminated from the phonetic categories in the L1, this assumption poses a challenge for the acquisition of certain L2 categories that are distal from L1 sounds. This difficulty is supported by research indicating poor sensitivity to L2 contrasts that do not exist in the listeners' L1. For instance, Aoyama et al. (2004) examined the Japanese speaker's perception and production of English phonemes, /l/-/ɪ/, /r/-/w/, /s/-/θ/, and /b/-/v/, which are considered challenging for Japanese speakers. The participants' ability to differentiate and produce these contrasts was evaluated in two tests with a 1.1-year gap between the tests. Findings revealed a significant increase in the production ability of the English /w/ phoneme, which was attributed to the existence of an equivalent sound in Japanese, which may have eased the assimilation of the L2 sound. On the other hand, results from Aoyama et al. (2004) also showed greater improvement in identifying and producing the /ɪ/ sound compared to /l/. These results provided evidence in favour of the notion that a greater dissimilarity between an L2 phoneme and the nearest corresponding phoneme in the L1 enhances the learnability of the target language sound (which was proposed in the SLM model by Flege (1995), which is presented below).

#### **4.3.2. SLM.**

The speech learning model (SLM, Flege, 1995) has been highly influential in understanding L2 speech perception and production. This model proposes that L2 learners perceive L2 sounds based on the phonetic categories established in their L1 repertoire. Like PAM, SLM posits that L2 learners can establish new phonetic categories for L2 sounds that differ from those in their L1. Supporting this idea are studies on the learners' production of L2 sounds (Aldamen & Al-Deaibes, 2023; Du & Chen, 2023; Flege, 1980). For instance, in a study conducted by Flege (1980), Saudi Arabic speakers learning English produced L2 voiced and voiceless stops with similarities to Arabic phonetic dimensions, while also approximating some English rules. These findings provide

evidence that L2 learners have the ability to produce the necessary L2 sounds and can establish new phonetic categories.

### 4.3.3. DMAP.

PAM and SLM models have focused on the relevance of L2 sound categories and their impact on perception and production. However, in the introduction of this dissertation, we have highlighted the importance of sublexical cues during speech processing (in L1 as well as in L2). Darcy et al. (2012) present findings that contradict the claims from the SLM model, showing that certain learners who could establish lexical distinctions in the L2 struggled to accurately discriminate them. The key empirical observation reported in their study involves English learners of French who can differentiate between lexical items relying on a /y/–/u/<sup>5</sup> contrast but exhibit poor performance when discriminating [y] from [u] in an ABX task.

The authors proposed the model known as DMAP (Direct Mapping of Acoustics to Phonology), which focuses on the direct association between acoustic properties and phonological representations. According to the DMAP model, the perception of speech sounds involves a direct connection between the acoustic information of the sound signal and the phonological representations in the mental lexicon. The model assumes that the detection of acoustic properties can prompt phonological restructuring based on principles of economy in phonological inventories, resulting in a lexical contrast.

As learners become more proficient in the L2, their perception and categorisation of L2 sounds gradually align with the target language's phonological system. However, the phonetic categories may not yet resemble the L2 forms. The learners are assumed to rely on their existing interlanguage (L1-L2) hierarchy of features to establish contrasting lexical representations. This hierarchy of features is based on the idea that phonological distinctions rely on discrete features organised hierarchically, which has been established by several researchers (e.g., Clements, 2001; Halle, 1992; McCarthy, 1988). The hierarchical structure “minimises redundancy, expresses universal tendencies, and serves as the basis for phonological processes” (Darcy et al., 2012, p. 7). In L2 acquisition, a feature is selected if it is relevant to the L2 processing, such as lexical contrasts (Clements,

---

<sup>5</sup> The L1 English lacks the /y/ sound in their repertoire but not the /u/ sound.

2001). This allows for the abstraction of irrelevant phonetic details in the process of phonological acquisition in L2 (note that this also applies to the acquisition of an L1).

The fact that English learners of French in the study by Darcy et al. (2012) were able to distinguish items that included the contrast /y/-/u/ but were not able to discriminate between the items, led to the conclusion that learners who are able to correctly categorise L2 contrasts, may inaccurately discriminate them. This aligns with a previous study conducted by Sebastian Galles & Baus (2005), who found that performance is at a native-like level in the phonetic task (categorisation) compared to tasks that involve lexical processing (lexical decision and word identification tasks). It was suggested that even learners with accurate perception experience difficulties with L2 lexical encoding. According to DMAP, these findings would be explained by the fact that learners are capable of identifying acoustic cues that correspond to phonological features in the L2 speech, and the extraction of relevant features would be explained by the fundamental language component that enables the acquisition of language.

#### **4.3.4. The Cue-weighting Theory.**

A common aspect of most proposals on how L2 phonological representations are processed by learners is the influence of the listener's L1 while segmenting an L2. For instance, the difficulty in the perception of L2 sounds by Japanese learners of English has been also explained by a theoretical approach known as the cue-weighting theory, which is based on previous evidence showing the relative weight of cues in the listener's L1 (Francis et al., 2000, 2008; Francis & Nusbaum, 2002; Holt & Lotto, 2006; Ingvalson et al., 2011; Iverson et al., 2003).

The cue-weighting theory of speech perception posits that speech perception involves the simultaneous use of multiple acoustic cues. These cues are weighted based on their informativeness in signalling sound contrasts within the L1. As a result, the perception of sound contrasts in the L2 is influenced by the listener's transfer of cue weightings from their L1 to the L2. This theory highlights the multidimensional nature of speech perception and emphasises the role of different cue weightings in shaping the perception of L2 sound contrasts.

In essence, this theory suggests that the stronger the functional importance of a specific cue in the L1, the more likely L2 listeners are to establish an association between the same cue and a distinct function in the L2. Tremblay et al. (2017) pointed out that,

according to the cue-weighting theory, it is expected that when exposed to L2 speech, listeners would prioritise cues that are significant for distinguishing lexical identity in their L1, even if the cues are used differently in their L1 and L2. In an eye-tracking study using a visual-world paradigm, Tremblay et al. (2016) explored the impact of L1 segmentation strategies during L2 segmentation for Korean and English learners of French. Korean and French prosodic patterns are very similar, having a H(igh) tone on the AP-final syllable, and cueing word-final boundaries in the AP-final position. However, F0 cues to word-final boundaries peak slightly later in French than in Korean; particularly, the phrase-final F0 rises in Korean peaks before the syllable offset and starts decreasing thereafter, whereas in French the F0 begins decreasing after the accented syllable. In their study, the authors aimed to determine whether the assimilation of the L2 pattern would make it more challenging for Korean listeners to exploit F0 cues for word-final boundaries in French. Results revealed that, unlike English listeners, Korean listeners were unable to effectively use F0 cues for word-final boundaries in French. This advantage of the L1 segmentation routines was also shown in a subsequent eye-tracking study by Tremblay et al. (2017), showing that the greater functional weight of F0 cues for lexical identity in Dutch, compared to English, gave Dutch listeners an advantage over English listeners in acquiring the use of F0 cues for word-final boundaries in French.

Hence, findings from Tremblay et al. (2016, 2017) extend the weighting-cue theory proposing that the functional weight of a prosodic cue in the L1 can also predict its exploitation in L2 speech segmentation, specifically in learning the association between a given prosodic cue and word-edge boundaries. The authors argued for a transfer of strategies from L1 to L2 during segmentation, even if the strategies served separate functions, such as stress vs. word-boundary signalling.

As demonstrated in this chapter, L2 speech segmentation largely depends on the L1 features and segmentation routines. In the following section, we will particularly focus on the use of low-level cues in English and Spanish, which have been well-documented to display differing speech segmentation strategies from French and are the focus of Experiment 4 (Chapter 8).

#### **4.4. The use of acoustic cues during L2 processing for English and Spanish speakers**

The reliance on acoustic cues during speech segmentation has been proven different for these different groups of speakers. The most notable difference is that, while stress is non-contrastive in French, it is contrastive (i.e. it carries lexical information) in both English and Spanish, and can also indicate differences in grammatical functions (e.g., in English *CON.tract* (noun) vs. *con.TRACT* (verb), example extracted from Fry, 1958); in Spanish *BE.be* “(s/he) drinks” vs. *be.BE* “baby”, example extracted from Sebastian Galles et al., 1992). In the subsequent section, we introduce the main findings showing the use of acoustic cues by English and Spanish speakers.

##### **4.4.1. The exploitation of acoustic cues for L1 English.**

Word stress appears to provide a reliable cue to English speakers to locate boundaries in word-initial positions (Mattys, 2004; McQueen et al., 1994; Vroomen & de Gelder, 1995). As an example, in stress-timed languages, like English, contrastive stress placement is common (such as in “impact”: *IMPact* vs. *impACT*; Fry, 1958). As discussed by Zeng et al. (2022) L1 speakers of English tend to exhibit a trochaic bias in word segmentation, which is linked to stress patterns. This has been shown in studies where disyllabic words were segmented faster and more accurately when the stress was on the initial syllable (Cutler & Norris, 1988a). Moreover, words with initial stress are frequent in English, they occur approximately three times more frequently than words without initial stress (Cutler & Carter, 1987). For example, around 90% of content words begin with a strong stressed syllable, such as *DA.ddy* or *BA.by* (Cutler & Carter, 1987). The reliability of stress on word-initial edges has been demonstrated in studies with infants. For instance, as early as 9 months old, English-learning infants prefer words with a trochaic (strong–weak) syllable pattern over those with an iambic (weak–strong) syllable pattern, aligning with the typical prosodic structure found in English (Jusczyk et al., 1993).

In English, stress is signalled by multiple cues; specifically, F0, intensity, vowel quality, and duration (Altenberg, 2005; Chrabaszcz & Gor, 2014; Fry, 1958, 1965; Haggard et al., 1981; Ito & Strange, 2009; Sluijter & Van Heuven, 1996; Wang, 2008; Zhang & Francis, 2010; Zora et al., 2015). For instance, stress predicts vowel quality, as full vowels can only occur under stress in English (except for some cases which can be unstressed in the final syllable) (e.g., Cutler, 1986; Lunden, 2017).

F0 is shown to be a relevant cue for English listeners. Fry (1958) synthesised noun–verb word pairs such as *subject*, for which F0, intensity, and duration were manipulated. English speakers had to identify whether they heard the noun or the verb form of the word. Results showed that F0 variations were the most robust stress cue, followed by vowel duration. Moreover, F0 serves as an indicator of sentence intonation in English. The F0 is aligned to the stressed syllable within a given word, but it adjusts to the overall sentence intonation, potentially resulting in the alignment of F0 to a different word (Gordon, 2011). Furthermore, sometimes the combination of the aforementioned cues has been observed to indicate stress. A study conducted by Beckman & Edwards (1994) examined recordings of the sequence /'papa/ at different speech rates. They found that stressed syllables were longer and were articulated with greater width and speed.

Once the main findings of the use of cues in English are reviewed, it is important to consider some evidence of the segmentation strategies these listeners exploit during L2 segmentation. Several studies reveal that English speakers use their L1 strategies during the processing of an L2. For instance, in an artificial language learning study by Tyler & Cutler (2009), English speakers were found to rely on duration and F0 to locate word boundaries. These speakers identified higher F0 values as word onsets, while longer durations in final position were a powerful cue to segment speech. Moreover, duration, F0, and vowel quality were used by native English speakers when they discriminated between Mandarin Chinese tones of varying durations and pitch contours for low and high vowels ([a] and [i]) (Chen et al., 2017). In a recent eye-tracking study, English-speaking learners of Korean were also shown to apply their L1 segmentation routines during the identification of target words differing in tonal boundary information (Tremblay, Kim, et al., 2021). In Korean, the canonical AP boundaries are AP-final H tone followed by an AP-initial L tone. However, English learners of Korean were more prone to look at targets when hearing an AP-initial H tone (non-canonical in Korean) compared to the canonical AP-initial L tone. These findings were interpreted as evidence that English speakers rely on their L1's strategy by using an F0 rise as a cue for word-initial boundaries during the processing of L2 Korean. Overall, previous research provides evidence that English speakers rely on their L1 segmentation routines to segment L2 speech, regardless of the specific language involved.



#### 4.4.2. The exploitation of acoustic cues for L1 Spanish.

In Spanish, like in English, stress is marked by a variety of cues. Specifically, a combination of F0, duration, and intensity seems to signal syllable stress to Spanish speakers (e.g., Estebas, 2008). Llisterra et al. (2003) found that an association between higher F0 values with either longer syllable durations or higher intensities guides Spanish listeners during speech segmentation, whereas a single cue does not seem to be a sufficiently robust source. Empirical evidence has been obtained from the sensitivity to stress in studies testing infants, which have shown that Spanish-learning infants (L1) can discriminate between speech segments with differing stress patterns (Pons & Bosch, 2010; Skoruppa et al., 2009).

In contrast to English, Spanish speakers do not rely on vowel quality since no vocalic reductions are found in the language and no differences are observed in vowel characteristics between stressed and unstressed vowels (Navarro, 1966; Sebastian Galles et al., 1992). The stress pattern in Spanish although varies in location, it is a quite stable and predictable prosodic feature, with the stressed syllable occurring at any of the final three syllables of a word (Peperkamp et al., 2010; Sebastian Galles et al., 1992). Specifically, stress on nouns seems to display a rather consistent pattern: it falls on the penultimate syllable in vowel-final nouns (e.g. *MEsa* “table”), while stress falls on the final syllable in consonant-final nouns (e.g. *aniMAL*) (Baković, 2016). Exceptions to these patterns also seem to have a predictable stress pattern. For example, vowel-final nouns with penultimate stress such as *candiDato* (“candidate”) or *paTata* (“potato”), end in a terminal element, generally a vowel (Harris, 1983; see also Baković (2016) for recent work on the stress placement in Spanish).

Spanish is an intriguing language case for investigating the perception of stress and prosodic patterns. This interest stems from the fact that contrary to many languages, stress in Spanish does not align with the boundary edges of words (Soto et al., 2007; Soto-Faraco et al., 2001), and the rising pattern of F0 seems to be misaligned with the accented syllable (Ortega-Llebaria & Prieto, 2011; Prieto et al., 1996).

After examining the main findings regarding the use of prosodic cues during segmentation in Spanish, it becomes crucial to examine the evidence concerning segmentation strategies employed by these speakers during the segmentation of an L2. Multiple studies indicate that Spanish speakers, as English speakers, employ strategies from their L1 when processing an L2. In a series of ABX discrimination tasks using non-

words, Dupoux et al. (1997) demonstrated the sensitivity of Spanish speakers to stress patterns when processing unfamiliar items. Specifically, when non-words exhibited differences at the level of accentuation, such as *boPElo* vs. *bopeLO*, Spanish speakers showed improved performance compared to items that differed in both stress and phonemic levels, such as *BOpelo* – *soPElo* – *boPElo*. Reliance on stress in L2 processing was further supported by a subsequent study using a recall task (Peperkamp et al., 2010). Participants were required to learn two CVCV non-words differing in stress location (*NUmi* vs. *nuMI*), which presented differences in either a combination of duration, F0, and intensity or a combination lacking duration. Spanish speakers made few errors in identifying non-words that differed in stress location, which was enhanced when duration was present as a cue compared to when it was absent.

In summary, similar to English speakers, evidence from these studies reveals that Spanish speakers apply their L1 segmentation strategies during L2 processing, which seems to align with the cue-weighting theory (e.g., Francis et al., 2000; Francis & Nusbaum, 2002; Holt & Lotto, 2006; Tremblay et al., 2017).

We have discussed several studies that examine the sensitivity to stress patterns in an L2, but limited research has focused on the role of specific acoustic-phonetic cues, such as F0, in L2 speech processing during word recognition. In the following section, we will review evidence regarding the use of cues by learners of French when exposed to ambiguous phrases resulting from resyllabification in French.

#### **4.5. Processing French resyllabification as an L2**

French can be a challenging language for L2 speakers since it is characterised by phonological phenomena that can further complicate the task of speech segmentation. As mentioned in Chapter 2, sandhi phenomena such as liaison, *enchaînement*, and elision lead to the mismatch of syllable and word boundaries.

Some studies have addressed the question of the extent to which French learners rely on acoustic cues in liaison contexts. In an eye-tracking study, Tremblay (2011) tested English learners of French of different proficiency levels. Participants were presented with two-word phrases of either real or nonce words. L2 speakers listened to items that either began with a vowel (liaison) or with a consonant /z/ (e.g., *fameux élan* [fa.mø.z#e.lã], meaning "infamous swing," vs. *fameux zélé* [fa.mø.#ze.le], meaning "infamous zealous one"). Participants were asked to identify the target word, which was orthographically displayed.

Results showed that L2 listeners preferred liaison targets over consonant-initial ones when a lexical competitor was also displayed (here, *élan* and *zélé*). This preference was shown for both real and nonce words, regardless of the proficiency level in French.

These findings were further confirmed in a subsequent eye-tracking study. Tremblay and Spinelli (2014) tested the ability to recognise words in liaison environments but included frequent and infrequent pivotal consonants. For instance, the frequent /z/ in *curieux érable* [ky.Rjø.z#e.rabl] meaning “strange maple” vs. *curieux zéro* [ky.Rjø.#ze.ro] meaning “strange zero”, and the infrequent /t/ in *parfait abri* [paʁ.fe.t#a.bʁi] meaning “perfect shelter” vs. *parfait tableau* [paʁ.fe.#ta.blo] meaning “perfect painting”. The task required participants to select the corresponding picture associated with the target word. Results showed that in infrequent liaison contexts, consonant-initial words (*ta.bleau*) were preferred over liaison words (*par.fai.t#a.bri*). However, no differences were observed in frequent liaison contexts. These patterns were consistent for both L1 French speakers and English learners of French. The level of proficiency in French did not interact with the effects.

Findings from both studies are interpreted as evidence of the acquisition of the phonological representation for liaison consonants in English learners of French, particularly since English does not possess resyllabification processes akin to those found in French liaison. Moreover, the authors proposed that the ability to track the occurrence probabilities of liaison and consonant-initial words guided both French natives and English learners to accurately identify the target.

Few studies have focused on the segmentation of French liaison, and to our knowledge, no study has investigated the sensitivity to cues of French learners in elision contexts and their use of segmentation strategies. There are major differences in acoustic properties between liaison and elision (see Chapter 2). Furthermore, while liaison involves pivotal consonants, elision is associated with misalignments to the pivotal vowel /a/. Research has revealed differences in the processing of L2 sounds depending on their nature (i.e. being a vowel or a consonant) (e.g., Aoyama et al., 2004; Darcy et al., 2012; Mora, 2005). For instance, in an AX discrimination task, Mora (2005) showed that Spanish-speaking advanced learners of English performed better when discriminating words based on vowel contrasts (such as [i:]–[ɪ] in *feet–fit*) than in consonant contrasts (such as [s]–[z] in *loose–lose*). Consequently, findings from liaison may not be fully representative of elision.

Differences between consonant and vowel characteristics have not been directly examined in the context of L2 speech processing. Consequently, exploring the extent to which listeners rely on cues associated with liaison and elision contexts may shed light on differences in the effects of their functional roles on language processing in general. In particular, following proposals from Nazzi & Cutler (2019), Nespor et al. (2003), and New et al. (2008), pivotal phonemes involved in liaison are likely to encompass differences that are associated with lexical processing, while elision may involve variations in the processing of structural information. Differences between the functional role of consonants and vowels and their potential implication in resyllabified speech processing will be further discussed at the end of this dissertation, in the General Discussion.

#### **4.6. Summary and conclusion**

Segmenting speech in an L2 is to a certain extent similar to the segmentation process in the L1, as both rely on similar fundamental skills. However, the process of segmentation is less efficient for L2 listeners due to a myriad of factors, including an increased activation and competition of lexical candidates or inefficient sound detection. Consequently, segmenting L2 speech may be more cognitively demanding compared to segmenting L1 speech.

Proposals on the L2 speech segmentation process have mainly considered the mapping and perception of contrastive L2 sounds, such as [l] vs. [r] (PAM, SLM, and DMAP models). Moreover, approaches from DMAP and cue-weighting theory include the listeners' reliance on low-level cues.

On the other hand, recent models of L2 speech perception have also considered the impact of the similarities and differences in segmentation routines between the speakers' L1 and L2, as seen in the cue-weighting theory. This approach explores how individuals learning an L2 exploit non-contrastive features from their L1 to decode speech and recognise words in an L2. This theory accounts for transient ambiguities such as those arising from resyllabification environments in French, which cannot be attributed to phonological or allophonic distinctions but rather rely on acoustic-phonetic information.

After introducing the main research findings about speech segmentation in general and French segmentation in particular, a series of studies were conducted to address some gaps in three different populations: L1 French speakers, French learners, and individuals with dyslexia. There are three main research questions guiding this dissertation:

- 1) To what extent are French listeners attuned to fine-grained acoustic information to segment continuous speech in French when lower- and higher-level cues are available in the speech signal? (Experiments 1 and 2, Chapter 6)
- 2) What is the degree of sensitivity of adults with dyslexia to acoustic cues in the speech signal in French and to what extent are they able to exploit this information? (Experiments 3, Chapter 7)
- 3) Are French learners able to use the acoustic cues present in the signal to segment French? To what extent the L1 segmentation strategies will influence the sensitivity to perceive fine-grained acoustic details in the L2? (Experiments 4, Chapter 8)

To tackle the interests of this dissertation, two sets of the same auditory speech stretches were used across the 4 experiments in the subsequent chapters. Therefore, before delving into these studies, we will provide an overview of the materials used.

# EXPERIMENTAL PART

## CHAPTER 5. Acoustic cues of the material in this dissertation

To address our research questions, we used pairs of phonologically identical determiner phrases (DPs) in elision environments in French. The pairs of homophonic DPs consisted of a definite article in its singular form followed by a noun. Each pair consisted of an elided DP (e.g., *l'amie*) and a DP without elision (e.g., *la mie*). Hence, the noun in one member of each pair was vowel initial and the other member was consonant initial.

DPs were grouped into pairs that differed only by the beginning of the noun (e.g., /ami/ vs /mi/, /afi/ vs /fi/). See (1) and (2) for examples.

- |  |                          |
|--|--------------------------|
| (1) <i>l'amie</i> (“the friend”, elided DP)    | both produced as [lami]  |
| <i>la mie</i> (“the crumb”, non-elided DP)     |                          |
| (2) <i>l'affiche</i> (“the poster”, elided DP) | both produced as [lafɪʃ] |
| <i>la fiche</i> (“the sheet”, non-elided DP)   |                          |

Two different sets of recordings were used in our experiments, each recorded by a different female native speaker. DPs embedded in long sentences were used in Experiments 1 and 2 (Chapter 6). DPs were embedded in the neutral context *C'est* (“It is”) for recordings in Experiments 3 (Chapter 7) and 4 (Chapter 8), although the neutral context *C'est* was excised from the sentences in Experiment 4. Moreover, we generated three additional item conditions for Experiment 3 by manipulating the pivotal phoneme of consonant-initial items.

Differences between pair items were found in duration, formant, and F0 values. All acoustic measurements were made from spectrogram and waveform displays in Praat, using both Praat scripts and visuoacoustic inspection. Statistical analyses were performed using one-way ANOVA tests to compare measures across the items within each experiment, as well as paired samples t-tests to compare between item conditions within each experiment. All statistical analyses of the acoustic measures were performed in R (version 4.0.2).

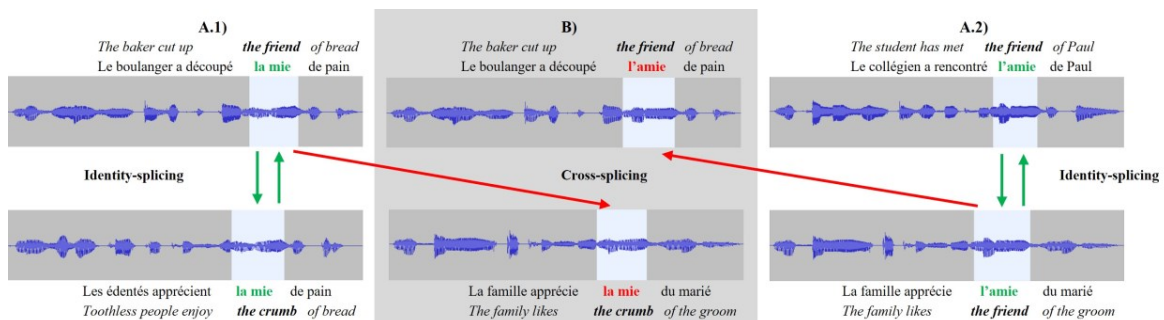
First, we will analyse the acoustic measures of DPs used in Experiments 1 and 2 (Chapter 6), which were embedded in sentences in two different conditions (congruent and incongruent). Then, we will analyse the DPs from Experiments 3 (Chapter 7) and 4 (Chapter 8). Finally, resynthesised manipulations in Experiment 3 will be addressed at the end of this chapter.

## 5.1. DPs in Experiments 1 and 2

### 5.1.1. Materials.

Twenty-four pairs of natural productions of phonemically identical DPs were used, such as /lami/ or /lafif/. The DPs were embedded in sentences with a strong semantic restriction in favour of one of the candidates for each homophone pair. Half of the DPs were placed in a congruent context that favoured their meaning (*congruent condition*) and half in an incongruent context, favouring the meaning of the other homophone (*incongruent condition*), resulting in a total of 96 experimental sentences. As an example, in *Le collégien a rencontré l'amie de Paul* (“The high-school student met Paul's friend”), the meaning of *l'amie* (“the friend”) is selected by the context in which it is presented (congruent condition); however, the meaning of *l'amie* in *Le boulanger a découpé l'amie de pain* (“The baker cut up the bread friend”) is not appropriate in the given context (incongruent condition). Please refer to the Annex I for a complete list of the stimuli.

Two acoustic manipulations were performed for each DP using Audition software (see Figure 3). The first, *cross-splicing*, consists of exchanging the target DPs between sentences. Incongruent sentences were created by cross-splicing each DP into congruent sentences and placing them in the context favouring the other DP. This was done to avoid introducing acoustic differences resulting from the surprise effect caused by words contained in unexpected contexts. The second manipulation, *identity-splicing*, consists of exchanging a target word with another production of the same word, which was applied to sentences in the congruent condition. The aim of this manipulation is to enable the presentation of comparable stimuli, i.e. all of which have undergone the same acoustic manipulations.



**Figure 3.** Illustration of the acoustic manipulations performed to generate our stimuli: A) cross-splicing, B) identity-splicing. Green arrows show the generation of the congruent condition, red arrows show the generation of the incongruent condition. Panel (1) illustrates the two original productions of *la mie*, panel (2) illustrates those of *l'amie*.



Below, we analyse the duration, formant, and F0 values of the DPs in each condition. For the sake of clarity, we will categorise the tokens as either in congruent or incongruent contexts.

### 5.1.2. Measurements and analyses.

#### *Duration*

Durational differences were found between vowel-initial (*l'affiche*) and consonant-initial (*la fiche*) members of the homophonic DPs (see Table 2). In both contexts, mean duration of the first syllable [la] of vowel-initial items was longer ( $M_{\text{CONGRUENT}} = 176$  ms,  $M_{\text{INCONGRUENT}} = 176.3$  ms) than that of consonant-initial items ( $M_{\text{CONGRUENT}} = 167.59$  ms,  $M_{\text{INCONGRUENT}} = 173.5$  ms). However, this effect was not statistically significant ( $ps = .16$ ). The initial phoneme [l] in elided items (e.g., [l] of the definite article in *l'affiche*) was longer than the [l] of non-elided items (like *la fiche*). This was true for both congruent ( $M = 86.5$  ms vs.  $74.6$  ms, respectively) and incongruent ( $M = 86.95$  ms vs.  $81.92$  ms, respectively) contexts. However, this effect was significant only for congruent sentences ( $F(1, 23) = 2.32, p = .03$ ). Reversely, the initial [a] in consonant-initial items was slightly longer ( $M_{\text{CONGRUENT}} = 92.97$  ms,  $M_{\text{INCONGRUENT}} = 91.59$  ms) than that of vowel-initial items ( $M_{\text{CONGRUENT}} = 89.47$  ms,  $M_{\text{INCONGRUENT}} = 89.36$  ms), although these differences were not statistically significant ( $ps > .25$ ).

The second syllable of homophonic DPs (e.g., [fiʃ] in [lafɪʃ] (*la fiche – l'affiche*)) also showed differences in duration. Second syllables in vowel-initial items were longer ( $M_{\text{CONGRUENT}} = 306.16$  ms,  $M_{\text{INCONGRUENT}} = 310.3$  ms) than those in consonant-initial items ( $M_{\text{CONGRUENT}} = 239.7$  ms,  $M_{\text{INCONGRUENT}} = 294.4$  ms). However, this difference was significant only for the incongruent context ( $F(1, 23) = 3.8, p = .0009$ ).

**Table 2.** Mean duration and F0 values (and SDs) for [l] and [a] phonemes in vowel-initial and consonant-initial items.

ITEM	CONGRUENT			INCONGRUENT		
		Duration	F0		Duration	F0
Vowel-initial (e.g., <i>Pamie</i> )	[l]	86.5 (15.74)	165.23 (10.46)	[l]	85 (19)	165.47 (8.79)
	[a]	89.47 (13.3)	184.7 (9.38)	[a]	89.36 (11.36)	184.78 (7.3)
	[la]	176 (20.17)	175 (9.27)	[la]	176.3 (22)	175.1 (7.16)

Consonant- initial (e.g., <i>la mie</i> )	[l]	74.6 (23.5)	173.22 (12)	[l]	81.9 (26.62)	169.35 (14.2)
	[a]	93 (18.38)	160.4 (8)	[a]	91.59 (17.36)	159.9 (8.7)
	[la]	167.59 (35.1)	166.8 (9.3)	[la]	173.5 (34.3)	164.6 (10.17)

Hence, both the initial phoneme [l] and the first syllable [la] in both DP productions exhibit longer durations when they appear in vowel-initial phrases compared to consonant-initial ones, i.e. when the pivotal vowel /a/ is part of the content word (here, the noun). On the other hand, both the pivotal vowel [a] and the following syllable of these items were shorter compared to vowel-initial ones.

### *Formants*

First and second formants (F1 and F2, respectively) were measured at the vowel midpoint. Vowel-initial items had higher values for both F1 ( $M_{\text{CONGRUENT}} = 728.78$  Hz,  $M_{\text{INCONGRUENT}} = 740.06$  Hz) and F2 ( $M_{\text{CONGRUENT}} = 1876.63$  Hz,  $M_{\text{INCONGRUENT}} = 1867.95$  Hz) compared to formants of consonant-initial items (F1:  $M_{\text{CONGRUENT}} = 724.61$  Hz,  $M_{\text{INCONGRUENT}} = 729.57$  Hz; F2:  $M_{\text{CONGRUENT}} = 1851.49$  Hz,  $M_{\text{INCONGRUENT}} = 1846.43$  Hz). However, differences were not significant ( $ps > .45$  and  $p > .39$ , respectively).

### *Fundamental frequency*

Mean F0 values of [l] and [a] in the first syllable for both vowel-initial and consonant-initial items were calculated. Results showed that the F0 of the initial phoneme [l] in non-elided items was higher ( $M_{\text{CONGRUENT}} = 173.2$  Hz,  $M_{\text{INCONGRUENT}} = 169.4$  Hz) compared to the [l] in elided items ( $M_{\text{CONGRUENT}} = 165.2$  Hz,  $M_{\text{INCONGRUENT}} = 165.5$  Hz). However, this difference was significant only for congruent DPs ( $F(1, 23) = -2.39$ ,  $p = .025$ ). Inversely, the F0 of the initial [a] of non-elided items was significantly lower ( $M_{\text{CONGRUENT}} = 160.4$  Hz,  $M_{\text{INCONGRUENT}} = 159.9$  Hz) than the initial [a] of elided items ( $M_{\text{CONGRUENT}} = 184.7$  Hz,  $M_{\text{INCONGRUENT}} = 184.8$  Hz) ( $ps < .0001$ ).

Importantly, intonational features of syllable onsets are influenced by characteristics of the succeeding syllable. Since the second syllable varied across DP pairs, phonotactic phenomena may introduce some signal perturbations in the production of some syllable onsets, such as the voiceless obstruents [t] in *l'atout – la toux* (“the asset/the cough”) and

[p] in *l'appel – la pelle* (“the call/the shovel”). Consequently, to enable adequate comparisons of items, F0 differences between the first vowel (always /a/) and the second vowel midpoints (e.g., /i/ in *l'amie – la mie*) were also quantified, following procedures in (Spinelli et al., 2007). This analysis revealed F0 differences between vowel and consonant initial items, with vowel-initial items having a higher F0 ( $M_{\text{CONGRUENT}} = 189.46$  Hz,  $M_{\text{INCONGRUENT}} = 192.95$  Hz) than that of consonant-initial items ( $M_{\text{CONGRUENT}} = 174.12$  Hz,  $M_{\text{INCONGRUENT}} = 169.58$  Hz). This difference was significant for the incongruent items ( $p = .004$ ), but it was not significant for the congruent ones ( $p > .08$ ).

To account for the articulatory variations that arise from vowel properties, further analyses were performed on F0 values of [a] contrasting across second-vowel types: vowel height (high/mid-high, mid-low/low) and backness (front, back). We found differences for items with both high and low vowels. For items with a high vowel in the second syllable, F0 values were higher in items like *l'affiche* ( $M_{\text{CONGRUENT}} = 180.78$  Hz,  $M_{\text{INCONGRUENT}} = 184.53$  Hz) than in those like *la fiche* ( $M_{\text{CONGRUENT}} = 161.34$  Hz,  $M_{\text{INCONGRUENT}} = 165.66$  Hz) ( $ps < .005$ ). For items with a low vowel in the second syllable, F0 values were also higher in vowel-initial items, like *l'alarme* ( $M_{\text{CONGRUENT}} = 188.14$  Hz,  $M_{\text{INCONGRUENT}} = 186.48$  Hz), than in consonant-initial items, like *la larme* ( $M_{\text{CONGRUENT}} = 159.16$  Hz,  $M_{\text{INCONGRUENT}} = 156.54$  Hz) ( $ps < .0001$ ). Backness of word-medial vowels also revealed differences in F0 values. Vowel-initial items with a frontal vowel, such as *l'amie*, had higher F0 values ( $M_{\text{CONGRUENT}} = 186.92$  Hz,  $M_{\text{INCONGRUENT}} = 186.66$  Hz) compared to consonant-initial items, such as *la mie* ( $M_{\text{CONGRUENT}} = 159.65$  Hz,  $M_{\text{INCONGRUENT}} = 159.44$  Hz) ( $ps < .0001$ ). Similarly, vowel-initial items with back word-medial vowels (e.g., *l'accroche*,  $M_{\text{CONGRUENT}} = 181.02$  Hz,  $M_{\text{INCONGRUENT}} = 182.66$  Hz) had higher F0 values compared to consonant-initial ones (e.g., *la croche*,  $M_{\text{CONGRUENT}} = 160.76$  Hz,  $M_{\text{INCONGRUENT}} = 160.1$  Hz) ( $ps < .03$ ).

Hence, the mean F0 values of the onset phoneme ([l]) in non-elided items were higher than for elided items. On the other hand, mean F0 of the initial syllable [la], the first [a], and the subsequent syllable of items were higher in elided DPs than in non-elided DPs.

## 5.2. DPs in Experiments 3 and 4

### 5.2.1. Materials.

Thirty pairs of natural productions of phonemically identical DPs were used in each of these experiments. As in Experiments 1 and 2, the pairs consisted of a noun preceded by a definite article (e.g., *la mie/l'amie*, 'the crumb/the friend'). However, for these experiments, the DPs were framed in a neutral context (*c'est*, 'it is'). Please refer to the Annex II and III for a complete list of the stimuli.

### 5.2.2. Measurements and analyses.

#### *Duration*

Durational differences were found between vowel-initial (*l'amie*) and consonant-initial (*la mie*) members of the homophonic DPs (see Table 3). Mean duration of the first syllable [la] of vowel-initial items was longer (M = 157.69 ms) than that of consonant-initial items (138.68 ms) ( $F(1, 29) = 4.57, p < .0001$ ). Precisely, the initial phoneme [l] in elided items (e.g., [l] of the definite article in *l'amie* items) was longer (M = 68.6 ms) than the [l] of non-elided items (like *la mie*) (M = 57.36 ms) ( $F(1, 29) = 4.36, p = .0002$ ). Moreover, the duration of the initial [a] in vowel-initial items was greater (M = 89.09 ms) than that one of non-elided (consonant-initial) items (M = 81.32 ms) ( $F(1, 29) = 2.30, p = .029$ ).

Moreover, the second syllable of homophonic DPs (e.g., [fiʃ] in [lafɪʃ] (*la fiche – l'affiche*)) also showed differences in duration. Second syllables in consonant-initial items were longer (M = 271.11 ms) than those in vowel-initial items (M = 243.51 ms). However, this difference was not significant ( $p = .2$ ).

Thus, both the onset phoneme [l] and the vowel [a] were longer in content-initial word positions compared to those in function words. On the other hand, the second syllable of DPs was longer in consonant-initial than in vowel-initial items.

#### *Formants*

F1 and F2 were measured at the vowel midpoint. Vowel-initial items had higher values for both F1 (M = 720.08 Hz) and F2 (M = 1439.64 Hz) compared to formants of consonant-initial items (M = 708.23 Hz and M = 1409.2 Hz, respectively). However, differences were not significant ( $p = .5$  and  $p = .71$ , respectively). It is important to note that the first syllable was kept constant across items (always /la/), however vowels of the

second syllable varied (such as /i/ in *l'amie* – *la mie* or /u/ in *l'atout* – *la toux*). Thus, we performed separate analyses in function of vowel height (2 groups: high/mid-high, mid-low/low) and vowel backness (2 groups: front, back), as in Spinelli et al. (2007).

Higher F1 values are associated with lower vowels and higher F2 values with more frontal vowels. Thus, F1 values were contrasted to vowel height and F2 with vowel backness. Analyses showed that F1 was significantly lower (M = 454.43 Hz) for high-mid word-medial vowels in vowel-initial content words (e.g., *l'affiche*) than for low-mid word-medial vowels in vowel-initial content words (e.g., *l'amarre*) (M = 626.67 Hz) ( $F(1, 27) = -2.66, p = .013$ ). F1 was also significantly lower (M = 454.43 Hz) for words like *l'affiche* (high-mid word-medial vowels) than for low-mid word-medial vowels in consonant-initial words (e.g., *la marre*) (M = 658.2 Hz) ( $F(1, 28) = -3.36, p = .002$ ). F1 values in words like *la fiche* (high-mid word-medial vowels) were significantly lower (M = 464.98 Hz) than those of low-mid word-medial vowels in vowel-initial words (e.g., *l'amarre*) (M = 626.67 Hz) ( $F(1, 27) = -2.55, p = .017$ ). F1 values of DPs like *la fiche* (high-mid word-medial vowels) were also significantly lower (M = 464.98 Hz) than those of low-mid word-medial vowels in consonant-initial words (e.g., *la marre*) (M = 658.2 Hz) ( $F(1, 27) = -3.26, p = .003$ ).

Significant differences were found for F2 values of word-medial vowels. F2 of front word-medial vowels in vowel-initial content words (e.g., *l'amarre*) were significantly higher (M = 1598.24 Hz) than those of back word-medial vowels (e.g., *l'alocution*) (1219.14 Hz) ( $F(1, 16) = 2.23, p = .04$ ). F2 of DPs like *l'amarre* were significantly higher than those of back word-medial vowels in consonant-initial content words (e.g., *la locution*) (M = 1185.49 Hz) ( $F(1, 16) = 2.42, p = .028$ ). F2 values of DPs like *la marre* (front medial vowels) were significantly higher (1795.3 Hz) than those of back word-medial vowels in vowel-initial content words (e.g., *la locution*) ( $F(1, 17) = 3.55, p = .003$ ). F2 was higher for items like *la marre* (front word-medial vowels) than for items like *la locution* (back word-medial vowels) (M = 1185.49 Hz) ( $F(1, 17) = 3.76, p = .002$ ).

#### *Fundamental frequency*

Mean F0 Hz of [l] and [a] in the first syllable for both vowel-initial and consonant-initial items were calculated (see Table 3). F0 of the initial phoneme [l] in elided items was higher (M = 191.39 Hz) compared to the [l] in non-elided items (M = 183.46 Hz). This difference was marginally significant ( $F(1, 29) = 1.97, p = .059$ ). Similarly, the F0 of the

initial [a] of elided items was higher (200.52 Hz) than the initial [a] of non-elided items (M = 181.99 Hz) ( $F(1, 29) = 8.76, p < .0001$ ).

We also considered the impact of the characteristics of the succeeding syllable in the intonational features of syllable onsets. F0 differences between the first vowel (always /a/) and the second vowel midpoints (e.g., /i/ in *l'amie – la mie*) were also quantified, following procedures in (Spinelli et al., 2007). Analyses revealed that the F0 in consonant-initial items was higher (M = 41.75 Hz) than that in vowel-initial items (M = 25.71 Hz) ( $F(1, 29) = -4.27, p < .0002$ ). Moreover, F0 values of second-syllable vowels were higher in vowel-initial items (M = 226.23 Hz) than in consonant-initial items (223.74 Hz). Yet, these differences were not significantly different ( $p = .58$ ).

Further analyses were performed on F0 values of [a] contrasting across second-vowel types: vowel height (high/mid-high, mid-low/low) and backness (front, back). We found differences for items with both high and low vowels. For items with a high vowel in the second syllable, F0 values were higher in items like *l'amie* (M = 204.31 Hz) than in those like *la mie* (M = 184.52 Hz) ( $F(1, 9) = 6.64, p < .0001$ ). For items with a low vowel in the second syllable, F0 values were also higher in items like *l'amarre* (M = 198.63 Hz) than in those like *la toux* (M = 179.44 Hz) ( $F(1, 19) = 7.33, p < .0001$ ). Backness of word-medial vowels also revealed differences in F0 values. Vowel-initial items with a frontal vowel, such as *l'affiche*, had higher F0 values (M = 199.4 Hz) compared to consonant-initial items, such as *la fiche* (M = 181.41 Hz) ( $F(1, 20) = 7.45, p < .0001$ ). Similarly, vowel-initial items with back word-medial vowels (e.g., *l'ajout*, M = 203.14 Hz) had higher F0 values compared to consonant-initial ones (e.g., *la joue*, M = 183.31 Hz) ( $F(1, 8) = 5.78, p = .0004$ ).

**Table 3.** Mean duration and F0 values (and SDs) for [l] and [a] phonemes in vowel-initial and consonant-initial items.

		Duration	F0
Vowel-initial (e.g., <i>l'amie</i> )	[l]	68.6 (10.71)	191.39 (8.7)
	[a]	89.09 (19.36)	200.52 (10.3)
	[la]	157.69 (15.04)	195.89
	[l]	57.36 (14.15)	183.46 (20.78)
	[a]	81.32 (14.55)	181.99 (5.49)

Consonant- initial (e.g., <i>la</i> <i>mie</i> )	[la]	138.68 (14.35)	182.81
--	------	-------------------	--------

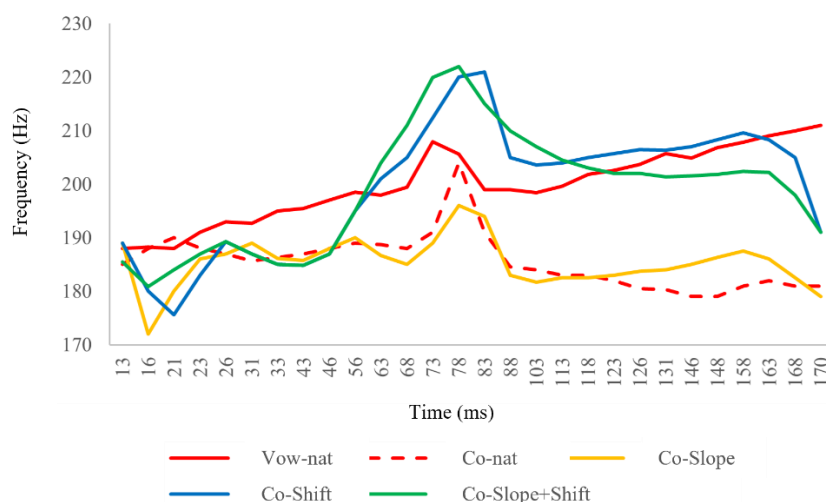
---

### 5.3. Analyses of the F0 trajectory of DPs in Experiment 3

#### 5.3.1. Materials.

For this experiment, we were interested in observing the impact of the F0 trajectory on the listener's perception. For that, we focused on two acoustic properties of the F0: its mean and its slope values. Five conditions were included in the experiment: two natural productions and three resynthesised stimuli generated from the consonant-initial items. The two natural productions were vowel initial (Vow-nat condition) and consonant initial (Co-nat condition) DPs, for which the F0 mean was already analysed in the previous section. In this section, we proceed to the analyses of the F0 trajectories for vowel initial and consonant initial DPs, as well as analyse the F0 mean and slope for the resynthesised stimuli used in Experiment 3 (Chapter 7).

Manipulations on the resynthesised conditions aimed to disentangle the role of the mean and the slope during the perception of ambiguous DPs. For that, both the mean F0 value (Hz) and the F0 slope (Hz/ms) of /a/ were calculated from vowel-initial words (e.g., *l'amie*) to modify the F0 value of the first vowel /a/ in each of the 3 new tokens. These parameters were taken from their paired vowel-initial counterpart and were applied to the F0 contour of Co-nat (e.g., *la mie*) in the time window belonging to /a/ with 20 ms cosine ramps. The F0 curve of the initial vowel /a/ in Co-nat was (1) multiplied by a scaling factor to reach the same F0 mean value as that in the /a/ in Vow-nat (Co-Shift condition); (2) rotated to reach the slope value of the /a/ in Vow-nat while keeping its F0 value (Co-Slope condition); and (3) both shifted and rotated (Co-Slope+Shift condition). The resynthesis process was performed on the STRAIGHT software (McClelland & Elman, 1986). See Figure 4 for an illustration of the parameters of each condition.



**Figure 4.** Illustration of the F0 manipulations of the first two phonemes (/la/) for the pair “*la mie/l’amie*”. Conditions are represented in separate lines: Vow-nat (continuous red), Co-nat (discontinuous red), Co-Slope (yellow), Co-Shift (blue), and Co-Slope+Shift (green).

### 5.3.2. Measurements and analyses.

To account for the dynamicity of the F0 trajectory in every condition, we took 5 acoustic measures of the F0 Hz of the pivotal /a/ phoneme: at its onset, at its midline point, at its offset, and both its maximum and minimum peaks. However, only 2 items (“*joue*” in the Co-Slope+Shift condition and “*version*” in the Co-nat condition) had a minimum peak that was different from the onset, midpoint, or offset values. Therefore, we only kept a total of 4 measurements for the analyses: onset, midpoint, offset, and maximum peak. See all measures in Table 4.

Analyses revealed that differences between natural productions were both at the mean and slope values. The vowel /a/ of the article in the Co-nat condition (e.g., *la mie*) had a consistently lower F0 value (M = 183.77 Hz) than that of Vow-nat items (e.g., *l’amie*) (M = 199.7 Hz). Particularly, the onset F0 mean value in Vow-nat DPs was higher (M = 198.67 Hz) than that in Co-nat DPs (M = 185.99 Hz) ( $t(29) = 5.75, p < .0001$ ). It was also higher than in Co-Slope resynthesised DPs (M = 183.39 Hz) ( $t(29) = 7.65, p < .0001$ ). However, both Co-Shift (M = 202.04 Hz) and Co-Slope+Shift (M = 198.2 Hz) items had mean F0 onsets that were higher than those in Vow-nat items ( $ps < .02$ ). Co-Shift had the highest mean F0 onset on average, being statistically significant compared to all conditions ( $ps < .001$ ). Specifically, the F0 onset of Co-Shift items was +3.84 Hz higher than Co-Slope+Shift items, +5.37 Hz than Vow-nat items, +16.37 Hz than Co-nat items, and +18.66 Hz than Co-Slope items.



F0 rises were also different across item conditions. The F0 maximum peak was the highest for Co-Slope+Shift items (M = 203.09 Hz). This was statistically significant compared to the maximum peak of both Co-Slope (M = 186.85 Hz) and Co-nat (M = 185.99 Hz) items ( $ps < .0001$ ). No significant differences were found between the F0 maximum peaks of Co-Slope+Shift items and those of Co-Shift and Vow-nat ones ( $ps > .26$ ). However, we observed that the F0 maximum peak appeared after the onset and before the midpoint for all item conditions except for Vow-nat items, for which the F0 maximum peak appeared later, between the midpoint point and the offset. One-way ANOVA tests revealed differences in the time points of the F0 maximum peaks. Vow-nat items had later F0 rises (M = 0.147 ms) compared to the other conditions ( $F(4, 149) = 7.92, p < .0001$ ).

Mean F0 values at the midpoint were higher for Co-Slope+Shift DPs (M = 199.39 Hz) compared to other conditions ( $F(4, 149) = 31, p < .0001$ ). This was statistically significant compared to Co-Slope DPs (M = 182.98 Hz,  $t(29) = -7.769, p < .0001$ ) and Co-nat DPs (M = 182.10 Hz,  $t(29) = 8.023, p < .0001$ ).

At the offset, the mean F0 was higher for Vow-nat items (M = 201 Hz) compared to the other conditions ( $ps < .01$ ). Specifically, F0 offsets of the Vow-nat condition were +4.32 Hz higher than those of the Co-Slope+Shift condition, +8 Hz higher than those of Co-Shift items, +17.1 Hz higher than Co-Slope items, and +19.7 Hz higher than Co-nat items.

**Table 4.** Mean time points (in s) and frequency (in Hz) values (and SDs) for the four acoustic measures taken on the F0 (onset, midpoint, maximum peak, and offset) across conditions (Vow-nat, Co-nat, Co-Slope, Co-Shift, Co-Slope+Shift).

CONDITION	M TIME POINT				M Hz			
	Onset	Midpoint	Max. peak	Offset	Onset	Midpoint	Max. peak	Offset
<i>Vow-nat</i>	0 (0)	0.078 (0.12)	0.147 (0.24)	0.156 (0.24)	196.670 (9.58)	198.633 (9.73)	202.496 (9.85)	201.002 (11.12)
<i>Co-nat</i>	0 (0)	0.053 (0.07)	0.009 (0.02)	0.106 (0.15)	185.668 (6.1)	182.103 (5.5)	185.992 (5.97)	181.303 (5.34)
<i>Co-Slope</i>	0 (0)	0.053 (0.04)	0.042 (0.08)	0.106 (0.15)	183.385 (5.78)	182.979 (6.27)	186.846 (6.55)	183.904 (7.29)

<i>Co-Shift</i>	0 (0)	0.053 (0.08)	0.005 (0.02)	0.106 (0.16)	202.042 (10.15)	199.187 (10.22)	202.521 (10.38)	193.040 (10.17)
<i>Co-Slope+Shift</i>	0 (0)	0.052 (0.08)	0.047 (0.03)	0.106 (0.15)	198.203 (9.97)	199.385 (11.18)	203.317 (10.28)	196.687 (11.33)

The F0 slope or trajectory was analysed considering the F0 peaks and valleys (Hz/ms) from onset to maximum peak and from maximum peak to offset. The F0 trajectory with the mean values for items for each condition are illustrated in Figure 5. The vowel /a/ in the Co-nat condition (e.g., *la mie*) had a consistently lower F0 slope (-0.041 Hz/ms) than that of Vow-nat items (e.g., *l'amie*) (F0 slope = +0.028 Hz/ms). In Vow-nat items, we observed a relatively high F0 onset (M = 197.67) that increased by +5.83 Hz over an average of 146.5 ms to reach its maximum peak and then decreased by -1.49 Hz until reaching its offset (see panel A in Table 4 for an illustration of its mean trajectory). The slope of Vow-nat items was +0.04 Hz/ms until their maximum peak and -0.159 Hz/ms until offset.

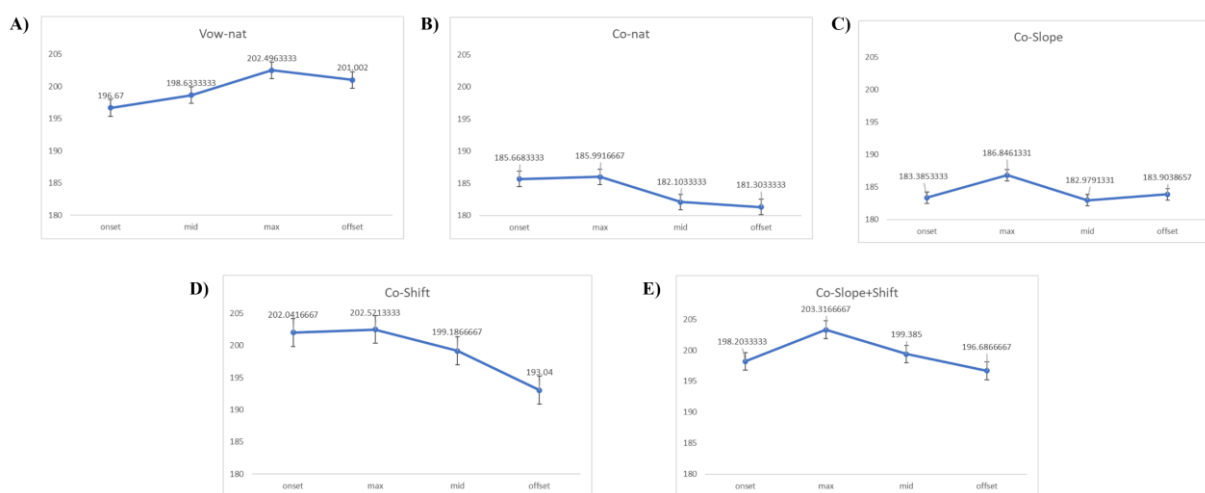
Co-nat items increased their F0 by +0.32 Hz at their maximum peak quite rapidly (during 9.17 ms on average). Subsequently, the F0 decreased by -3.89 Hz until the midpoint and continued to decrease by -0.8 Hz until reaching the offset. The offset value (M = 181.3 Hz) was lower than the onset (M = 185.67 Hz). The slope of Co-nat items was +0.035 Hz/ms until their maximum peak and -0.048 Hz/ms until offset. See panel B in Table 4 for an illustration of its mean trajectory.

The resynthesised conditions differed from both Vow-nat and Co-nat conditions. In the Co-Slope items, there was an initial F0 increase of +3.46 Hz from the onset, which lasted an average of 41.8 ms. Subsequently, it rapidly decreased by -3.87 Hz over 11.39 ms until reaching the midpoint of /a/. Afterward, there was a slight increase (M = +0.93 Hz) until its offset, occurring 53.2 ms later. See panel C in Table 4 for an illustration of its mean trajectory. The slope of Co-Slope items was +0.083 Hz/ms until their maximum peak and -0.046 Hz/ms until offset. See panel B in Table 4 for an illustration of its mean trajectory.

The F0 trajectory of the Co-Shift condition exhibited a decline. Initially, it experienced a moderate but swift increase (M = +0.089 Hz/ms), followed by a gradual decrease of -9.48 Hz until reaching its offset (M = -0.094 Hz/ms). The slope of Co-Shift items was +0.0873

Hz/ms until their maximum peak and -0.094 Hz/ms until offset. See panel D in Table 4 for an illustration of its mean trajectory.

Finally, Co-Slope+Shift items showed a steep F0 rise, increasing +5.13 Hz in 46.91 ms. The F0 then dropped -3.93 Hz until its midpoint (M = 199.39 Hz, -0.71 Hz/ms), which continued to decrease until its offset (M = 196.69 Hz, -0.05 Hz/ms). The slope of Co-Slope-Shift items was +0.109 Hz/ms until their maximum peak and -0.112 Hz/ms until offset. See panel E in Table 4 for an illustration of its mean trajectory.



**Figure 5.** Illustration of the F0 mean trajectory of the five conditions in Experiment 3, including four measures: onset, midpoint, maximal peak, and offset. Panel A shows the mean F0 measures of the natural vowel-initial items (Vow-nat condition); panel B shows the mean F0 measures of the natural consonant-initial items (Co-nat condition); panel C shows the mean F0 measures of the slope-resynthesised items (Co-slope condition); panel D shows the mean F0 measures of the mean-resynthesised items (Co-Shift condition); panel E shows the mean F0 measures of the mean-and-slope-resynthesised items (Co-Slope+Shift condition). Error bars show SDs.

Overall, we observe that the pivotal /a/ phoneme in Co-Slope+Shift items had a steeper slope increase from onset to maximum peak (M = 0.085 Hz/ms) compared to other conditions. However, this difference was only statistically significant compared to Co-Slope (M = 0.049 Hz/ms,  $p = .005$ ) and Co-nat (M = 0.005 Hz/ms,  $p < .001$ ) conditions. In contrast, the F0 in Co-nat items had a less steep trajectory (M = -0.005 Hz/ms) compared to the other conditions. This was significant compared to all conditions ( $ps < .001$ ), except for Co-Shift ( $p = 0.075$ ).

The pivotal /a/ phoneme in Vow-nat items had a steeper slope decrease from maximum peak to offset (M = -0.282 Hz/ms) compared to other conditions, although no significant differences were found ( $ps > .35$ ). In contrast, the F0 in Co-nat items had a less steep

trajectory ( $M = -0.061$  Hz/ms) compared to the other conditions. This was significant compared to two of the resynthesised conditions (Co-Shift:  $p < .001$ ; Co-Slope+Shift:  $p > .005$ ), but it was not significant compared to Co-Slope ( $p = 0.56$ ).

#### 5.4. Summary and conclusion

The analyses performed on our material revealed differences in several acoustic properties of the productions. Specifically, there were variations in duration, formant values, and F0 of the pivotal phoneme [a] and of the subsequent syllable of items. Additionally, we observed differences in duration and F0 values for the initial [l] of DPs.

Across all four studies, we found some consistent patterns. The initial phoneme [l] and the first syllable [la] were longer in vowel-initial phrases. Furthermore, the pivotal [a] and the first syllable [la] exhibited higher mean F0 values in elided DPs compared to non-elided DPs.

However, it is important to note two important differences between the set of stimuli. In consonant-initial items, the pivotal vowel [a] was longer in Experiments 1 and 2 but shorter in Experiments 3 and 4. The onset [l] had a higher F0 value in non-elided items in the sets used for Experiments 1 and 2, while it was higher in elided items in Experiments 3 and 4. Additionally, the duration of second syllables was not consistent between sets. In Experiments 1 and 2, they were longer in vowel-initial items, but in Experiments 3 and 4, they were longer in consonant-initial items.

Analyses conducted on the F0 trajectories in both the natural and resynthesised conditions of Experiment 3 have revealed how the intonational patterns vary across conditions. The steepest increase in F0 from the DP onset was observed in items where both slope and mean F0 were manipulated (Co-Slope+Shift), which was followed by natural vowel-initial productions (Vow-nat). Moreover, Vow-nat showed the steepest decrease in F0 until the offset, followed by Co-Slope+Shift. On the other hand, in items where only the F0 slope was manipulated (Co-Slope) and in natural consonant-initial productions (Co-nat), the F0 trajectories exhibited less pronounced rises and falls. Finally, the F0 of items where only the mean F0 was manipulated (Co-Shift) consistently decreased from onset to offset.

## CHAPTER 6. The interplay of acoustic and semantic cues during the processing of homophonic DPs in French

### 6.1. Introduction

The existing body of research on L1 processing has widely agreed that prosody plays a crucial role in segmentation, as it is one of the first auditory language patterns infants perceive and acquire (Friederici & Wessels, 1993; Koponen, 2001). However, the listener's perceptual system simultaneously integrates information from multiple levels as the stream of speech unfolds. To locate word boundaries, listeners rely on a variety of cues including sentence-level information such as syntactic structure or semantic content (Dahan & Brent, 1999; Sanders & Neville, 2000), suprasegmental features such as allophonic variation or stress patterns (Quené, 1992; Sanders & Neville, 2003a), and/or sublexical information such as phonotactics (Dal Ben et al., 2021; McQueen, 1998; Mehler et al., 1981).

Research conducted by Mattys and colleagues (Mattys, 2004; Mattys et al., 2005, 2007; Mattys & Melhorn, 2007; White et al., 2012) has explored the weighting of multiple segmentation strategies, aiming to develop a hierarchical framework that assigns relative importance to each distinct cue in the process of speech processing. Using a word monitoring task, Mattys et al. (2005) evaluated the processing of sublexical and lexical cues in English listeners. Real words (e.g., *male*) followed either a word (*calculus male*) or a nonword (*baltuluf male*) in intact and truncated contexts. Participants were asked to identify when a specific target word was present or absent in a subsequent utterance. The authors found that under normal listening conditions (intact contexts) and when all cues were optimally available, listeners relied on lexical information to segment speech, with semantic and syntactic information contributing to the segmentation of lexical content. Additionally, the authors show in a series of studies the substantial effect of noisy environments in the detection of and reliance on cues (Mattys, 2004; Mattys et al., 2012; Mattys & Melhorn, 2007; White et al., 2012).

The main findings of these studies point to a predominant role of suprasegmental and segmental cues, such as stress or acoustic-phonetic information, over lexical and contextual cues when the signal is degraded. However, this cue reliance is reversed under intact listening conditions. Overall, when other (higher-level) cues are available in intact listening conditions, the model of the hierarchy of cues proposed by the authors

establishes a rather secondary role of subphonemic cues during segmentation. Yet, the authors found that low-level information plays a relevant role when lexical and contextual information are limited. This was demonstrated in a follow-up experiment using a lexical decision task (Mattys et al., 2005). Words like *creMATE* were embedded in contexts favouring the full word (here, *cremate* and not *mate*), such as in *An alternative to traditional burial is to creMATE the dead*. Results revealed that contextual information outweighed stress. However, when the phonotactic information was manipulated to favour *mate* instead of *cremate*, no effects were found. The authors argued that segmental cues such as phonotactics seem to play a crucial role in facilitating speech segmentation under conditions where lexically driven segmentation is weakened or limited.

The variability across languages in the role of cues during speech segmentation can complicate the understanding of the extent to which listeners rely on certain low-level cues and their interaction with other cues. Consequently, it is essential to evaluate the specific role of the cues involved in speech segmentation in different languages, enabling a comparative assessment of their relative importance. This will provide a theoretical model that accurately captures the interaction of cues across languages, thus contributing to a deeper understanding of linguistic differences. In this study, we wanted to provide evidence of the interplay between acoustic-level and contextual-level cues in French.

The present study focuses on the exploitation of acoustic-phonetic and semantic cues during sentence processing in French and their role in lexical access and meaning disambiguation. Elision is a phonological phenomenon that occurs when a function word with a vowel in its final position is followed by a vowel-initial (or [h]-initial) word, such as in *l'affiche* – produced as [l.afiʃ] instead of [lə.afiʃ] (*\*la affiche*)–. This resyllabification makes both DPs phonemically identical and underscores the importance of accurate segmentation for successful comprehension. Therefore, the particular case of elision may hinder speech segmentation in French (Gaskell et al., 2002; Gustafson & Bradlow, 2016; Shoemaker, 2010).

Previous studies have found certain acoustic differences between pairs of phonemically identical DPs such as *la fiche* – *l'affiche* or *la mie* – *l'amie* (see Spinelli et al., 2007, 2010; Welby, 2003, 2007). Particularly, a higher F0 across the first vowel has been shown to indicate the beginning of content words, i.e., at the beginning of [la] in *l'amie* but later in *la mie*, at [mi]. Moreover, durational differences were found between both items. The initial [l] was longer in elided DPs (e.g., *l'amie*), compared to non-elided ones (e.g., *la*

*mie*), and reversely for the first vowel [a], which was longer in non-elided items than in elided ones. These studies have found that acoustic correlates of homophonically identical DPs are salient enough to French listeners during speech processing. Spinelli et al. (2010) showed that French listeners were able to discriminate between ambiguous DPs such as /lafɪ/ and identify much better than chance which word they heard. This evidence has proven that, in addition to lexical knowledge, listeners rely on other types of word boundary cues, such as acoustic-phonetic information, to compensate for resyllabification processes during speech segmentation (Gaskell et al., 2002; Shoemaker, 2014; Spinelli et al., 2002, 2003; Tremblay & Spinelli, 2013, 2014). Yet, the exploitation of these cues in natural speech remains understudied.

Most of these studies have used behavioural measures that involve tasks revealing limited effects during speech segmentation. The event-related potentials (ERPs) technique, in contrast to behavioural measures, captures subtle changes in brain activity related to the detection of subtle changes in the signal stream. For instance, the MMN ERP component reflects the processing of unexpected changes in some features of a regular continuous auditory stream. In an MMN study, Do Carmo-Blanco et al. (2019) addressed the question of whether non-contrastive features present in ambiguous DPs (e.g., /lafɪ/) are relevant for speech segmentation in French. A rare sound (deviant) is displayed in a series of frequent stimuli (standard). The component is a negative amplitude wave that peaks between 100-200 ms after the deviance onset in fronto-central scalp sites. This component reflects memory traces from input regularities. The authors first tested the syllables *l#a* and *la#* from *l'affiche* and *la fiche*, respectively. In a second study, they tested the complete DPs (*l#affiche* – *la#fiche*). Results showed MMN effects for deviants in both syllable and phrase items, meaning that listeners were able to detect acoustic differences at both syllable and phrase levels. These findings revealed the importance of such acoustic-phonetic information even though duration and pitch are non-contrastive cues in French, i.e. these cues do not modulate meaning on their own. Moreover, the effect was observed despite intra-speaker variations between productions. The authors associated these effects with the idea that the representations of speech units in memory are stored with subphonemic information, which is likely processed automatically without activating different semantic networks for the disambiguation of the targets. However, word recognition and meaning disambiguation of phonemically identical DPs have not yet been tested within contextual information.

## 6.2. Goals and hypotheses

What cues lead to lexical and semantic access for homophonic DPs in situations of ambiguity? How does the sensitivity to fine-grained acoustic details change at the sentence level? The present study aimed to build upon the findings of Do Carmo-Blanco et al. (2019) and Spinelli et al. (2010) and explores whether listeners process subphonemic cues of phonemically identical DPs (e.g., /lafɪʃ/) even when contextual information is provided, or if the context functions as a predominant cue, overriding these effects while processing ambiguous speech. Additionally, the interplay of low- and high-level cues in French was explored to evaluate the applicability of the hierarchical framework proposed by Mattys et al. (2005) in the specific context of ambiguity processing within elision.

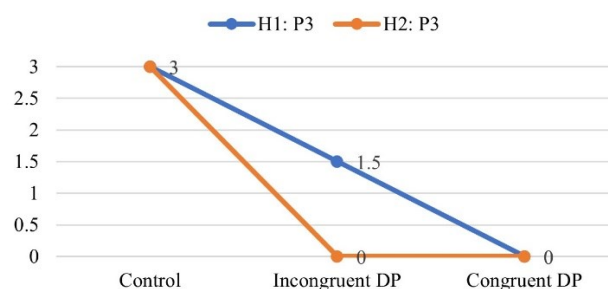
To evaluate this, we administered homophonic DPs such as /lami/ or /lafɪʃ/ embedded in sentences that were either favouring (congruent) or disfavouring (incongruent) their meaning. They were included in two different tasks that modulated attention toward the DPs. Participants were presented with a semi-passive sentence judgement task (Experiment 1) and a word identification task (Experiment 2). As control condition, semantically anomalous sentences were included. See Method for more detailed information. If acoustic information from homophonic DPs is salient enough for French speakers, as expected from Do Carmo-Blanco et al. (2019), low-level information (acoustics) will outweigh high-level information (context) (H1). Under these circumstances, early and late ERP components will reflect differences in the processing and lexical retrieval of DPs embedded in congruent and incongruent context conditions. However, if contextual information plays a major role in speech processing, as expected from Mattys et al. (2005), homophonic DPs will be identified by means of the context and no differences will be found between DPs in congruent and incongruent conditions (H2).

Following H1, we hypothesised that if acoustic-phonetic information of homophonic DPs is perceived and processed by listeners, attention may automatically be drawn to fine-grained acoustic cues in Experiment 1 (semi-passive sentence judgement task). Specifically, it has been shown that the P3a ERP component can be generated when processing non-target rare stimuli if sufficient attention is paid (Comerchero & Polich, 1999; Katayama & Polich, 1998; G. McCarthy et al., 1997; Verbaten et al., 1997). The detection of deviant stimuli produces a positive-going waveform with a maximum over the parietal brain regions, peaking around 250–500 ms (Katayama & Polich, 1998; Polich,



2007; Volpe et al., 2007). We, therefore, expected a larger P3a amplitude to homophones in incongruent contexts compared to those in congruent contexts. Moreover, given that the oddity of semantically anomalous words is more prominent than that of incongruous homophonic DPs, we expected a greater positivity (P3a) for the control compared to the incongruent conditions, with the P3a amplitude being larger for control compared to congruent conditions (Friedrich & Friederici, 2004; Lau et al., 2008). While in Experiment 1 the focus of attention was directed towards the contextual meaning (i.e. semantic information) through a sentence judgement task, in Experiment 2 a word identification task focused participants' attention on the lexical forms, i.e. the DPs. Under these circumstances, the processing of acoustic cues may elucidate a P3b component instead. The P3b component arises when the attentional focus required during a task is directed towards a rare target stimulus (Comerchero & Polich, 1999; Volpe et al., 2007). The component has a different spatial distribution than P3a. The P3b has a maximum over the parietal regions peaking around 250–500 ms (Polich et al., 1997; Simons et al., 2001; Spencer et al., 2001). Therefore, we expected to observe the same effects in Experiment 2 as in Experiment 1, but with different scalp locations –precisely, over central to parietal sites–.

On the other hand, following H2, if listeners fail to process the acoustic-phonetic information of homophonic DPs, control sentences will yield larger P3 amplitudes than both congruent and incongruent sentences in both experiments. No differences will therefore be observed for this component between the homophonic DP conditions. An illustration of H1 and H2 hypotheses for the P3 component is given in Figure 6.

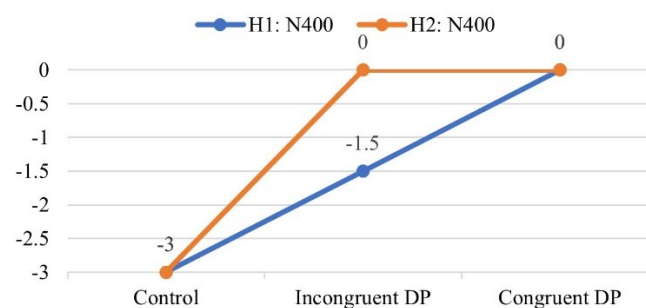


**Figure 6.** Illustration of H1 (blue line) and H2 (orange line) hypotheses for the P3 amplitudes across conditions (control, incongruent and congruent).

Moreover, higher levels of information (semantics) are crucial to achieving understanding during speech processing. If our expectations from H1 are met, integrating acoustic-phonetic details of target DPs may affect word segmentation and may, therefore, hinder

lexical access to some extent, rendering speech comprehension less efficient. Neural correlates to lexical-semantic processing are typically reflected by the N400 component, which peaks between 300–500 ms after stimulus onset and is centro-parietally distributed (Kutas & Hillyard, 1980). Stimuli that are semantically incongruous in a given context elucidate greater N400 amplitudes than those semantically congruous. It is important to note that the amplitude of the N400 does not serve as an indicator of stimulus anomaly; rather, it reflects the absence of contextual support. This effect reflects the effort for stimulus integration into semantic context, being stronger for unexpected than for expected stimuli (Friedrich & Friederici, 2004; Lau et al., 2008). Hence, we hypothesised that if homophonic DPs in incongruous contexts are processed as deviants (i.e. unexpected within the given context), they will yield a larger N400 amplitude compared to homophonic DPs in congruous contexts. Since the control condition is strongly semantically incongruent, we expected a larger negativity (N400) for the control condition compared to the incongruent one. Moreover, we expected that control utterances elucidate a larger N400 effect compared to congruous utterances.

However, if expectations of H2 are met, the control condition will yield larger N400 amplitudes than both congruent and incongruent DPs in both experiments, and no differences will be found for this ERP component between DP conditions. An illustration of H1 and H2 hypotheses for the N400 component is given in Figure 6 and Figure 7.



**Figure 7.** Illustration of H1 (blue line) and H2 (orange line) hypotheses for the N400 amplitudes across conditions (control, incongruent and congruent).

Overall, we expected that homophonic DPs elicit different amplitudes of both ERP components of interest (P3a/b and N400) if listeners processed acoustic-phonetic information. Moreover, we expected to find differences between critical conditions in the behavioural measures (Experiment 2). We expected an increased number of error rates

and slower reaction times (RTs) when identifying DPs in incongruent contexts, while fewer errors and faster RTs identifying DPs in congruent conditions.

Conversely, if no differences were found between homophonic DPs, we hypothesised that listeners failed to integrate such information, potentially overshadowed by the prominence of contextual cues. Under these circumstances, we expected to find no differences between critical conditions in the behavioural measures (Experiment 2). That is, we expect similar accuracy rates and RTs identifying DPs in congruent and incongruent conditions.

Our study aims to provide additional empirical evidence concerning the use of fine-grained acoustic and contextual cues in French. The outcomes will shed light on the neural mechanisms implicated in the processing and comprehension of ambiguous speech in French speakers. Specifically, our results will unveil the ERP components that are pertinent during the encoding of ambiguous speech, thereby enhancing our comprehension of the interplay of top-down and bottom-up processes. Both experiments provide valuable insights into the mechanisms underlying the activation and selection of word meanings when encountering acoustically (and semantically) ambiguous words while holding other lexical and pre-lexical factors constant.

### **6.3. Experiment 1: Electrophysiological evidence to online segmentation during auditory perception with a judgement task (focus on semantics)**

This experiment explored the cost of ambiguity during the segmentation of speech signal in French. The interplay between low- and high-level cues during the processing of spoken sentences was examined through a task that drew focus onto the semantics of the utterance.

#### **6.3.1. Method.**

##### **6.3.1.1. Participants.**

Twenty-three French native speakers (13 females, mean age  $21 \pm 2.42$ ) were admitted to this experiment. At the time of testing, they were all students at the Côte d'Azur University. Prior to the experiment, we collected demographic data using a brief questionnaire (see Annex IV). All participants reported being right-handed based on the

Edinburgh Handedness Inventory (Oldfield, 1971) and had corrected or normal vision. None of the participants reported a history of psychiatric or neurological disorders or current use of medications that might affect their central nervous system. All participants reported having had enough hours of sleep before the experiment and were not under the influence of alcohol. None of them reported hearing or visual difficulties. The participants received 15 euros in remuneration for their time.

### 6.3.1.2. *Materials.*

Forty-eight phonologically identical pairs of DPs were selected (e.g., *l'amie / la mie*, both /lami/; or *l'affiche / la fiche*, both /lafif/). See a complete list of the stimuli in the Annex I. The DPs were embedded in sentences with a strong semantic restriction in favour of one of the two candidates for each pair of the DPs. Half of the DPs were placed in a congruent context that favoured their meaning (*congruent condition*) and half in an incongruent context, favouring the meaning of the other homophone (*incongruent condition*), resulting in a total of 96 experimental sentences. As an example, in *Le collégien a rencontré l'amie de Paul* (“The high-school student met Paul's friend”), the meaning of *l'amie* (“the friend”) is selected by the context in which it is presented (congruent condition); however, the meaning of *l'amie* in *Le boulanger a découpé l'amie de pain* (“The baker cut up the bread friend”) is not appropriate in the given context (incongruent condition).

In addition, 48 non-congruent, non-homophonic sentences served as a control. These sentences consisted of grammatical sentences containing semantic violations. Semantically anomalous words in control sentences had the same characteristics as the experimental DPs: half of the words were preceded by the elided definite article (*l'*) and half were preceded by the non-elided definite article (*la*). For example, *Le chien tire sur l'amidon pendant la promenade* (“The dog pulls on the starch during the walk”) and *Le cordonnier répare la tendresse du directeur* (“The shoemaker repairs the director's tenderness”), respectively. See Annex I for the complete list. Semantically anomalous sentences were included as a control to enable comparisons between conditions in the analysis of EEG data. As mentioned in the hypotheses of the current study, we expected to observe an N400 ERP component yielding different amplitudes across conditions. However, to our knowledge, no previous studies have tested the processing of acoustic and semantic cues of homophonic DPs in elision environments in French. Hence, we

chose to incorporate sentences that ensured an N400 effect and enabled proper interpretation of the data.

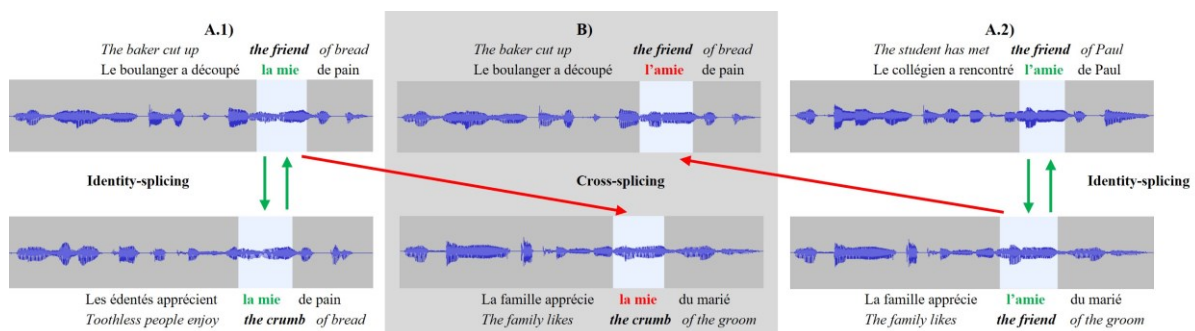
Each member of the pair was recorded in two different sentences, resulting thus in four sentences per pair: two for each elided DP and two for each DP without elision. All sentences were recorded by a French speaker, naïve to the goals of the study and normalised at 65 dB-A, broadcasted at a comfortable listening level (65 dB SPL). The length of the sentences was kept between 6 and 13 words ( $M = 8.54 \pm 1.48$  words, see Table 5 for more details) and was normalised to 44.1 Hz. A one-way ANOVA test was performed to evaluate possible effects of sentence length for each context condition. However, no significant differences were found ( $F(2,141) = 1.50, p = 3.06$ ). Another aspect that has been widely tested in studies focused on auditory sentence processing is the influence of the position of the target. A large body of research has shown how altering the target position can lead to variations in the effects tested (e.g., Adda-Decker et al., 2008; Ernestus et al., 2006; Meunier & Espesser, 2011). No significant differences were found for relative positions of DPs in sentences across conditions ( $F(2,141) = 1.82, p = 3.06$ ). The relative positions of target DPs are shown in Table 5 for every context condition.

**Table 5.** Mean (M), Standard deviation (SD), and range values across conditions (congruent, incongruent, control) for sentence total length, duration of target DPs, and target DPs' relative position within sentences.

	CONTEXT CONDITION								
	Congruent			Incongruent			Control		
	Sentence length	Target position	Target duration	Sentence length	Target position	Target duration	Sentence length	Target position	Target duration
M	8.46	4.90	0.64	8.38	4.85	0.66	8.88	5.23	0.78
SD	13	1.04	0.14	13	1.01	0.21	12	1.12	0.11
Range	6	4 – 8	0.37 – 0.9	6	4 – 8	0.36 – 1.85	6	4 – 8	0.53 – 1.02

Two acoustic manipulations were performed for each DP using Audition software (see Figure 8). The first, *cross-splicing*, consists of exchanging the target DPs between sentences. Incongruent sentences were created by cross-splicing each DP into congruent sentences and placing them in the context favouring the other DP. This was done to avoid introducing acoustic differences resulting from the surprise effect caused by words

contained in unexpected contexts. The second manipulation, *identity-splicing*, consists of exchanging a target word with another production of the same word, which was applied to sentences in the congruent condition. The aim of this manipulation is to enable the presentation of comparable stimuli, i.e. all of which have undergone the same acoustic manipulations. We evaluated possible differences across context conditions. Table 5 shows the duration means and standard deviations of DPs in each condition. Differences between DP durations were evaluated through a one-way ANOVA test, evidencing that there were no statistically significant differences between durations across context conditions ( $F(2,141) = 10.36, p = 3.06$ ).



**Figure 8.** Illustration of the acoustic manipulations performed to generate our stimuli: A) cross-splicing, B) identity-splicing. Green arrows show the generation of the congruent condition, and red arrows show the generation of the incongruent condition. Panel (1) illustrates the two original productions of *la mie*, panel (2) illustrates those of *l'amie*.

### 6.3.1.3. Procedure.

Prior to starting the experiment, participants were provided a one-page description of the experiment, followed by consent forms, a linguistic, and a medical questionnaire. The questionnaire was designed to verify that the participants did not fall within our exclusion criteria.

The experiment took place in an electrically and acoustically shielded booth, sitting in a comfortable armchair in front of an LCD. The stimuli were programmed using Psychopy (Peirce et al., 2019, version 2020.2.5) and were presented aurally to participants. Each trial started with a fixation cross displayed for 1s in the centre of the screen followed by the presentation of each sentence which the participants were asked to listen to attentively. A sentence judgment task was used as an attention check. In this task, a written sentence was displayed at the offset of the auditory sentence and participants were asked to judge

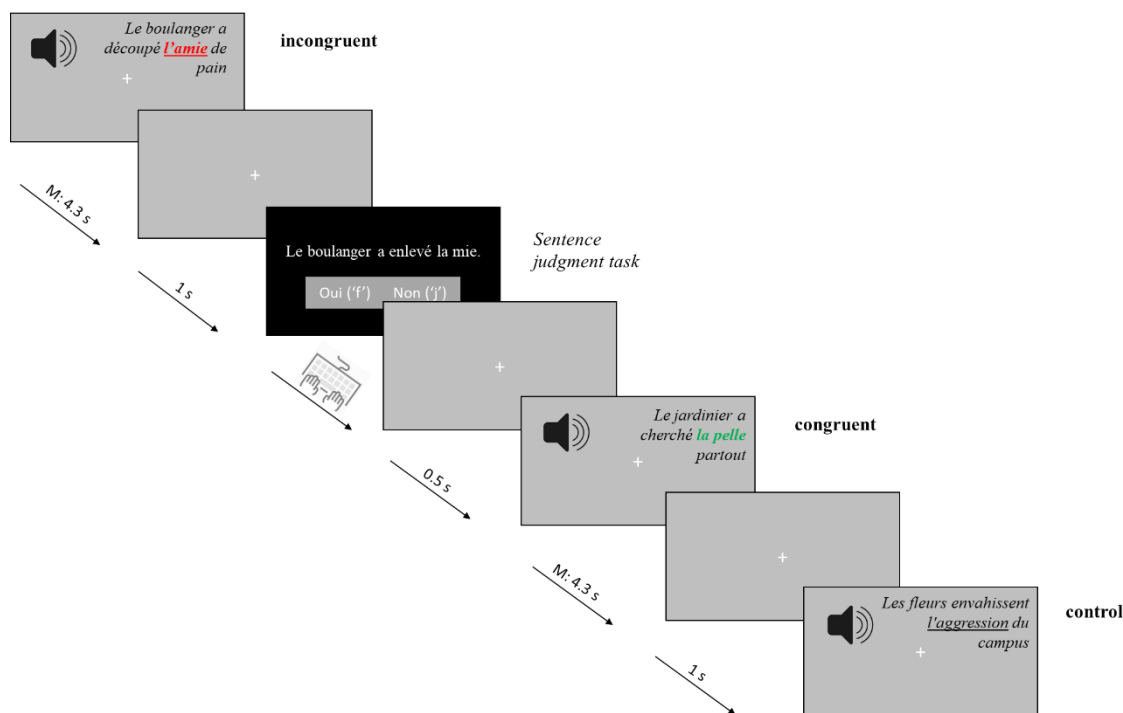
whether they were semantically related by a two-alternative *forced choice* (*yes-no answer*). Participants used two keyboard buttons to respond, one on the left side (*f*) and one on the right side (*j*). The choice of button for yes-no answers was balanced among participants. *The sentence judgment task* appeared randomly but not after each trial. The ratio was 15 out of 36 sentences, i.e. 42% of the time. The order of appearance was pseudorandomised to prevent anticipatory effects (see Figure 9 for an illustration of the procedure).

This setting was repeated along 5 blocks: 1 training block consisting of 11 spoken sentences (4 incongruent, 4 congruent, and 3 control), after what followed 4 testing blocks each containing 12 congruent sentences, 12 incongruent ones, and 15 controls. Critical DPs were split into four different experimental blocks to avoid the proximity of homophone candidates, such that every block contained a candidate of each condition (see Table 6 for an illustration). All stimuli sets were randomised among participants, keeping the training stimuli consistent but displayed in a different order. Each block thus consisted of a total of 36 sentences.

**Table 6.** Illustration of the distribution of sentences along the 4 experimental blocks (example pair *l'amie–la mie*)

<b>Experimental block</b>	<b>Sentence (context condition)</b>
Block 1	<i>La famille apprécie <u>la mie</u> du marié</i> (non-elided incongruent)
Block 2	<i>Les édentés apprécient <u>la mie</u> de pain</i> (non-elided congruent)
Block 3	<i>Le boulanger a découpé <u>l'amie</u> de pain</i> (elided incongruent)
Block 4	<i>Le collégien a rencontré <u>l'amie</u> de Paul</i> (elided congruent)

A pause of up to 2 minutes followed every block to enable participants to relax and blink freely. Interstimuli intervals (ISI) were added after every presented stimulus. During ISIs, a fixation cross was displayed at the centre of the screen to keep the participants' sight focused. One-second ISI was inserted between sounds to ease the perception of every utterance; one-second ISI was added before each sentence presentation; a half-second ISI was presented after keypress to minimise muscle artifacts in the EEG data recollection (M. X. Cohen, 2014). Participants were instructed to answer as quickly as possible to promote responses after early perception and to avoid metalinguistic awareness processes. ERPs were recorded from the onset of target DPs.



**Figure 9.** Example of the sequence of the presentation of stimuli for Experiment 1.

### 6.3.2. EEG recording and preprocessing.

Neurophysiological measures were acquired using the Curry 8.0 XS software from Compumedics, version 8.0.6 X. EEG was recorded from 64 active channels (FP1, FPZ, FP2, AF3, AF4, F11, F7, F5, F3, F1, FZ, F2, F4, F6, F8, F12, FT11, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, FT12, T7, C5, C3, C1, CZ, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, M1, M2, P7, P5, P3, P1, PZ, P2, P4, P6, P8, PO7, PO3, POZ, PO4, PO8, O1, OZ, O2, CB1, CB2), mounted in an elastic cap (Neuroscan Quik-Cap Neo Net). The elastic cap was fitted according to the standard international 10/20 positioning system, with the vertex electrode (Cz) placed at 50% of the distance between inion and nasion and the left and right ear. Each elastic cap had 4 integrated bipolar leads for vertical and horizontal eye movements (VEOG and HEOG channels, respectively). The left VEOG channel was placed 1 cm below the left eye to record blinks. The cap was directly connected to a Neuroscan SynAmps 2/RT amplifier, which uses a 24-bit sampling method. Recording bandwidth was 0.1–400 Hz and the sampling rate was 1000 Hz.



EEG sessions were recorded for  $22.02 \pm 1.06$  seconds on average. The data was referenced online to both a reference and a ground integrated into the cap. Impedances were maintained below 5 k $\Omega$ . No online filter was used.

EEG data were preprocessed and analysed in EEGLAB version 2019 (Delorme & Makeig, 2004) within MATLAB (R2019b). Data were first filtered offline with a bandpass filter of 0.1–48 Hz (Hamming windowed sinc FIR filter) and bad channels were removed from continuous data using the Artifact Subspace Reconstruction (ASR, *clean\_rawdata* plugin, implemented for EEGLAB –version 2.3–, Delorme & Makeig, 2004), and verified visually. Channels were rejected if characterised by a flatline during more than 20s, if a minimum of nearby channel correlation of 0.8 was detected, and if the high-frequency noise exceeded 4 SD (Line Noise criterion) (see Heidlmayr et al., (2021) for a similar procedure). Channels in the range of 0%–5% for each participant were rejected. Epochs were extracted for a span of -200 to 1100 ms relative to the critical DPs' onset (i.e. from the beginning of the definite article). After a visual inspection among different pre-stimulus baseline periods (from -50 ms to -200 ms), a 200-ms baseline period was chosen for baseline correction using the mean after epoching. As recommended by Luck (2014), this was done to minimise the effects of alpha oscillations and avoid post-stimulus contamination of overlapping activity from previous trials or preparatory activity occurring before the onset. We rejected epochs with excessive muscle artifacts ( $>100 \mu\text{V}$ ), which resulted in the rejection of between 0% and 9% of epochs per participant.

Next, a Picard decomposition (Ablin et al., 2018) was performed, followed by an automatic EEG-independent component classifier (ICLabel; Pion-Tonachini et al., 2019) to correct for eye, muscle, and channel noise artifacts present in a high percentage ( $>80\%$ ) in EEG data. After that, the data was checked again on the level of epochs to find shorter periods of noisy signal. Interpolation and re-reference procedures were performed as a final step to minimise signal-to-noise ratio (Cohen, 2014). We ran an interpolation (spherical) for missing channels and the reference was recomputed to average.

We discarded 4 pairs of critical DPs due to technical issues and 2 participants due to high trial rejection. A total of 21 participants and 88 critical items were included in further analyses. Participants had on average  $32.53 \pm 2.97$  trials for the congruent context condition,  $32.32 \pm 3$  trials for the incongruent condition, and  $34.84 \pm 2.46$  for the control condition.

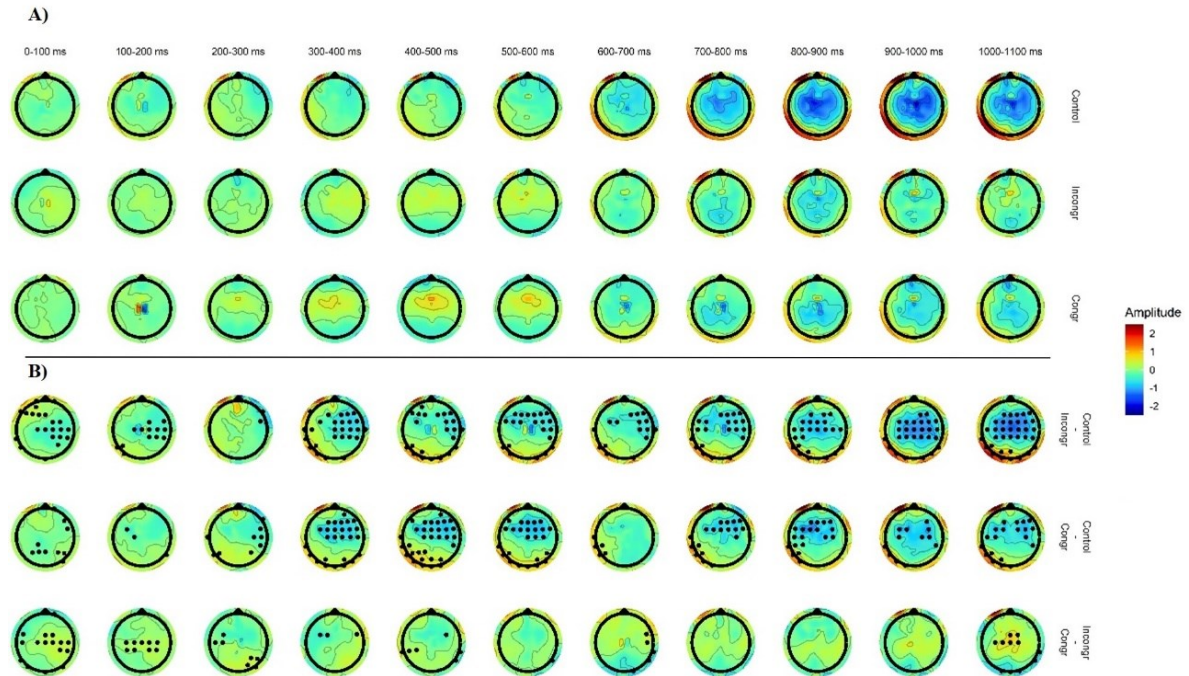
### 6.3.3. EEG data analysis.

Due to the noisy nature of the EEG signal during auditory processing and the noise present in the baseline period of epochs, we performed an analysis to increase the signal-to-noise ratio and minimise Types I, II, S, and M errors (Luck, 2014). We used a data-driven approach that enabled us to find the ERP components of interest based on time windows best suited our data. To this end, permutation tests were performed in R with the *permutes()* package (Voeten, 2022). The rationale for adopting this approach was threefold. Firstly, the stimuli used were natural spoken sentences and our aim was to keep sentences as natural as possible, keeping some variability across sentences. Consequently, the words preceding and following DPs of interest were not identical, and the sentence structure was not fully maintained across sentences (neither within nor between conditions). This variability can introduce noise due to the overlap of different cognitive processes. Secondly, sentences contained homophonic DPs, such as /lami/, that were either congruent or incongruent with the context. The incongruent condition was limited to the acoustic-phonetic level. Hence, such subtle differences may not yield large ERP effects. Thirdly, the choice of materials and tasks in EEG studies has implications in the time windows within which ERP components occur (e.g., Katayama & Polich, 1998; Kutas & Federmeier, 2011; Polich, 2007). Specifically in the current study, no previous evidence could guide our expectations regarding the time windows within our paradigm. These reasons made the data-driven approach appropriate for our experiment.

To this end, groups of electrodes displaying significant effects were determined via permutation tests (Maris & Oostenveld, 2007) in steps of 100-ms time windows. In our study, two ERP components were of interest: P3 and N400. Permutation tests found early significant differences (P3a) across conditions between 300 and 600 ms. For the later component, the N400, a time window between 600–900 ms was found. Scalp maps of these effects are shown in Figure 10. Our analyses seem to exhibit late time windows of both ERP components. This will be addressed in the Discussion.

Linear mixed models (LMM) were performed for every time window of interest. LMM tests included Condition (congruent, incongruent, control), Spatial distribution (anterior, central, posterior), and Laterality (left, midline, right) as fixed factors. Moreover, participants were included as random factors. Electrodes were chosen for each region of interest (ROI), which were defined by crossing the two factors of scalp regions: Region (frontal vs. central vs. parietal) and Hemisphere (left vs. midline vs. right). Each ROI

included three electrodes: frontal left (F7, F5, F3), frontal right (F8, F6, F4), central left (T7, C5, C5), central right (T8, C6, C4), parietal left (P7, P5, P3), and parietal right (P8, P6, P4). Midline ROIs included 1 electrode only, frontal midline (Fz), central midline (Cz), and parietal midline (Pz).



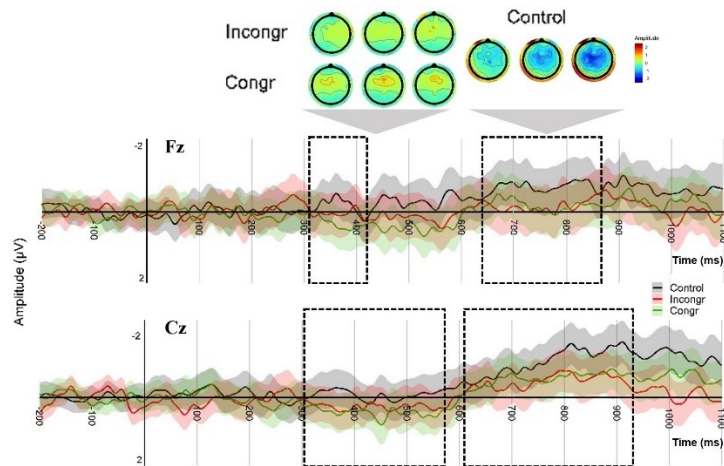
**Figure 10.** Scalp-maps of the ERP waveforms in Experiment 1 covering a time window from 0 to 1100 ms. Panel (A) shows the signal amplitude (in  $\mu\text{V}$ ) for control, incongruent, and congruent conditions. Panel (B) shows the amplitude differences resulting from permutation tests: control-incongruent, control-congruent, and incongruent-congruent. The blue dots indicate the electrodes showing a statistically significant effect, corresponding to channels in which  $p$ -values reached significance at  $\alpha \leq .05$ .

#### 6.3.4. ERP results.

Results of ERP waveforms and scalp-maps time-locked to the stimulus onset are presented in Figure 11.

##### *Late P3a*

There was a main effect of Condition ( $F(2, \text{Inf}) = 554.163, p < .0001$ ), of Spatial distribution ( $F(2, \text{Inf}) = 859.728, p < .0001$ ), and of Laterality ( $F(2, \text{Inf}) = 264.335, p < .0001$ ). Interactions were found between these factors: Condition  $\times$  Spatial distribution ( $F(4, \text{Inf}) = 614.001, p < .0001$ ), Condition  $\times$  Laterality ( $F(4, \text{Inf}) = 517.901, p < .0001$ ), and Spatial distribution  $\times$  Laterality ( $F(4, \text{Inf}) = 23.819, p < .0001$ ). Moreover, a double interaction was found significant, Condition  $\times$  Spatial distribution  $\times$  Laterality ( $F(8, \text{Inf}) = 12.568, p < .0001$ ).



**Figure 11.** Mean ERP amplitudes (and SDs) of frontal midline electrode Fz and central midline electrode Cz in Experiment 1. Black lines show ERP amplitudes for the control condition, red lines for incongruent, and green lines for congruent. Scalp maps displaying the effects of interest are shown on the top (P3a on the left, N400 on the right).

Interestingly, contrary to what was expected, the congruent condition elicited larger positivities (M amplitude = 0.35  $\mu\text{V}$ , SD 4.07) than the incongruent (M amplitude = 0.02  $\mu\text{V}$ , SD 3.32) and the control conditions (M amplitude = -0.54  $\mu\text{V}$ , SD 3.01). Mean ERP amplitudes were more positive in the congruent condition than those in the control condition ( $ps < .01$ ). This effect was maximal over anterior to central electrodes at midline scalp sites ( $ps < .01$ ). Moreover, mean amplitudes were more positive for congruent than for incongruent conditions over anterior midline brain regions ( $ps < .01$ ). Although the observed effect was small, a positivity was found larger for the incongruent condition than the control condition ( $ps < .01$ ). This effect was largest over anterior to central midline electrodes ( $ps < .01$ ).

#### *Late N400*

There was a main effect of Condition ( $F(2, \text{Inf}) = 500.239, p < .0001$ ), of Spatial distribution ( $F(2, \text{Inf}) = 942.576, p < .0001$ ), and of Laterality ( $F(2, \text{Inf}) = 2542.712, p < .0001$ ). We found significant interactions between these factors: Condition  $\times$  Spatial distribution ( $F(4, \text{Inf}) = 328.583, p < .0001$ ), Condition  $\times$  Laterality ( $F(4, \text{Inf}) = 329.590, p < .0001$ ), and Spatial distribution  $\times$  Laterality ( $F(4, \text{Inf}) = 124.122, p < .0001$ ). Moreover, the model revealed a significant double interaction Condition  $\times$  Spatial distribution  $\times$  Laterality ( $F(8, \text{Inf}) = 16.858, p < .0001$ ).

Results showed that the mean ERP amplitudes of our experimental conditions did not significantly differ ( $p > .5$ ). Nevertheless, as expected, the control condition yielded an

N400 effect. The control condition had negative-going mean amplitudes that were larger (M amplitude = 0.7  $\mu$ V, SD = 3.89) than those of both congruent (M amplitude = 0.23  $\mu$ V, SD = 4.02) and incongruent (M amplitude = 0.18  $\mu$ V, SD = 4.14) conditions ( $ps < .02$ ). For both contrasts, the effect was maximal over anterior to central scalp sites, at midline electrodes slightly right lateralised ( $ps < .01$ ).

### 6.3.5. Discussion.

We wanted to examine the interplay of cues involved in processing word boundaries in elision environments in French. Precisely, we explored whether high-level information interferes in the processing of low-level information when acoustic differences between items have been found to be robust in isolation. To this aim, we used phonemically identical DPs, such as *l'amie-la mie*, that can be differently segmented and should lead to the activation of different lexical entries to capture meaning. Contextual information was presented under two different conditions: congruent and incongruent.

Our findings of the P3a component are in line with H1. Significant differences were found between the experimental conditions. However, contrary to our hypothesis, the congruent condition elicited larger positivities (maximal at frontocentral electrodes) compared to both incongruent and control conditions. Interestingly, a P3a effect was found larger for incongruent sentences compared to control sentences. Overall, these findings suggest that listeners were able to process variations of homophonic DPs even when embedded in a sentential context under intact conditions, which contradicts previous findings of Mattys et al. (2005, 2007).

Conversely, N400 effects were in agreement with H2. As expected, control sentences yielded larger N400 effects than both experimental conditions. However, no significant differences were found between experimental conditions. This reveals that lexical access and meaning retrieval processes were performed similarly regardless of the congruity between the DPs' acoustic-phonetic information and the sentential information. Thus, although fine-grained details are perceived by French listeners, context overrides this effect and guides disambiguation during word recognition, which is in line with Mattys et al. (2005, 2007).

## **6.4. Experiment 2: Behavioural and electrophysiological evidence to online segmentation during auditory perception with an identification task (focus on word form)**

In a second experiment, we wanted to verify the effects obtained in Experiment 1 and observe whether focusing attention on the DPs of interest will affect the perception of acoustic-phonetic cues. Thus, in Experiment 2, we modified the focus of attention of the task. While in Experiment 1 attention was driven to the contextual information, in Experiment 2 attention was paid to the lexical forms (i.e. words) embedded within sentences. To this end, we used a cross-modal word identification task where attention was focused on the target words and, thus, on the acoustic-phonetic information.

### **6.4.1. Method.**

#### **6.4.1.1. Participants.**

Twenty-eight participants (22 females, mean age  $22.5 \pm 4.7$ ) were rewarded at the end of the experiment for their participation. None of them participated in Experiment 1. All were right-handed and all had normal or corrected vision. All the participants reported no history of psychiatric or neurological disorders or current use of medications thought to affect the central nervous system. All participants reported having had enough hours of sleep before the experiment and were not under the influence of alcohol. None of the participants had hearing or visual difficulties.

#### **6.4.1.2. Materials.**

The same materials were used as in Experiment 1. For the word identification task, the auditory sentences from Experiment 1 were presented as primes and were matched with three different written word conditions: i) the target homophone (only the noun) displayed in the prime sentence (*homophone condition*), ii) an expected, non-target noun that appeared in the prime sentence (*present condition*), iii) an unexpected, non-target noun that was not displayed in the prime sentence (*absent condition*) (see Table 7 for an illustration of the experimental design). Each participant was presented with only one of the three matching conditions for each sentence.

**Table 7.** Design illustration of auditory primes and written word associations (example pair *l'amie-la mie*).

Context condition	Homophone type	Written target condition
Congruent	Vowel initial	<i>collégien</i> (present)
	e.g., <i>Le collégien a rencontré l'<u>amie</u> de Paul.</i>	<i>menace</i> (absent)
		<i>amie</i> (homophone)
	Consonant initial	<i>édentés</i> (present)
	e.g., <i>Les édentés apprécient la <u>mie</u> de pain.</i>	<i>fontaine</i> (absent)
		<i>mie</i> (homophone)
Incongruent	Vowel initial	<i>pain</i> (present)
	e.g., <i>Le boulanger a découpé l'<u>amie</u> de pain.</i>	<i>menace</i> (absent)
		<i>amie</i> (homophone)
	Consonant initial	<i>famille</i> (present)
	e.g., <i>La famille apprécie la <u>mie</u> du marié.</i>	<i>fontaine</i> (absent)
		<i>mie</i> (homophone)
Control	Vowel initial	<i>chien</i> (present)
	e.g., <i>Le chien tire sur l'<u>amidon</u> pendant la promenade.</i>	<i>baron</i> (absent)
	Consonant initial	<i>directeur</i> (present)
	e.g., <i>Le cordonnier répare la <u>tendresse</u> du directeur.</i>	<i>chasseur</i> (absent)

To avoid misleading effects due to possible memory factors, half of the nouns chosen for the present condition appeared at the beginning of the sentence (before the homophone) and half of them at the end (after the homophone). To avoid semanticity effects between prime context and written word conditions, nouns in the absent condition were kept the same for each homophone prime. For example, *menace* (“threat”) was matched with *amie* and *fontaine* (“fountain”) with *mie* in both congruent and incongruent conditions; for *affiche*, *refuge* (“refuge”) was displayed and *gâteau* (“cake”) was associated with *fiche* in both experimental conditions. To reduce attention to homophonic DPs and avoid bias to our stimuli of interest, we presented 27.8% more written words in non-homophonic conditions than in homophonic conditions (i.e. present and absent conditions).

#### 6.4.1.3. Procedure.

A similar procedure was followed as in Experiment 1. After the setup, participants were instructed to listen attentively to French sentences. In Experiment 2, the task was done

after every trial, i.e. after each sentence a written word was displayed at the center of the screen. Participants were asked to identify whether the written target was present or not in the sentence they had just heard.

For increased accuracy in response latencies, responses were given with a controller, using the upper left and upper right buttons. The choice of button for Yes/No answers was balanced among participants. Importantly, participants were instructed to answer as quickly as possible to promote responses after early perception and to avoid metalinguistic awareness processes. Responses were collected and their reaction times (RTs) were recorded after the onset of the DPs.

As in Experiment 1, to avoid the proximity of homophone candidates during the presentation to participants, they were split into four different experimental blocks (see Table 6). To ensure a balanced presentation of every word condition, stimuli were distributed along three lists, one with each condition, and lists were counterbalanced among participants. Within each list, the order of block and item presentation was randomised.

#### **6.4.2. EEG recording and preprocessing.**

EEG sessions were recorded for  $30 \pm 2.26$  seconds on average. The same EEG recording was conducted as in Experiment 1.

Four participants were rejected from further analyses due to technical issues, representing 14.3% of the data. Data from 24 participants was included in further analyses. The same EEG data preprocessing was followed as in Experiment 1. On average, 8% of electrodes were rejected by participant ( $M = 5.11 \pm 0.85$ ). On average, after extracting epochs from -200 ms to 1100 ms relative to the onset of critical DPs (i.e. from the onset of the definite article), 14% of the trials were rejected by participant ( $M = 19.29 \pm 14.09$ ). There were on average  $42.68 \pm 5.06$  final trials for the congruent condition,  $40.64 \pm 4.99$  trials for the incongruent condition, and  $41.88 \pm 5.21$  trials for the control one.

#### **6.4.3. Data analysis.**

##### **6.4.3.1. Behavioural measures.**

For the behavioural analysis, error rates and RTs were calculated. Only answers for Present and Homophone written word conditions were included, therefore we only considered results for which the correct answer was affirmative. Control sentences were



excluded from further analyses. *Yes* answers were coded as 1 and *no* as 0. Reaction times (RTs) were computed by participant, context condition, and written word condition. RTs shorter or longer than the mean  $\pm$  2 SD were excluded from the analyses. Despite the positive skewness of the RTs, data was not normalised following the recommendations of Wilcox (2017). RTs to correct answers were analysed through LMM, while participants' scores were analysed through generalised linear mixed-effects models (GLMM). Statistical analyses were performed using *R* and the *lme4* package (Bates et al., 2015).

To select the model that best explained the data, the interaction between our conditions of interest (Context and Written word) was tested ( $\chi^2(1) = 30.006, p < .0001$ ). Context (congruent, incongruent) and Written word (target, other word) were added as fixed factors and Participant and Item as random factors. Both intercepts and slopes of random factors were tested on the models. Moreover, word frequency, number of letters (of written targets), and orthographic uniqueness point were added to the model as covariates when possible, with all possible combinations. Comparisons of each model's combination were performed through the ANOVA function to select the best-adapted model as indicated by the lowest Akaike information criterion and the most parsimonious model was kept (Matuschek et al., 2017). Contrasts were assessed with the method of estimated marginal means (Searle et al., 1980) using *emmeans* (R. Lenth, 2016) and *multcomp* (Hothorn et al., 2008) packages. All *p*-values were calculated using the Satterthwaite's model of approximation with the *lmerTest* package (Kuznetsova et al., 2020).

For both RTs and error rate models, the inclusion of random slopes for participants significantly improved the models compared to all other combinations. However, for error rates, the incorporation of random slopes for items improved to a greater extent the model. In order to keep coherence in the analyses and make the best decision, the goodness of fit for each model was assessed through the adjusted  $R^2$  (*r.squaredGLMM*, Bartoń, 2020). RTs' best model, which encompassed random slopes for participants, explained 41.6% of the variability while the model with random slopes for items explained 1.3% less. The best model for error rates, containing random slopes for items, explained 41.9% of the variability while the model with random slopes for participants explained 4.8% less. Hence, models including random slopes for items conveyed overall a greater percentage of data variability for both RTs and error rates, thereby ensuring that

the model's selection did not significantly compromise the goodness of fit for our dependent variables.

#### 6.4.3.2. EEG data.

The same analyses were performed as in Experiment 1. Permutation tests found significant differences for the P3b component between 400 and 500 ms across conditions. For the later component (N400), a time window between 500–900 ms was found significant across conditions. Our analysis revealed a delayed time window for the N400 component, which will be addressed in the discussion.

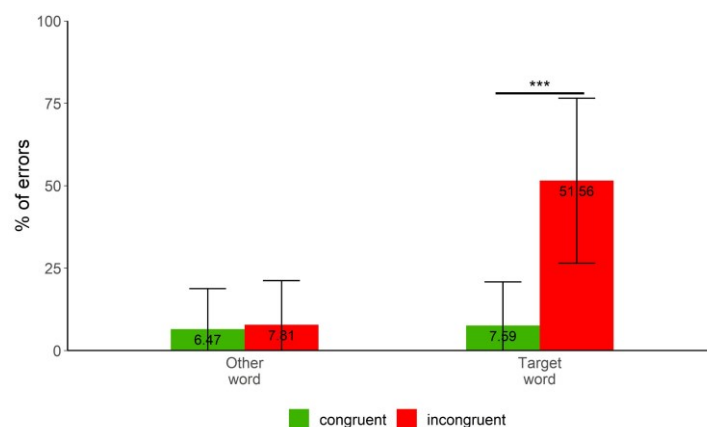
#### 6.4.4. Results.

##### 6.4.4.1. Behavioural measures.

###### *Error rates*

Accuracy rates of the control condition were verified to ensure proper performance in the task. All participants paid attention to the task as indicated by high accuracy rates in the condition (94.7% on average).

On average, participants made few errors when exposed to non-homophonic words that were expected in the sentence, regardless of the context condition (6.47% of errors in congruent sentences and 7.81% in sentences with incongruous DPs). During the identification of homophone targets, few errors were made in the congruent condition (7.59%). However, error rates increased drastically when identifying targets in the incongruent condition (51.56%). The percentage of errors for non-homophonic words and homophone targets is reported in Figure 12.



**Figure 12.** Percentage of errors (and SDs) made during the identification task. Left bars show errors when identifying non-homophonic words (Other word), right bars show errors to

homophone targets (Target word). Green bars show results in the congruent condition, red bars show results in the incongruent condition.

Results from GLMM revealed a main effect of Written word ( $\chi^2(1)= 12.643$ ;  $F= 37.325$ ;  $p <.0001$ ) and a main effect of Context ( $\chi^2(1)= 12.643$ ;  $F= 32.814$ ;  $p <.0001$ ). An interaction was reported Written word  $\times$  Context ( $\chi^2(1)= 12.643$ ;  $F= 32.66$ ;  $p <.0001$ ). To evaluate the effects within contrasts, post hoc tests were performed for the interaction (see the model's coefficients in Table 8). Results revealed statistically significant differences in the identification of homophone targets in function of the context condition. Particularly, participants made more errors in incongruent than in congruent conditions ( $z= -9.272$ ,  $p <.0001$ ).

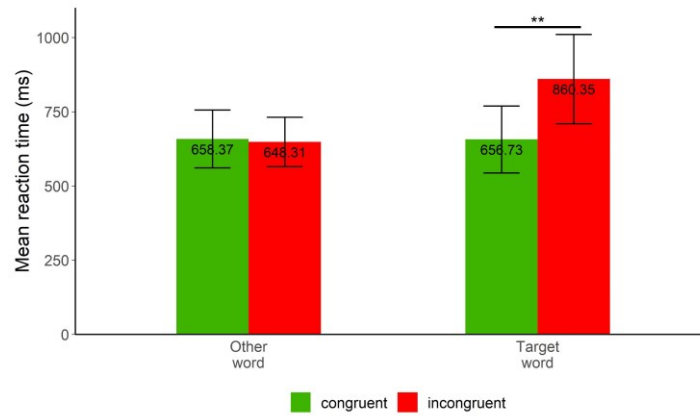
**Table 8.** Error contrasts performed in post hoc tests for Context and Written word.

<i>contrast</i>	<i>estimate</i>	<i>SE</i>	<i>z.ratio</i>	<i>p.value</i>
present congruent - homophone congruent	0.2816	0.4331	0.650	0.914
<b>homophone congruent - homophone incongruent</b>	<b>3.5223</b>	<b>0.3799</b>	<b>9.272</b>	<b>&lt;.0001</b>
present incongruent - homophone congruent	-0.0933	0.46524	-0.200	0.997
<b>present congruent - homophone incongruent</b>	<b>3.8039</b>	<b>0.4466</b>	<b>8.519</b>	<b>&lt;.0001</b>
present congruent - present incongruent	0.3749	0.4886	0.767	0.867
<b>present incongruent - homophone incongruent</b>	<b>3.4291</b>	<b>0.3854</b>	<b>8.89</b>	<b>&lt;.0001</b>

*Note:* Bold indicates statistically significant coefficients.

### *RTs*

Participants' reaction times mirrored the results in error rates (see Figure 13). Similar RTs were found when they had to identify non-homophonic words, regardless of the context condition ( $M = 659.37$  ms in congruent sentences, and  $M = 648.31$  ms in sentences with incongruous DPs). Crucially, participants had similar RTs when identifying homophone targets in the congruent condition ( $M = 656.73$  ms). However, they were slower when presented with homophones in the incongruent condition ( $M = 860.35$  ms).



**Figure 13.** Mean reaction times in milliseconds (and SDs) to correct responses are shown to non-homophonic words (Other word, left bars) and to homophone targets (Target word, right bars). Green bars show results in the congruent condition, red bars show results in the incongruent condition.

Results from the LMM revealed a main effect of Context ( $\chi^2(1)= 35.737$ ;  $F= 17.744$ ;  $p <.0001$ ) and a main effect of Written word ( $\chi^2(1)= 35.737$ ;  $F= 11.860$ ;  $p =.0014$ ). An interaction was reported Context  $\times$  Written word ( $\chi^2(1)= 35.737$ ;  $F= 31.032$ ;  $p <.0001$ ). Results of post hoc tests revealed significant differences in the RTs to homophone targets. Precisely, participants were significantly faster when presented with homophone targets in the congruent condition than in the incongruent condition ( $t=-7.796$ ,  $p <.0001$ ). See Table 9 for the model’s coefficients.

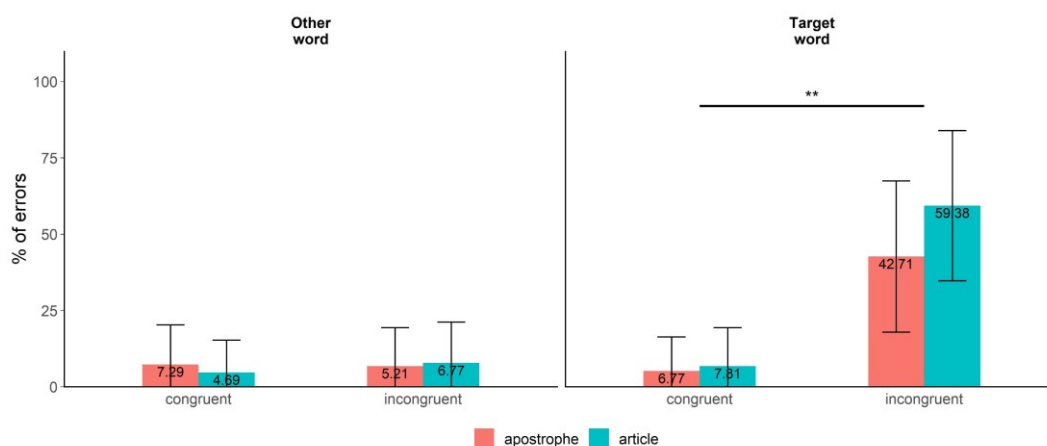
**Table 9.** RT contrasts performed in post hoc tests for Context and Written word.

<i>contrast</i>	<i>estimate</i>	<i>SE</i>	<i>t.ratio</i>	<i>p.value</i>
present congruent - homophone congruent	-5.948	19.696	-0.302	0.990
<b>homophone congruent - homophone incongruent</b>	<b>-93.529</b>	<b>11.997</b>	<b>-7.796</b>	<b>&lt;.0001</b>
present incongruent - homophone congruent	-18.933	19.638	-0.964	0.759
<b>present congruent - homophone incongruent</b>	<b>-99.478</b>	<b>19.696</b>	<b>-5.051</b>	<b>&lt;.0001</b>
present congruent - present incongruent	12.985	14.871	0.873	0.810
<b>present incongruent - homophone incongruent</b>	<b>-112.463</b>	<b>19.638</b>	<b>-5.727</b>	<b>&lt;.0001</b>

*Note:* Bold indicates statistically significant coefficients.

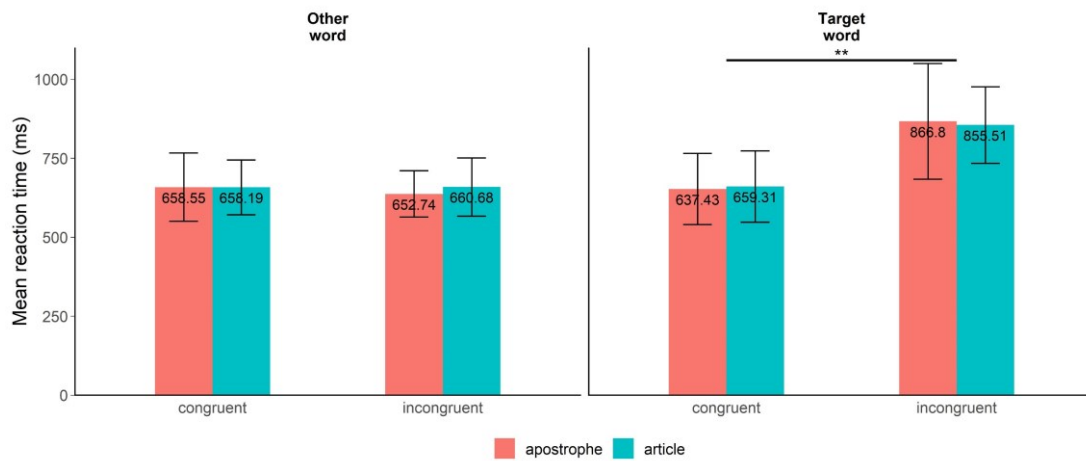
Overall, the findings indicated that participants exhibited a higher error rate when identifying homophone targets in the incongruent condition and were slower to give a response. This indicates that congruity seems to modulate the perception of homophonic DPs in elision environments, which impacts their correct identification. Even though contextual information was available as a cue to integrate the meaning of words within sentences, their acoustic-phonetic attributes were robust enough to influence the participants' perception of targets. Contrary to what was observed in Experiment 1, at least at the behavioural level, sentences that disfavoured homophonic DPs led to confusion during the identification of the lexical candidate.

We further assessed the perception of homophone targets while considering differences between elided and non-elided items. Article type (apostrophe, article) was added as a fixed factor into the model. Figure 14 shows error rates for written words considering the Article type factor. As a reminder, written homophone targets presented to participants corresponded to the same auditory DP displayed in the prime sentence (e.g., heard: *l'amie*, written: *amie*). Again, it can be clearly observed that more errors were made to homophone targets in incongruent contexts (on average, 51.1%) compared to congruent contexts (on average, 5.8%). Specifically, it seems that more errors were made to non-elided items (59.4%) than to elided items (42.7%). However, no main effect of Article type ( $\chi^2(1)= 19.36$ ;  $F=2.913$ ;  $p =.088$ ) was found, and no interaction was found significant either ( $ps >.2$ ), suggesting no differences between items.



**Figure 14.** Percentage of errors (and SDs) made during the identification task. The left panel shows errors in identifying non-homophonic words (Other word), right panel shows errors in homophone targets (Target word). Red bars show results for elided DPs (apostrophe), blue bars show results for non-elided DPs (article). This is shown for both congruent and incongruent conditions.

The same analysis was performed on RTs based on Article type. Overall, results showed that participants were slower in identifying homophone targets in incongruent contexts ( $M = 866.8$  ms for apostrophe,  $M = 855.51$  ms for article) compared to other conditions. This was confirmed by the LMM model, revealing no main effect of Article type ( $\chi^2(1) = 1.563$ ;  $F = 0.076$ ;  $p = .8$ ), as well as no interaction ( $ps > .3$ ).



**Figure 15.** RT means in ms (and SDs) to correct responses in the identification task. The left panel shows RTs when identifying non-homophonic words (Other word), right panel shows RTs to homophone targets (Target word). Red bars show results for elided DPs (apostrophe), blue bars show results for non-elided DPs (article). This is shown for both congruent and incongruent conditions.

Overall, results from behavioural measures support findings in Experiment 1 indicating that the perception of items in elision seems to be influenced by its congruence with the context given. Specifically, French speakers demonstrated the ability to detect intended items in elision, although they were less accurate when items were incongruent with sentential information.

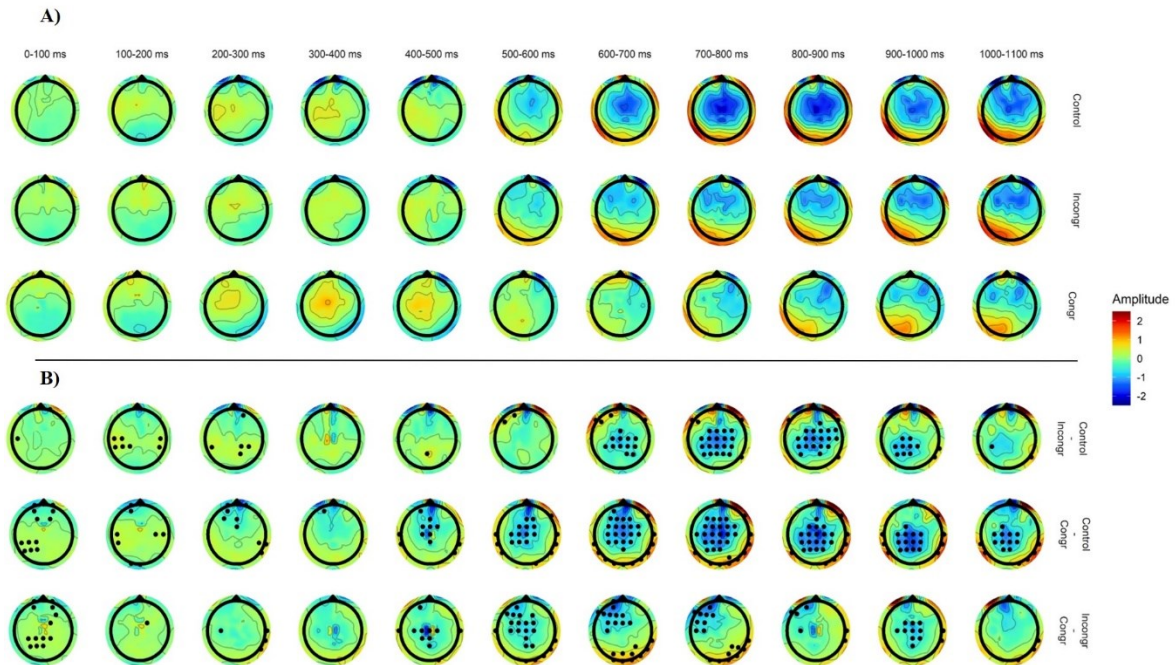
#### 6.4.4.2. ERP results.

Scalp maps of the effects for the complete time window (from 0 to 1100 ms) are shown in Figure 16. ERP waveforms and scalp-maps time-locked to the stimulus onset are presented in Figure 17.

#### P3b

There was a main effect of Condition ( $F(2, Inf) = 56.710$ ,  $p < .0001$ ), Spatial distribution ( $F(2, Inf) = 963.152$ ,  $p < .0001$ ), and Laterality ( $F(2, Inf) = 1518.164$ ,  $p < .0001$ ). An interaction was found between these factors: Condition  $\times$  Spatial distribution ( $F(4, Inf) =$

84.054,  $p < .0001$ ), Condition  $\times$  Laterality ( $F(4, \text{Inf}) = 126.472$ ,  $p < .0001$ ), and Spatial distribution  $\times$  Laterality ( $F(4, \text{Inf}) = 167.435$ ,  $p < .0001$ ). Moreover, a double interaction Condition  $\times$  Spatial distribution  $\times$  Laterality ( $F(8, \text{Inf}) = 32.881$ ,  $p < .0001$ ) was found.



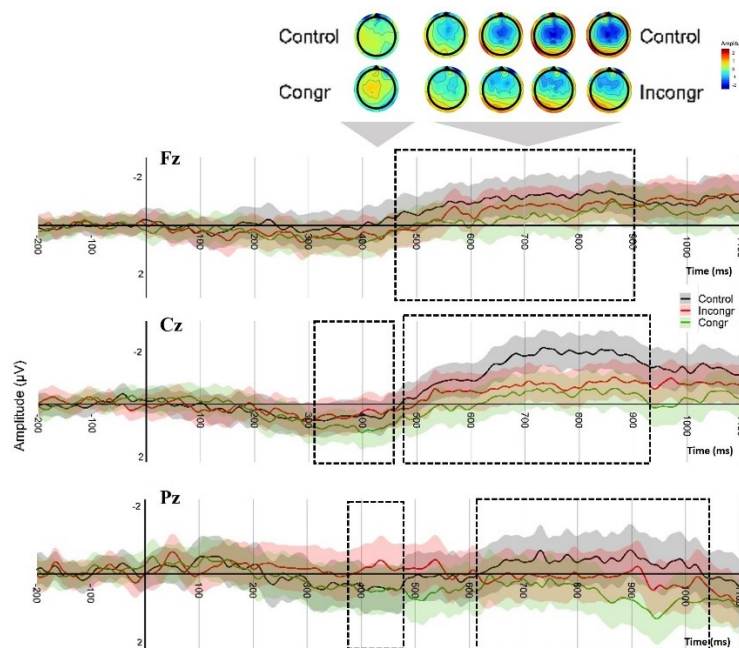
**Figure 16.** Scalp maps of the ERP waveforms covering a time window from 0 to 1100 ms. Panel A) shows the signal amplitude (in  $\mu\text{V}$ ) for control, incongruent, and congruent conditions. Panel B) shows the amplitude differences resulting from permutation tests: control-incongruent, control-congruent, and incongruent-congruent. The blue dots indicate the electrodes showing a statistically significant effect, corresponding to channels in which  $p$ -values reached significance at  $\alpha \leq .05$ .

Mean amplitudes in the congruent condition (Cz: M amplitude =  $0.68 \mu\text{V}$ , SD = 2.78) were more positive compared to those in both incongruent (M amplitude =  $0.32 \mu\text{V}$ , SD = 2.2) and control (M amplitude =  $0.4 \mu\text{V}$ , SD = 2.09) conditions ( $ps < .01$ ). These effects were largest over anterior to posterior electrodes at midline scalp locations and were slightly left-lateralised for both contrasts ( $ps < .01$ ). These effects confirm results from Experiment 1, both reflecting a larger positivity for congruent contexts. The control condition was somewhat more positive than the incongruent condition, the effect being significant only over posterior sites at midline electrodes ( $ps < .01$ ). However, ERP waveforms and scalp maps in Figure 17 do not seem to reflect the presence of a P3b-like component for control sentences.

*Late N400*

There was a main effect of Condition ( $F(2, \text{Inf}) = 802.701, p < .0001$ ), Spatial distribution ( $F(2, \text{Inf}) = 5703.848, p < .0001$ ), and Laterality ( $F(2, \text{Inf}) = 3287.437, p < .0001$ ). An interaction was found between the factors: Condition  $\times$  Spatial distribution ( $F(4, \text{Inf}) = 479.607, p < .0001$ ), Condition  $\times$  Laterality ( $F(4, \text{Inf}) = 479.607, p < .0001$ ), and Spatial distribution  $\times$  Laterality ( $F(4, \text{Inf}) = 235.929, p < .0001$ ). Moreover, a double interaction was found significant for Condition  $\times$  Spatial distribution  $\times$  Laterality ( $F(8, \text{Inf}) = 120.629, p < .0001$ ).

As expected, the control condition was more negative (Cz: M amplitude =  $-1.79 \mu\text{V}$ , SD = 2.01) compared to both incongruent (M amplitude =  $-0.69 \mu\text{V}$ , SD = 2.2) and congruent (M amplitude =  $-0.45 \mu\text{V}$ , SD = 2.43) conditions, the effect being more prominent over anterior to posterior scalp sites ( $ps < .01$ ). The effect was maximal at midline electrodes and was slightly right lateralised for both contrasts ( $ps < .01$ ). Crucially, contrary to Experiment 1, the incongruent condition was more negative compared to the congruent condition ( $ps < .01$ ), largest over anterior to central scalp locations at midline electrodes ( $ps < .01$ ).



**Figure 17.** Mean ERP amplitudes (and SDs) of frontal midline electrode Fz, central midline electrode Cz, and parietal midline electrode Pz. Black lines show amplitudes for the control condition, red lines for incongruent, and green lines for congruent. Scalp topography maps displaying the effects of interest are shown on the top (P3b on the left, N400 on the right).

#### 6.4.5. Discussion.

In this second experiment, we focused on the auditory speech processing of identically homophonic DPs by modifying the task paradigm. We wanted to assess the extent to



which French listeners exploit acoustic-phonetic cues that may be favoured or unfavoured by contextual information when attention is automatically drawn to such cues.

As in Experiment 1 and in line with H1, significant differences were found between experimental conditions for the P3 component. Precisely, the congruent condition yielded larger P3b effects compared to the other two conditions. This confirms our interpretation in Experiment 1 that congruous phrases were processed as rare, unexpected stimuli, yielding a larger positivity. Interestingly, we found a small effect for control sentences compared to incongruous sentences, although differences were scarce between the two conditions. This could prove the existence of a bias towards incongruity in the design of this work that could, consequently, explain that congruent sentences yielded a P3b effect.

Crucially, consistent with H1, the larger positivity present in one of the two critical conditions reflects a distinct perception of our DPs of interest; in other words, homophones that were congruous to sentences were processed differently than those that were incongruous. This suggests that acoustic-phonetic cues present in homophonic DPs in elision environments were processed by listeners, even when contextual information was provided. This confirms our findings in Experiment 1.

Similar to Experiment 1, the control condition yielded a larger negativity compared to the two experimental conditions. This is in line with previous studies reporting a large N400 effect for targets bearing a semantic violation (e.g., Kutas & Federmeier, 2011; Lau et al., 2008; Steinhauer et al., 2017; Steinhauer & Connolly, 2008). Remarkably, contrary to Experiment 1, mean ERP amplitudes in the incongruent condition were significantly more negative than those in the congruent one. This validates our H1. These results indicate that homophonic DPs were processed differently depending on whether the contextual information favoured or disfavoured their meaning. This is linked to the idea that specific semantic meaning is not a necessary condition for the elicitation of the N400 component (Deacon et al., 2004; Erlbeck et al., 2014).

Moreover, results from both behavioural suggest that French listeners have difficulties identifying homophonic DPs in incongruent contexts. This condition yielded responses close to chance level, suggesting that the tendency to hear the other homophone candidate (fitting the context of the presented sentence) is almost as likely as hearing the auditory target.

## 6.5. Tasks comparison

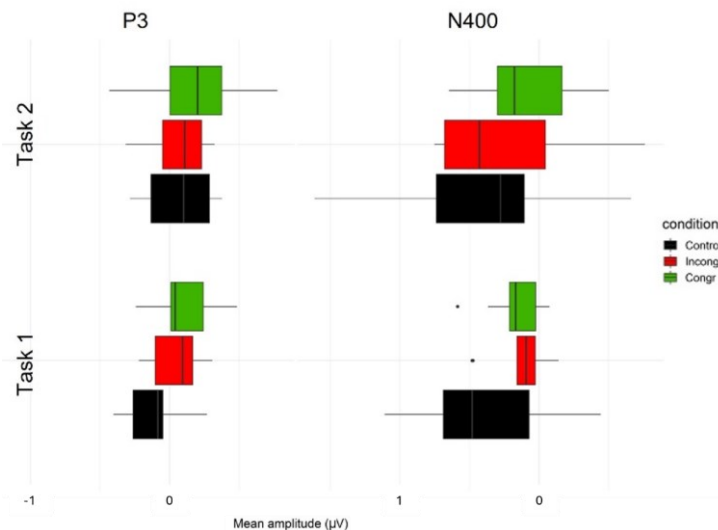
We proceed for a comparison between tasks (Experiment 1 and Experiment 2) in order to provide insight into the differences that exist between our findings. This enables us to thoroughly contrast the outcomes from both tasks in a more objective manner, avoiding task bias and comparing the effects obtained in the ERP components of interest.

### 6.5.1. Analysis.

LMM were performed on the mean ERP amplitudes across conditions obtained for each component of interest at both tasks, separately. The same model as in Experiments 1 and 2 was computed, and the variable Task (first, second) was included as a fixed factor.

### 6.5.2. Results.

Mean amplitudes in both tasks and across conditions are shown in Figure 18.



**Figure 18.** Mean ERP amplitudes of both components of interest (P3, N400) for each task (1, 2). Black boxplots show effects for the control condition, red boxplots for the incongruent condition, and green boxplots for the congruent condition. Error bars show SDs.

### P3

No main effect of Task was found ( $p > .1$ ). However, LMM revealed significant interactions between Task and all other factors: Condition  $\times$  Task ( $F(2, \text{Inf}) = 85.959, p < .0001$ ), Spatial distribution  $\times$  Task ( $F(2, \text{Inf}) = 77.770, p < .0001$ ), and Laterality  $\times$  Task ( $F(2, \text{Inf}) = 167.497, p < .0001$ ). Double interactions were also significant: Condition  $\times$  Spatial distribution  $\times$  Task ( $F(4, \text{Inf}) = 50.067, p < .0001$ ), Condition  $\times$  Laterality  $\times$  Task ( $F(4, \text{Inf}) = 106.499, p < .0001$ ), and Spatial distribution  $\times$  Laterality  $\times$  Task ( $F(4, \text{Inf}) =$

88.018,  $p < .0001$ ). Moreover, an interaction between the four factors was also found significant (Condition  $\times$  Spatial distribution  $\times$  Laterality  $\times$  Task ( $F(8, Inf) = 46.922$ ,  $p < .0001$ )).

Mean ERP amplitudes of control sentences were significantly more positive in the first task compared to those in the second one ( $p < .009$ ), being more prominent over central to posterior scalp locations ( $ps < .03$ ). This effect was also larger at midline electrodes in Task 2 compared to Task 1 ( $p < .01$ ). Moreover, the congruent condition was more positive in posterior scalp sites in Task 2 than in Task 1 ( $p < .03$ ). The positivity was greater over midline electrodes in the second task than in the first task ( $p < .01$ ). However, no significant differences were found for the incongruent condition between tasks ( $ps > 1$ ).

#### *Late N400*

No main effect of Task was found ( $p > .7$ ). Yet, all interactions between factors were found significant: Condition  $\times$  Task ( $F(2, Inf) = 145.096$ ,  $p < .0001$ ), Spatial distribution  $\times$  Task ( $F(2, Inf) = 2372.542$ ,  $p < .0001$ ), and Laterality  $\times$  Task ( $F(2, Inf) = 182.364$ ,  $p < .0001$ ). Double interactions were also found significant: Condition  $\times$  Spatial distribution  $\times$  Task ( $F(4, Inf) = 433.397$ ,  $p < .0001$ ), Condition  $\times$  Laterality  $\times$  Task ( $F(4, Inf) = 392.497$ ,  $p < .0001$ ), and Spatial distribution  $\times$  Laterality  $\times$  Task ( $F(4, Inf) = 190.005$ ,  $p < .0001$ ). Moreover, an interaction between all factors was also found significant (Condition  $\times$  Spatial distribution  $\times$  Laterality  $\times$  Task ( $F(8, Inf) = 48.264$ ,  $p < .0001$ )).

Mean ERP amplitudes of the control condition were more negative over central to posterior scalp regions in the second task than in the first task ( $ps < .01$ ). Moreover, this negativity was larger at midline electrodes in the second task compared to the first one ( $ps < .01$ ). The incongruent condition had larger negative amplitudes in Task 2 than in Task 1 ( $p < .003$ ). Notably, incongruent sentences elicited more negativities over anterior to posterior scalp locations in the first task than in the second task ( $ps < .01$ ).

The analysis of the comparison between tasks revealed that the effects of both ERP components were more prominent in Experiment 2 than in Experiment 1. This provides further evidence that the choice of task can modulate effects in EEG studies (e.g., Katayama & Polich, 1998; Kutas & Federmeier, 2011; Polich, 2007).

### **6.5.3. Discussion.**

Overall, the results confirmed the presence of differences between the tasks. The control condition had mean amplitudes that were more positive over central to parietal sites in the second experiment. Nevertheless, topomaps and ERP amplitudes do not show a P3b-like component in Task 2.

Most differences were found in results from N400 effects. Results showed that the (active) identification task (Experiment 2) yielded larger and broader N400-like effects for the incongruent and control conditions compared to those of the semi-passive semanticity judgment task (Experiment 1). Task seems to modulate the listeners' speech processing when exposed to DPs in elision environments. Exposing participants to an active task of target identification seemed to increase the detection of targets compared to when exposed to a semi-passive task of semantic sentence judgment. This was reflected by the presence of a negative component yielded by incongruous homophones in Experiment 2 compared to an absence of it in Experiment 1. However, the effect was less spread and had disappeared earlier compared to the effect elucidated by non-homophonic semantically anomalous DPs (control sentences). This reflects a reduced difficulty in integrating targets when their acoustic-phonetic cues are incongruent with the context compared to integrating semantically anomalous targets.

### **6.6. General discussion and conclusion**

We tested the perception of acoustic-phonetic information along two experiments that manipulated how attention was driven to information within sentences, by either focusing on the sentential meaning (Experiment 1) or the lexical meaning (Experiment 2). The exposure to homophonic DPs enabled us to disentangle the real-time processing of acoustic-phonetic information and semantic information. Our initial hypothesis was that the processing of acoustic-phonetic cues embedded in disfavoured contexts would result in the presence of P3 and N400 components compared to favoured contexts (H1). Alternatively, context information could override these effects if listeners relied on it as a cue to segmentation rather than on acoustic-phonetic details (H2).

Differences were found in P3 amplitudes between our experimental conditions, a larger positivity being elicited by congruent sentences compared to incongruent ones. This suggests that, when listening to continuous speech, the participants' perceptual system processed homophones that match the context (congruent condition) differently from

those that do not match it (incongruent condition). Moreover, this effect was also revealed by response latencies in Experiment 2. Specifically, listeners were faster at identifying congruent utterances compared to utterances that encompassed incongruent words. These findings are in agreement with H1 and support previous findings on the saliency of acoustic-phonetic information associated with DPs in elision environments in French (Do Carmo-Blanco et al., 2019; Spinelli et al., 2007, 2010).

Contrary to our expectations, semantically anomalous DPs from control sentences did not yield a P3 component in any of the experiments. Surprisingly, P3 effects were found for congruent contexts in both experiments. Prior results have shown that the detection of deviant stimuli produces a positive-going waveform, the P3 component (e.g. Polich, 2007; Volpe et al., 2007). Hence, although it would be logical that sentences containing a word that is semantically anomalous (control sentences) yielded a P3 component, no such effect was found. A possible reason for this is an existent bias towards incongruity in the experimental design, in which 2/3 of the stimuli were incongruent (either semantically or acoustically in control and incongruent conditions, respectively). This bias is reflected in the results of both experiments, the positivity being greater in the second experiment as revealed by the task comparison. This idea is supported by the fact that no differences were found for incongruent or control sentences between tasks, implying a bias towards incongruity in the design of our experiment.

While our findings were similar for the P3 component in Experiments 1 and 2, N400 effects were not. As expected, control sentences yielded an N400 effect in both experiments. Interestingly, the incongruent condition elicited an N400 effect in Experiment 2 only. While experimental conditions did not significantly differ in Experiment 1, findings from Experiment 2 suggest a graduality in the level of difficulty while integrating DPs as a function of context. The most challenging elements were semantically anomalous DPs from control sentences, followed by unexpected homophonic DPs from incongruent sentences, and finally decreasing when exposed to expected homophonic DPs within congruent sentences. This pattern indicates that the effort required for retrieving targets and integrating them into the sentences increases relative to their congruity with the contextual information, which seems to be modulated by the amount of attention directed toward targets. Effects in Experiment 1 align to H2 whereas effects in Experiment 2 align to H1. Our findings show that word recognition and meaning disambiguation processes are temporally disturbed when attention is drawn

to the lexical form (Experiment 2) whereas no such effect is found when attention is driven towards the global meaning, i.e. to sentential information (Experiment 1). These investigations provided further evidence for differences in N400 effects based on the task used (e.g., Deacon & Shelley-Tremblay, 2000; Küper & Keil, 2009). Our findings regarding the segmentation of multi-level information in DPs within elision indicate that while early sensory processing may be relatively sensitive to variations in the speech stream irrespective of attentional focus, attention plays a significant role during cognitive processes unfolding auditory processing such as lexical retrieval and semantic integration. This is consistent with previous work showing that directing participants' attention to the experimental stimuli is relevant for the elicitation of greater amplitudes of the N400 component (e.g., Deacon & Shelley-Tremblay, 2000).

Although Experiment 2 yielded larger ERP amplitudes compared to Experiment 1, these amplitudes remained relatively modest, being below  $|2|$  (and even  $|1|$ ). Previous studies have shown how changing from active to passive task paradigms influences the ERP amplitude effects. Precisely, P3 effects observed from passive tasks have considerably smaller amplitudes compared to P3 effects from active tasks, and ERP amplitudes reduce further with increasing task difficulty (Polich, 1989, 2007). N400 effects are also shown to be affected by task demands and the attention required (Chwilla et al., 1995; see Kutas & Federmeier, 2011, for a review). For instance, O'Rourke and Holcomb (2002) examined word recognition of both words and pseudowords in English using an active lexical decision task and a passive listening task. N400 amplitudes in the passive task were smaller than in the active task for both words and pseudowords. The authors attributed this pattern to the participants' reduced number of decisions in the passive task and their decreased engagement in post-lexical checking processes. In the current study, exposing participants to a semi-passive task in which they had to sometimes make semanticity judgments on relatively long sentences in Experiment 1 may have increased the difficulty of the task.

It is worth discussing that analyses revealed a rather late time window for both P3 and N400 components compared to previous studies. Prior research has shown P3 effects peaking between 250–500 ms after stimulus onset. However, while in Experiment 2 our P3 effects were found between 300 and 500 ms, a P3 component was found at 300-600 ms in Experiment 1. On the other hand, while the N400 has been typically found to peak between 300–500 ms after stimulus onset, our effect was found at 600–900 ms in

Experiment 1 and roughly between 500 and 900 ms in Experiment 2. This delay in the time window of the components in our experiments compared to previous findings may be due to the complexity of our stimuli and tasks, reflecting delayed cognitive processes. For instance, in the study by O'Rourke and Holcomb (2002), the offset of the N400 for pseudowords was delayed compared to words in both tasks. Authors argued for additional processing of pseudowords compared to words, which is consistent with the idea that the N400 is composed of several sub-components that are associated with distinct cognitive processes (e.g., Kounios & Holcomb, 1994; Nobre & McCarthy, 1994).

An alternate explanation for the delayed onset of our ERP components could stem from the selection of the onset point, which was located in the article rather than the word (e.g., at *la* instead of at *mie*). Shifting the onset point to the word would lead to an average reduction of 82.63 ms (SD = 21.9) for both congruent and incongruent conditions, corresponding to the duration of /l/. Additionally, it would result in an extra reduction of 92.44 ms on average (SD = 17.85) for consonant-initial items, equivalent to the duration of the initial /a/. Moreover, we suspect that the delay of components might reflect the variable word durations, since, as mentioned in the Materials, the sentence fragment that precedes DPs is rather variable. Moreover, the DPs have different durations. Therefore, our ERP results must be interpreted carefully since they may reflect an overlap of cognitive processes.

Despite the material limitations of the present study, we can argue that electrophysiological measures from Experiments 1 and 2 suggest that the acoustic-phonetic information present in homophonic DPs in elision environments is processed by French listeners, even when contextual information is available. This is supported by the results from behavioural measures of Experiment 2. French speakers made very few errors when identifying DPs in congruent contexts. However, errors increased by approximately 44% and decision latencies slowed down by more than 200 ms in the incongruent condition. These findings indicate that lexical retrieval is less efficient when French speakers encounter DPs in elision that are incongruent with the provided context compared to DPs that are congruent. This effect challenges the idea that low-level cues are overruled by high-level cues when both sources of information are available and are displayed under intact listening conditions (Mattys et al., 2005; Mattys & Melhorn, 2007; White et al., 2012).

While context is informational enough to recognise words and achieve comprehension, DPs in elision environments possess acoustic cues that may be harnessed during word recognition, thereby enhancing the successful segmentation of continuous speech. Therefore, both ERP and behavioural measures seem to reflect that, even though top-down cognitive processes are important during sentence processing (as proposed by Mattys & Melhorn, 2007), low-level information is salient enough to be processed (Experiments 1 and 2) and integrated (Experiment 2). Differences between critical DPs were perceived, as evidenced by the P3 effects in both experiments. This aligns with previous findings using the same set of homophonic DPs (Do Carmo-Blanco et al., 2019). Moreover, building on their findings, we have shown that differences between DP pairs undergo further processing during word recognition and lexical retrieval, as manifested by the N400 component in Experiment 2.

This study has provided further evidence on how materials and task design can influence ERP effects during auditory speech perception and speech segmentation. Importantly, the present study is the first one to explore French DPs in elision environments in interaction with contextual information. Our findings showed that the processing of acoustic-phonetic information elicited an early positivity (Experiment 1 and 2), which was followed by a negative component (Experiment 2) associated with the ease or difficulty accessing lexical representations and retrieving the correct meaning. Furthermore, our findings reflect that the hierarchy of cues in sentence processing may be more intricate. Notably, it seems that despite the syllable misalignment produced by elision, the perceptual abilities of French speakers are not compromised during their segmentation.

Our findings provide evidence of the processing of homophonic DPs in the context of elision in French and the interplay of low- and high-level cues. We have provided evidence of the extent to which French listeners exploit acoustic-phonetic cues when they are either favoured or disfavoured by contextual information. The exposure to homophonic DPs has enabled us to disentangle acoustic-phonetic information from semantic information embedded in lexical representations. Importantly, our findings provide evidence for phenomena that may challenge the hierarchy of cues and require to be approached differently during speech processing.



## **CHAPTER 7. The role of F0 trajectories during the segmentation of French in skilled readers vs. adults with dyslexia**

In this study, we explore the role of another acoustic feature of the F0: its temporal dynamics revealed by the value of its slope. As an extension of previous work, our purpose was to examine the role of different F0 characteristics while segmenting running speech in French. Furthermore, we wanted to examine the underexplored research question of whether the phonological deficits experienced by individuals with Developmental Dyslexia impact speech segmentation processes. Few studies have focused on word recognition situations with an emphasis on fine-grained cues or discriminating phonetic representations during speech processing (Virtala et al., 2020). In this study, we used homophonic DPs such as *l'amie* vs. *la mie* (/lami/) and manipulated the acoustic parameters of the F0 to test how French adults (with and without phonological impairments) may use acoustic-phonetic cues to properly segment ambiguous speech when no context is available. Will F0 variations lead to differences in the processing of homophonic DPs for adults with dyslexia as well as for controls?

### **7.1. Introduction**

Several studies have revealed that, particularly in the absence of semantic cues, some fine-grained acoustic information is used during online speech segmentation (K. B. Shatzman & McQueen, 2006; Spinelli et al., 2003). The detection of such cues during speech processing is critical for speech comprehension as it modulates the activation of possible (lexical) candidates. Moreover, segmentation is the basis of language acquisition and a divergent segmentation strategy, based on other cues for example, could be the source of difficulties during language development.

Research centered on the role of the F0 dynamics in speech intelligibility has traditionally contrasted natural speech with speech in which the F0 contours are artificially flattened (Assmann, 1999; Binns & Culling, 2007; Laures & Weismer, 1999). Overall, the flattening of F0 resulted in the detriment of sentence intelligibility. In the study by Laures & Weismer (1999), low-predictability sentences were resynthesised and the F0 contour was flattened to the mean F0 value of the whole sentence. Participants listened to both natural and monotonous sentences presented in a background white noise. Listeners found monotonous sentences less intelligible and made more errors identifying words compared to natural sentences. The authors concluded that flattening F0 contours eliminates

linguistic cues that drive attention to content words and guide listeners during speech comprehension (see also Cutler & Foss, 1977, and Binns & Culling, 2007).

Most of the studies conducted on this topic considered the F0 segmentation cue with its mean value. But natural spoken utterances carry patterns of gliding, in which F0 varies rather continuously. As a consequence, it seems interesting to consider the F0 segmentation cue from its temporal dynamics, i.e. the value of its slope. There is one recent study suggesting the role of the F0 slope value in speech segmentation in Korean. In Korean, the expected intonational pattern is L(HL)H tones. Thus, a word final H tone followed by an initial L tone would signal a word boundary. In an artificial language learning study, Tremblay et al. (2019) varied both the timing of the peak in the final H tones (early vs. late in the last syllable) and the value of the initial L tones (High, Mid, Low). Their results showed that Korean listeners mainly used the difference in pitch value (scaling) between final H and initial L tones (signalling a boundary) while segmenting the speech stream. Raising the value of the initial L tone disrupted segmentation performances. Moreover, in these specific cases where listeners could not rely on this scaling information (i.e. when initial L tones were raised), they seemed to interpret the slope between final H and initial L tones to achieve correct segmentation. When the slope between final H tones and initial L tones of the following words was steeper (as provided by the late peak in the last syllable), the segmentation performances were enhanced, suggesting that listeners used the slope values to locate word boundaries.

Building on these previous studies, our study focuses on the role of the F0 trajectory (slope) as a segmentation cue in French for two different populations: typically developing adult readers with no phonological processing impairments and adults with developmental dyslexia, an impairment that has been linked to deficiencies in speech processing.

Dyslexia is a learning disability that has been widely researched. It is primarily characterised by an impaired ability to learn to read, although the difficulties are not confined to reading. Dyslexia is associated with phonological processing deficits (Blomert, 2011b; Del Tufo & Earle, 2020; Ramus, 2014b; Ramus et al., 2003b; Szenkovits & Ramus, 2005b) that do not only affect the early stages of development. Indeed, research has shown that adults with dyslexia perform poorly on tasks that focus on phonological processing (e.g., Gruber, 2003). Nevertheless, these difficulties cannot be accounted for by low IQ level, poor education, or sensory or neurological deficits.

As suggested by Kujala et al. (2006), phonological impairments in dyslexia specifically target the capacity to effectively establish connections between the written language's code (orthography) and the language's sound structure (phonology). According to this perspective, some research proposes that deficiencies in phonological processing stem from a broader challenge in perceiving acoustic information accurately (Baldeweg et al., 1999; Farmer & Klein, 1995; Kujala et al., 2000; Kujala & Näätänen, 2001; McGivern et al., 1991; Reed, 1989; Tallal, 1980; Wright et al., 2000). This perspective suggests that a more general deficit in acoustic processing hampers the precise perception of essential acoustic components within spoken language, thereby disrupting the formation of an accurate phonological code. Apart from difficulties in distinguishing specific sound types, individuals with dyslexia encounter greater obstacles in tasks that involve intricate perceptual processes as opposed to simpler ones (Banai & Ahissar, 2006; Heiervang et al., 2002). For instance, in the study by Banai & Ahissar (2006), dyslexic individuals demonstrated regular performance in tasks requiring them to determine whether two sounds were identical or different while experiencing impairments in identifying the higher-pitched sound in a pair of two.

As mentioned by Goswami et al. (2002), the phonological deficits in dyslexia may go beyond phonological awareness and phonological representation of words, and underly processes of extraction of suprasegmental information. For example, it has been proposed that a major difficulty in speech perception in dyslexia is the perception of short and rapidly occurring elements (Farmer & Klein, 1995; McArthur & Bishop, 2001; Tallal, 1980; Temple et al., 2000). However, some studies did not report such findings (Adlard & Hazan, 1998; Rosen, 2003; Rosen & Manganari, 2001). Whereas the discrimination of some sound features (e.g., stimulus frequency) is impaired at the early, automatic level in information processing (e.g., Baldeweg et al., 1999; Kujala et al., 2003; Renvall & Hari, 2003), the processing of some other features (e.g., duration) is intact at this level (Kujala et al., 2006). However, when sounds are presented in natural contexts, such as complex tone patterns, it has been observed that the perception of changes in pitch and rhythm was poorer in individuals with dyslexia compared to neurotypical individuals (Kujala et al., 2000, 2003; Kujala et al., 2006; Schulte-Körne et al., 1999). When exposed to such contexts, even discrimination of the duration of long streams became compromised for dyslexics (Kujala et al., 2006). This reveals the multidimensionality of the perception

impairment present in dyslexia for which the specific role of F0 and its trajectory has not been tested yet.

The first aim of the present study is to further examine the role of F0 properties that characterise its dynamics while continuous speech is processed. Namely, we assessed the relative weights of the mean and the slope values of F0, in isolation and combined. As a second aim, we explored whether the phonological deficits experienced by individuals with dyslexia impact the processing of intonational information during speech segmentation.

For dyslexics, while no study has focused on the use of F0, the variations of the speech contour have been tested in several studies using artificial sinusoids. Overall, a deficit in the perception of AM has been found (see Leong & Goswami, 2014 for a review). For example, Leong and Goswami (2014), used processed tone-vocoded sentences to investigate difficulties in processing AM in speech using different rates (stress-only AM, syllable-only AM, and distinct Hz-rate AMs). They showed that dyslexics performed similarly to controls when exposed to AMs with a unique rate. However, when exposed to combined stimuli, the performance of dyslexic participants was poorer, suggesting an observable difficulty when auditory stimuli are complex. Thus, it is still unclear at which extent the phonological processing deficits in adults with dyslexia affect the processing of acoustic parameters of the speech signal.

## 7.2. Hypotheses

We hypothesised that manipulating the F0 acoustic features in our stimuli would modify speech segmentation for the group of typical readers. We expected that increasing the value of the F0 slope of /a/ in *la mie* would increase the perception of vowel-initial items (here, *amie*). Consequently, we expected the typical readers to be able to identify better than chance consonant- and vowel-initial items. However, it was expected that dyslexic adults would exhibit poorer performance in recognising intended targets. If the capacity of individuals with dyslexia to process such acoustic variations as cues for accurate phrase segmentation is compromised, it is expected that their ability to detect F0 contrasts will be as well impaired. As a result, performance of dyslexic participants is expected to not exceed chance levels. Under such circumstances, their responses to the resynthesised conditions should not mirror those of the typical readers (i.e. showing an interaction between group and condition). Conversely, if dyslexic adults adopt the same

segmentation strategy as typical readers, we should observe the same pattern of responses for both groups (i.e. no interaction between group and condition).

Our experiment allowed us to test whether an enhancement of perceiving vowel-initial items is due to an increase in the F0 slope, an increase in the F0 mean value, or an increase in both. Comparing the different experimental conditions allowed us to clarify which F0 characteristics between its slope and its mean value are more important for the segmentation process and whether both are useful. Examining the role of F0 variations during continuous speech segmentation will assess whether the phonological deficits experienced by dyslexic adults may also affect the exploitation of intonational cues, a process relevant during speech segmentation, which is crucial during earlier stages in language development and speech acquisition.

### **7.3. Method**

#### **7.3.1. Participants.**

Ninety-eight participants agreed to participate in this study, all students at the Côte d'Azur University (France) and native French speakers. The students were classified into two groups: those diagnosed with Developmental Dyslexia (DD group, hereafter) and those with normal literacy skills and no history of learning disabilities (Skilled Readers group, SR hereafter). All were retributed after their participation. The study was approved by the local ethics committee of Côte d'Azur University.

DD group included 49 adults (41 females) with a mean age of 21.29 (age range = 18–35, SD = 3). Forty-eight of the DD adults had a certified diagnosis (a formal statement) of dyslexia, the remaining participant showed severe literacy and phonological deficits according to the test battery which was administered to all participants. Moreover, all DDs reported having experienced major difficulties in learning to read when they were children.

The SR group included 49 adults (41 females) with a mean age of 21.37 (age range = 18–35, SD = 3). SR participants served as the control group and were matched to DD participants by chronological age ( $t(48) = -1.16, p = .25$ ), domain of studies ( $t(48) = -1.53, p = .13$ ) (DD: 73.5% social sciences, 26.5% sciences; SR: 83.7% social sciences, 16.3% sciences), and gender<sup>6</sup>. Importantly, the groups differed significantly in mean

---

<sup>6</sup> All participants were matched with gender except 1 female and 1 male from the DD group.

reading score ( $t(48) = -10.71, p < .001$ ) and reading-related skills. Results are given in Table 10 and Table 11.

We focused on adults with dyslexia who are university students. The main motivation for this choice was that, despite poor phonological and reading skills, this population has developed a relatively high level of vocabulary (Cavalli et al., 2016). They are constantly exposed to written material and most of them have likely developed adaptive reading strategies relying on oral comprehension or vocabulary skills that have allowed them to achieve a good level of reading comprehension (Callens et al., 2012, 2014; Cavalli et al., 2016; Olander et al., 2017). This hypothesis is supported by studies reporting deficits in acoustic/phonetic encoding in readers with dyslexia compared to SR despite intact phonological grammar (Berent et al., 2013, 2016).

None (DD and SR) reported visual or hearing problems or brain damage. Participants had no diagnosis of additional learning difficulties. Moreover, all participants had a nonverbal IQ within the normal range (i.e., above the 25th percentile on Raven's Matrices, Raven, 2008, Raven & Court, 1996). However, Raven's scores differed significantly ( $t(48) = -4.13, p < .001$ ), ranging from 48–60 for the DD group, and 53–60 for the SR group. This will be considered in the data analysis.

To confirm the presence of reading and phonological problems in DD participants, we compared performances between groups at standard reading, phonological, language, and cognitive tasks in the first session. To prevent fatigue and cognitive effort to influence their performance on the task, the pretest session and experimental task were run in two different sessions.

### **7.3.2. Cognitive and phonological assessment tests.**

All participants were given 3 reading tests, 1 phonological awareness test, 1 phonological short-memory test, 1 vocabulary test (EVIP: Dunn et al., 1993; the French adaptation of the Peabody Picture Vocabulary Test-Revised, PPVT-R, M. Dunn & Dunn, 1981), and 1 nonverbal test measuring intelligence and abstract reasoning (Raven's Progressive Matrices). Assessment tests on word and pseudoword reading, phonological awareness, and phonological short-memory are described below in this section. They were created by the research group ADUDYS at the Laboratoire de Psychologie Cognitive (LPC) at Aix-Marseille University. Results are shown in Table 11.

#### *Reading test*

Literacy skills of all participants were tested through the standardised French reading test Alouette test (Lefavrais, 1965, 2005), standardised for adults (Cavalli et al., 2018). Reading scores (CTL index) were calculated since they take into account both accuracy and speed (see also Cavalli et al., 2016; Martin et al., 2010, 2014).

#### *Word and Pseudoword reading*

The efficiency of phonological and sublexical decoding was assessed using a list of single real words (e.g., *broche*, “brooch”) and a list of single pseudowords (e.g., *ginve*). These tests are a major indicator of the orthographic processing skills involved in reading. Single-word tests assess the level of word identification independently of the context that may be facilitating, which shows a persistent deficit in single-word reading in adults (Ransby & Swanson, 2003; Swanson & Hsieh, 2009).

Both real and pseudoword tests are timed reading tests, the first lasting up to 1 minute and the latter up to 2 minutes. The score corresponds to the CTL index calculated from the reading time and number of errors. For the real word reading task, 80 bisyllabic words were used, for which written frequency, spoken frequency, regularity, and syllabic complexity of regular words were manipulated. Syllabically complex words are words containing digraphs (e.g., *ai*, *ou*) or contextual spellings, i.e. spellings whose grapheme-phoneme correspondences are not systematic and depend on the context in which they occur (e.g., “g” which is read “j” in front of an /e/ or an /i/). For the pseudoword reading task, 120 pseudowords (60 monosyllabic and 60 bisyllabic) were obtained from real words following the French grapho-phonological rules. See Brèthes et al. (2022) for more detailed information.

Participants were asked to read words written on a sheet as accurately and quickly as possible. In both reading tasks, the following errors were considered: a (pseudo)word read in more than 5 seconds, a transformed (pseudo)word, a syllabified (pseudo)word (i.e. read by unnaturally separating the syllables), an addition, an omission, and a self-correction. All words that had not been read once the time stopped were counted as errors as well.

#### *Phonemic awareness*

Several studies have shown that for languages with a shallow orthography such as French phonemic awareness is still impaired for dyslexic adults (Dufor et al., 2007; Szenkovits & Ramus, 2005a). To evaluate it, we presented participants with a phoneme deletion task.

Twenty-seven monosyllabic French pseudowords of CCV structure (consonant/consonant/vowel) were used following the grapho-phonological French rules (e.g., *glé*). In addition, 3 items were used as training before starting the real task. Participants were instructed to delete the initial sound (i.e. *lé*). Sixteen syllables started with an occlusive and 14 syllables with a fricative. Occlusives (p, t, c, b, d, g) were followed by liquids (l, r), fricatives (s, j), or a nasal (n). The fricatives (f, s, v) were followed by liquids (l, r), occlusives (p, t, c), fricatives (f), or nasals (n). Pseudowords were presented aurally and were not repeated. Both latency times and accuracy responses were collected.

At the end of the test, latency times were totalled for each item and the average was calculated. The time score is the average of the latency times in ms and the total number of correct responses is used to calculate the Performance Score (max = 27).

#### *Phonological short-term memory*

Phonological short-term memory is the system responsible for the temporary storage of verbal information. It has been shown to be impaired in dyslexics which reflects the persistence of the phonological impairment (Dufor et al., 2007; Paulesu et al., 2001). Phonological short-term memory is traditionally assessed through a pseudoword repetition test (e.g., Ahissar et al., 2006; de Carvalho et al., 2014; Cavalli et al., 2019).

For our assessment, a computerised test of the EVALEC Battery was used (Sprenger-Charolles et al., 2005) with a list of 24 pseudowords. They were divided into 4 groups of different syllable lengths ranging from 3 (e.g., *bartino*) to 6 syllables (e.g., *vardotivaruté*) progressively, each including 6 items. Both the time taken to perform the whole task (response time) and accuracy were measured. Participants were asked to repeat each item as accurately as possible, without time constraints. All items are performed independently of failures.

The performance score was the number of correctly repeated pseudowords. Any omission, inversion, or transformation of a phoneme during the pseudoword repetition was considered an error, even if it was subsequently corrected by the participant.



**Table 10.** Means (Standard Deviations) and *t* statistic of DD and SR groups for Chronological age, Reading level, and raw scores on Raven's Matrices.

		<b>DD group (N= 49)</b>	<b>SR group (N= 49)</b>	<b><i>t</i>(48)</b>
Chronological age	years	21.29 (3)	21.37 (2.96)	-1.16
Raven's matrices (max = 60)	raw scores	54.76 (3.5)	57.12 (1.8)	-4.13 ***
Alouette reading test	CTL	334.2 (74.8)	471.17 (48.5)	-10.72 ***

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

**Table 11.** Means (Standard Deviations), *t* statistic, and effect size (Cohen's *d*) of DD and SR groups for performance on Word reading, Phonological skills, and Vocabulary breadth.

		<b>DD group (N= 49)</b>	<b>SR group (N= 49)</b>	<b><i>t</i>(48)</b>	<b>Effect size (Cohen's <i>d</i>)</b>
Word reading (out of 120)	CTL	86.4 (21.4)	134.62 (18.4)	-13.56 ***	-2.42
	errors (%)	7.16 (9.3)	1.86 (1.4)	3.87 ***	0.8
	response time (ms)	52.38 (8)	35.45 (4.8)	14.25 ***	2.58
Pseudoword reading (out of 80)	CTL	77.73 (28.3)	148.37 (31.5)	-11.07 ***	-2.36
	errors (%)	46.12 (20.2)	16.12 (6.7)	9.28 ***	1.99
	response time (ms)	116.8 (9.9)	86.88 (14.9)	11.13 ***	2.37
Phonological awareness (out of 27)	accuracy (%)	76.6 (17.5)	88.84 (15.8)	-4.60 ***	-0.73
	response time (ms)	1881.33 (975.1)	1038.56 (272.2)	5.74 ***	1.18
Phonological Working-Memory (range = 3–6)	span	4.8 (1)	5.55 (0.5)	-4.47 ***	-0.95
	response time (ms)	1.13 (0.2)	0.89 (0.2)	5.61 ***	1.09
EVIP (max = 47)	raw scores	32.96 (5.5)	34.92 (3.7)	-2.27 **	-0.42

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

### 7.3.3. Stimuli.

We used thirty pairs of natural productions of phonemically identical DPs, framed in a neutral context (*c'est*, 'it is'). The members in each pair differed only in the onset of the noun, which was either vowel initial (Vow-nat condition) or consonant initial (Co-nat condition) (e.g., *amie* vs. *mie*). The vowel /a/ of the article in consonant-initial items (e.g., *la mie*) had a consistently lower F0 (M = 183 Hz) and F0 slope (-0.06 Hz/ms) than that of vowel-initial items (e.g., *l'amie*, M = 199 Hz; F0 slope = +0.12 Hz/ms).

Three additional items were created for each pair by resynthesising the F0 of the consonant-initial items. This technique was used to hold timing and spectral features constant while enabling the manipulation of F0 properties of slope and mean. The resynthesis process was performed on the STRAIGHT software (McClelland & Elman, 1986). Both the mean F0 (Hz) and the F0 slope (Hz/ms) values of /a/ were calculated from vowel-initial words (*l'amie*) to modify the F0 value of the first vowel /a/ in each of the 3 new tokens. These parameters were applied to the F0 contour of Co-nat (*la mie*) in the time window belonging to /a/ with 20 ms cosine ramps. The F0 curve of the initial vowel /a/ in Co-nat was (1) multiplied by a scaling factor to reach the same F0 mean value as that in the /a/ in Vow-nat (Co-Shift condition); (2) rotated to reach the slope value of the /a/ in Vow-nat while keeping its F0 value (Co-Slope condition); and (3) both shifted and rotated (Co-Slope+Shift condition). See detailed information about the acoustic measurements of the stimuli in Chapter 5.

Both duration and lexical frequency<sup>7</sup> were balanced between vowel-initial (M duration = 0.8 s ± 0.14; M lexical frequency = 22.21 ± 43.07) and consonant-initial items (M duration = 0.79 s ± 0.15; M lexical frequency = 43.45 ± 178.76). None of the factors showed significant differences between items ( $t(29) = 0.42, p = .68$ , for duration;  $t(29) = -0.62, p = .54$ , for lexical frequency). The number of phonemes and syllables of DPs (determinant + noun) were kept the same between item pairs (e.g., /lami/ or /lafif/). Number of phonemes ranged between 4–10 (M = 6.4 ± 2), and number of syllables between 2–4 (M = 2.8 ± 0.8).

Many studies examining word recognition in dyslexia have typically controlled for phonological (and orthographical) uniqueness points of items (Denis-Noël et al., 2020; J.

---

<sup>7</sup> Word frequencies extracted from *freqfilm2* (spoken word frequency) from the database Lexique3 (New et al., 2001, 2004).

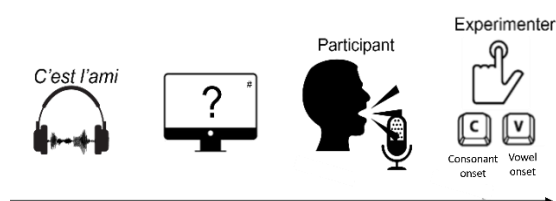
C. Ziegler, Muneaux, et al., 2003; J. C. Ziegler & Ferrand, 1998). However, due to material constraints, DPs could not be matched on the phonological uniqueness point (UP). This will be considered when analysing the results.

Additionally, 30 non-homophonic DPs were selected as a control. Half of them were vowel initial nouns (e.g., *allure*, “appearance”) and half were consonant initial (e.g., *glace*, “ice cream”). They were also included at the end of the same neutral context “*C’est*”. Mean length was  $0.76 \pm 0.10$  and mean lexical frequency was  $36.39 \pm 83.12$ .

To avoid possible bias induced by the experimental design, we constructed five lists including the same number of items per condition (i.e. 6 stimuli per condition). To ensure a balanced presentation of conditions by participant, the 150 stimuli were equally divided into the 5 lists. Lexical frequency was balanced across lists. No significant differences were found for the mean lexical frequency across lists ( $ps > .5$ ). Stimuli presentation within each list was randomised across participants, and lists were counterbalanced.

#### 7.3.4. Procedure.

Stimuli were presented aurally over Sennheiser HD headphones at a comfortable listening level. Each participant listened to only one member of each ambiguous pair from each condition. After that, participants were asked to give a verbal response after every trial and repeat the noun they heard. They were asked to answer according to their first impression. For example, if they heard “*C’est la mie*”, they had to answer “*mie*”. Participants could correct themselves if the article (*l’* or *la*) was also said aloud. Answers were recorded<sup>8</sup>. See Figure 19 for an illustration of the experimental design.



**Figure 19.** Illustration of the task design.

<sup>8</sup> While both answers and response times (RTs) were recorded during the experiment, only answers were included into the analysis, constituting the primary focus of this investigation. This approach was adopted to avoid potential ambiguities in the interpretation of RT data arising from the absence of time constraints. Given the absence of control over the diverse array of factors that could contribute to the variability of RTs within any of the participant groups, RTs were omitted from the data analysis.

Prior research has shown that sound discrimination performance in individuals with dyslexia may be influenced by task complexity (Banai & Ahissar, 2006; Heiervang et al., 2002). Thus, our design aims to minimise both the task's difficulty and working memory load. This verbal repetition task allowed participants to concentrate on the perception of the presented stimuli.

#### 7.4. Data analysis

To assess whether the manipulations of F0 features affect how French listeners segment speech, vowel-initial choices to items were analysed. Answers were coded as 0 if participants gave a consonant-initial response (e.g., *mie*) whereas they were coded as 1 if they answered a vowel-initial item. Results were analysed by comparing matched DD and SR individuals. An analysis of accuracy to natural productions (i.e. Co-nat and Vow-nat) showed that some items raised identification problems for all participants (the accuracy being 0% in both DD and SR groups). Thus, the items *amarre*, *avaleur* and *pareille* were discarded. In addition, *annotation* and *aposition* were discarded, the first word being difficult for the DD group, and the second one for the SR group. Consequently, we included a total of 55 items in further analyses for all participants, DD and SR groups.

A GLMM was performed using the *lme4* R package (Bates et al., 2015), with optimization by quadratic approximation (BOBYQA) with a set maximum of 200,000 iterations. GLMM were preferred over classical ANOVA models as GLMM simultaneously account for individual differences among participants, as well as possible stimuli variations (e.g., Vandermosten et al., 2010). It was also favoured over (multiple) linear regression models to avoid inflated Type I errors (Cohen et al., 2003).

We used maximum likelihood ratio tests to test whether a fixed factor explained a significant amount of the variance in the presence of the other factors and to test the final model with fixed factors against the null model including only the random factors (Faraway, 2016). Group (DD, SR) and Condition (Co-nat, Vow-nat, Co-Shift, Co-Slope, Co-Slope+Shift) were fitted into the model as fixed factors, and intercepts of both Participant and Target were fitted as random factors. We also fitted other variables as covariates into the model such as Lexical frequency, Difference lexical frequency between pairs (i.e. the absolute value of the difference in lexical frequency between homophonic items within each pair), Number of syllables, and Number of phonemes. Yet, the most parsimonious model included the main effects and interaction between Group

and Condition. The model selected was: *vowel responses* ~ *group* \* *condition* + (*1|participant*) + (*1|target*). The model's R-squared adjusted value was 44.66%. Main effects were computed through the *joint\_test()* function from the *emmeans* package (Lenth et al., 2023). Additionally, multiple comparisons were conducted using the *glht* function (Bretz et al., 2010).

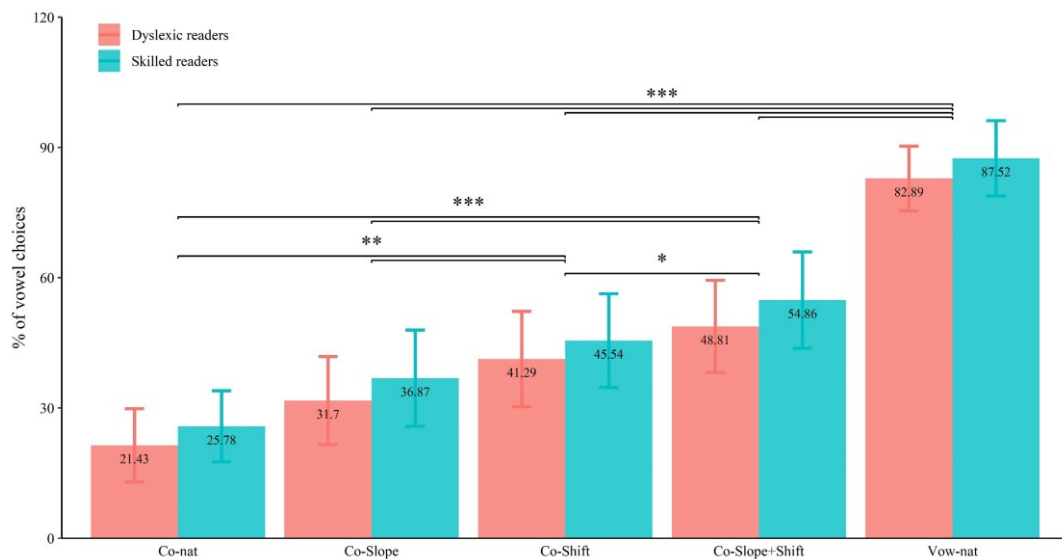
## 7.5. Results

### 7.5.1. Cognitive and phonological tests.

The DD group enrolled in this study performed significantly lower than the group of control participants ( $p < .05$ , see Table 11) on all tests.

### 7.5.2. Response to vowel targets.

All participants had correct responses close to 100% in the unambiguous control conditions (DD group = 98.84%, SR group = 99.25%). This was considered an indicator of good comprehension and performance on the task for both groups. Percentages of vowel-initial choices in the five conditions are presented in Figure 20.



**Figure 20.** Results from DD adults (red bars) and SR groups (blue bars) are shown for percentages of responses to vowel items for all 5 conditions by adults with dyslexia (red bars) and by skilled readers (blue bars). Stars show the significance level ( $p$  values: \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ ) of contrasts revealed by a GLMM.

Correct responses to consonant-initial items reached 74.22% in the SR group and 78.57% in the DD group. Correct responses to vowel items responses reached 87.52% in the SR

group and 82.89% in the DD group. This is in line with previous findings (Cordero et al., 2020; Spinelli et al., 2007, 2010).

A main effect was found for Group ( $F(1, Inf) = 9.558, p < .002$ ) and Condition ( $F(4, Inf) = 74.996, p < .0001$ ). The DD group was less likely to categorise targets as beginning with a vowel compared to SR. However, no interaction between Group and Condition was found ( $p = .97$ ) thus indicating that similar differences are found across conditions between groups. The model's estimated coefficients for each predictor, their standard errors,  $z$  statistic, and  $p$  values are shown in Table 12.

Comparisons across conditions showed that Vow-nat was more often categorised as vowel-initial compared to other conditions (all  $ps < .001$ ) (DD: 82.89%, SR: 87.52%). While no significant difference was found between Co-nat and Co-Slope (DD: 21.43%, SR: 25.78%) ( $p = .15$ ), more vowel-initial responses were given when participants were exposed to Co-Shift (DD: 41.29%, SR: 45.54%) (estimate = -1.1935, SE = 0.3163,  $z$ -value = -3.774,  $p = .001$ ) and Co-Slope+Shift (DD: 48.81%, SR: 54.86%) (estimate = -1.5882, SE = 0.3164,  $z$ -value = -5.019,  $p < .001$ ) when compared to Co-nat.

Crucially, significant differences were also found among the resynthesised conditions. Compared to Co-Slope, more vowel-initial responses were given to Co-Shift (estimate = -0.4774, SE = 0.1426,  $z$ -value = -3.348,  $p = .007$ ) and Co-Slope+Shift (estimate = -0.8721, SE = 0.1425,  $z$ -value = -6.119,  $p < .001$ ). Moreover, the difference between Co-Shift and Co-Slope+Shift was also statistically significant (estimate = -0.3947, SE = 0.1390,  $z$ -value = -2.840,  $p < .03$ ).

**Table 12.** Estimated coefficients from GLMM in function of Group and Condition for vowel responses.

<b>Fixed effects</b>	<b>Estimate</b>	<b>SE</b>	<b><math>z</math> value</b>	<b><math>p</math></b>
<b>(Intercept)</b>	<b>-1.61530</b>	<b>0.25602</b>	<b>-6.309</b>	<b>&lt; .0001</b>
Group [SR]	0.28414	0.21392	1.328	0.1841
Condition [Co-Slope]	0.73517	0.35323	2.081	0.0374
<b>Condition [Co-Shift]</b>	<b>1.22758</b>	<b>0.35046</b>	<b>3.503</b>	<b>&lt; .0001</b>
<b>Condition [Co-Slope+Shift]</b>	<b>1.56944</b>	<b>0.34997</b>	<b>4.484</b>	<b>&lt; .0001</b>
<b>Condition [Vow-nat]</b>	<b>3.53948</b>	<b>0.37236</b>	<b>9.506</b>	<b>&lt; .0001</b>

Group [SR] × Condition [Co-Slope]	-0.03800	0.29575	-0.128	0.8977
Group [SR] × Condition [Co-Shift]	-0.06806	0.29028	-0.234	0.8146
Group [SR] × Condition [Co-Slope+Shift]	0.03761	0.28941	0.130	0.8966
Group [SR] × Condition [Vow-nat]	0.16314	0.34873	0.468	0.6399
Random effects		Variance		SD
Target		1.03740		1.0185
Participant		0.09579		0.3095

*Note:* 2604 observations, 98 participants, 55 items. Bold indicates statistically significant coefficients.

These results provide further evidence that listeners can categorise homophonically identical DPs better than chance. This is in line with previous work indicating that listeners exploit acoustic features associated with word boundaries to correctly segment speech (Gow, 2001; Quené, 1992; K. B. Shatzman & McQueen, 2006). Moreover, these results show the crucial role of the mean F0 value and its slope (its trajectory) for correct segmentation. The lack of interaction between groups and conditions suggests that dyslexic adults make similar use of the F0 cues to segmentation than do typical readers. However, the DD group displays a lower tendency to categorise items as being vowel-initial, even in their natural productions.

### 7.5.3. Correlations with phonological UP.

To evaluate the extent to which responses to intended targets may depend on the UP of targets, we performed a correlation analysis between these factors. UP of targets ranged from 4 to 9, meaning that the earlier UP from targets was at the 4<sup>th</sup> phoneme, while targets with the later UP were recognised at the 9<sup>th</sup> phoneme.

Results showed a strong positive correlation that was significant for both groups ( $r_{DD}(4) = .96, p = .002$ ;  $r_{SR}(4) = .96, p = .002$ ). Results showed facilitatory effects during DPs' identification for those items possessing a later UP. Even though participants had no time constraints to respond, the question arises whether the impact of this factor may also correlate with decision-making processes. To investigate this possibility, we performed

the same correlation with reaction times instead. Interestingly, there was a strong negative correlation between RTs and UP significant for both groups ( $r_{DD}(4) = -.87, p = .02$ ;  $r_{SR}(4) = -.81, p = .05$ ).

UP modulated both responses and time latencies of participants, regardless of the group. However, our results are not consistent with previous research on the processing of UP during word recognition since previous work shows that items with earlier UPs increase correct responses and decrease RTs (Radeau et al., 1989; Samuel, 1987). It has been proposed that even high frequent words may require the auditory system to process the entire speech segment and the subsequent word boundary (if relevant to the recognition of the target) in order to be recognised (Luce, 1986b). Facilitatory effects of delayed UPs may reflect the difficulty in recognising our homophonic DPs in the absence of contextual information.

## 7.6. Discussion

We examined the influence of F0 characteristics in the segmentation of homophonic DPs in French. We explored the extent to which F0 trajectory and F0 mean influence the perception and segmentation of targets. Exposing listeners to identical homophonic phrases such as /selami/ allowed us to examine the differences in perception of variations purely at the level of prosody. The pairs of phrases varied only in their intonational information: duration, F0, and formants; while their structure, as well as their phonotactic and contextual information, were kept the same. Our design allowed us to test to what extent an increase in the F0 slope, in the F0 mean, or an increase in both may modulate the perception of vowel-initial targets. Additionally, this study provided evidence for the first time of the extent to which the F0 trajectory can influence the perception of French adults with DD. Overall, comparisons across conditions allowed us to clarify which of the manipulated F0 characteristics may be more important during segmentation and whether they are useful for word recognition processes in French.

All three resynthesised conditions lead to an increased segmentation of vowel-initial nouns (e.g., *l'amie*). Results reflected a gradual increase of vowel choices with respect to the F0 properties of targets, regardless of the group of participants. Crucially, when the F0 slope of /a/ in *la mie* was replaced by the F0 slope from *l'amie* while its mean value was kept the same, choices of initial-vowel items increased by 10.7% on average (from 21.4% to 31.7% for the DD group, and from 25.8% to 36.9% for the SR group). This



suggests that this subtle slope change ( $M = +0.18$  Hz/ms) may be used during speech segmentation processes in French. Additionally, a greater increase in vowel-initial responses was found when the F0 mean value was higher (DD: 41.3%, SR: 45.5%). Interestingly, the condition bearing a manipulation of the two cues (Co-Slope+Shift) had higher levels of vowel segmentation (DD: 48.8%, SR: 54.9%) than those conditions bearing only one-cue manipulation. However, it did not reach the degree of segmentation of vowel-initial natural productions, which was even higher (DD: 82.9%, SR: 87.5%). As hypothesised, results from typical readers are in line with previous work (Spinelli et al., 2010). These findings support the idea that even though low-level cues are processed by the listeners' auditory system, other cues are involved during segmentation and disambiguation processes, which ultimately lead to the correct understanding of speech (Spinelli et al., 2003).

The fact that results from unimpaired listeners are consistent with previous findings builds on the existing evidence of the exploitation of the acoustic-phonetic cues present in elided stimuli. Most importantly, it demonstrates the robustness of such cues under various methodological approaches. Unlike previous studies, our task has no time constraints, which can consequently affect decision-making processes. Despite this methodological difference, we found similar mean percentages across conditions for typical French readers, supporting previous evidence of stimuli robustness (Do Carmo-Blanco et al., 2019; Spinelli et al., 2007, 2010).

A main effect of group is found showing that dyslexics perceived on average fewer items as vowel initial compared to the SR group. However, no interaction is observed between the conditions and the group suggesting that both dyslexics and typical readers have similar segmentation strategies. The same gradual increase in vowel responses was found showing a beneficial use of the F0 properties. Moreover, this similarity effect between groups is also observable in the variability of performance within groups –i.e. the standard deviations–, which is similar for both groups across conditions.

It is important to note that our participants were not matched on IQ and a significant difference was found between groups. Even though the IQ level of individuals does not explain performance differences between dyslexics and controls, several studies have suggested the importance of controlling for non-verbal IQ levels (e.g., Goswami et al., 2010; Leong et al., 2011). Therefore, the same analysis was performed on a subset of participants with matching IQ (33 matching participant pairs, instead of the 49 pairs).

Interestingly, results from matching participants showed similar performance and no significant interaction Group  $\times$  Condition was found significant ( $ps > .9$ ). Our findings suggest that both the mean and the slope of the F0 are exploited during speech segmentation in French, even for adults with developmental dyslexia. We can argue that individuals with dyslexia may not use F0 cues differently from typical readers, which suggests an aspect of speech perception that may remain less affected in dyslexia.

While detection of pitch and rhythm changes has been found to be poorer for dyslexics compared to neurotypicals (Kujala et al., 2000, 2003; Kujala et al., 2006; Schulte-Körne et al., 1999), in our study we did not find significant differences in the use of F0 cues between DD and SR groups. The difference between previous work and our findings can be due to the material used. Kujala et al. (2006) employed fully synthesised pseudowords while we used natural recordings (vowel initial and consonant initial items) that were also used to generate three additional intermediate conditions. For intermediate conditions, only the F0 parameters of the pivotal /a/ were resynthesised. Thus, variations within and between our stimuli may be greater compared to stimuli used in Kujala et al. (2006) or other studies using tones instead of stretches of speech (e.g., Kujala et al., 2000, 2003).

Our findings exhibit similar effects to those of Leong & Goswami (2014), who manipulated the number of cues present in the signal. They showed that while dyslexics are impaired when cues are combined, they obtained similar performance to that of controls when only one cue was manipulated. In their experiment, the authors used processed tone-vocoded sentences of nursery rhymes, such as *Simple Simon met a pieman*, to investigate difficulties in processing amplitude modulations (AM) at onset in speech (i.e. the change in intensity of the signal or rise time). Participants discriminated sentences with either individual AM (stress-only AM, syllable-only AM, and sub-beat AM) or combined AMs. The sensitivity to AM patterns of dyslexic participants was poorer than those of controls when exposed to combined stimuli. However, dyslexics performed similarly to controls when exposed to AMs with a unique rate. The authors argued that greater modulations of prosodic patterns present in their sentences (from real speech) may have led to similar performances between groups. This could explain why no differences are found between groups in our study. Our homophonic phrases may exhibit complex AM patterns due to variations in the F0 patterns within individual DPs, which may be intertwined with AM from the preceding speech fragment (*C'est*).

In an additional analysis including the (phonological) UP factor, we observed that both DD and SR groups benefited from speech stretches with later UPs during the identification of intended DPs. This effect was observed through both an increased segmentation of intended targets (i.e. vowel-initial items such as *amie*) and shorter response times. Targets with UP at the 9<sup>th</sup> phoneme elicited more vowel responses than those with UP at the 4<sup>th</sup>, with the response rate declining from 60% to 37% for DD and from 68.4% to 35% for SR, respectively. Furthermore, response latency times decreased progressively, with both groups showing a decrease of more than 350 ms when identifying targets with UP at the 8<sup>th</sup> compared to those with UP at the 4<sup>th</sup>. On average, the response latency times decreased from 1131.6 ms to 1491 ms for DD and from 679.9 ms to 1139.6 ms for SR. In line with previous research (Radeau et al., 1989; Samuel, 1987), our findings support the notion that UP affects both low-level (perceptual) and high-level (decision-making) stages of speech processing. However, in our study later UPs facilitated word recognition, rather than early UPs. This may be attributed to the material employed. The contrast between our DP pairs is based on the pivotal vowel /a/, which appears either at the beginning of nouns or within the article "la". Previous research has suggested that information from vowels allows for a larger pool of possible word candidates compared to consonants (Cutler et al., 2000), which appears to be independent of phoneme inventory and vowel distinctiveness in a given language. Thus, the recognition of our stimuli may have been delayed due to their homophonic nature and the fact that the vowel carries crucial information needed for target recognition and disambiguation.

Crucially, we observed that dyslexics seem to have a bias toward consonant answers compared to controls. While the percentage of consonant-initial responses given in the SR group was 48.7%, the percentage given by the DD group was 54%. A t-test comparison between the overall responses of both groups showed that this difference was significant ( $F(2, 2524) = -2.67, p = .008$ ). Lexical frequency did not affect performance, regardless of the group. One may think that if F0 and duration differences are found between consonant- and vowel-initial items in the pivotal /a/, consonants that follow (e.g., /lami/ or /lafɪ/) may also carry different acoustic information, affecting the listeners' perception. Indeed, we found differences in both acoustic measures. Consonants were longer at the word onset ( $M = 119.1$  ms) than in vowel-initial items ( $M = 98.8$  ms) ( $t(29) = -5.12, p = .0001$ ). Moreover, F0 was lower for consonant-initial than vowel-initial items

( $t(27^9) = 8.52, p = .0001$ ). Therefore, the bias toward consonants may be due to the saliency of cues that mark word onsets in our homophonic DPs, both in the vowel /a/ and consonants.

Another way to look at this bias toward consonants would be to link it to the idea that consonants and vowels have different status and that consonants are relatively more informative for lexical distinctions than vowels (Caramazza et al., 2000; Hochmann et al., 2011; Peña et al., 2002; Toro et al., 2008; see Nazzi & Cutler, 2019 for a review). Listeners are able to segment words through consonants but not through vowels, which are used to extract structural regularities. This has been shown in studies using artificial language paradigms (Newport & Aslin, 2004) as well as real, continuous speech (Bonatti et al., 2005). Importantly, the differing role between consonants and vowels has been proven to exist regardless of the saliency of their acoustic differences (Toro et al., 2008). In the particular case of French, a consonant advantage is shown during lexical processing already at the age of 16-month-old infants (Havy & Nazzi, 2009). This persists later in time; a greater weight to consonants than to vowels is given in adulthood during lexical activation and access processes (Havy et al., 2014). The consonant bias found in our task for dyslexics may be due to the relevant role of consonants during lexical processing and word recognition, which may guide these individuals during the identification of homophonic phrases.

While we do not have a clear interpretation for this bias, we can alternatively relate this effect to the neuroscientific oscillation theory postulating that the phonological deficit in dyslexia found its origin in the atypical functioning of the left auditory cortex and that dyslexic brain samples auditory stimuli faster than typical readers (Lehongre et al., 2011, 2013). Indeed, some theoretical models propose that neural oscillations in the 25–35 Hz range could be the basic speech sampling rate and that this timing is linked to the phonemic temporal format (approximately 25–30 ms). In this framework, neuroimaging studies have identified, for dyslexics, a deficit in oscillatory activity in the low-gamma band (30-Hz) (Di Liberto et al., 2018; Lallier et al., 2017; Lizarazu et al., 2015; Marchesotti et al., 2020; Van Hirtum et al., 2019), as well as abnormally strong responses at higher frequencies (around 40 Hz), suggesting that auditory sampling could be faster than in typical readers (Lehongre et al., 2011, 2013). This deficit in oscillatory activity has already been linked to difficulties in the processing of rise time in AM sounds

---

<sup>9</sup> Note that the degrees of freedom are not the same in duration and F0 measures since two of the consonants [t] had no F0 values.

(Goswami, 2011, 2019; Van Hirtum et al., 2019). Therefore, it could be that this oversampling of speech by dyslexic brains give an advantage in our task to shorter events, as consonants, over longer ones such as vowels.

In summary, our findings contribute to a growing body of research that provides contradictory evidence on a deficient perception of short and brief acoustic elements (Blomert et al., 2004; Blomert & Mitterer, 2004; Bonte & Blomert, 2004; Hazan et al., 2009; Serniclaes et al., 2001). Particularly, the consonant bias found for dyslexics contrary to controls may indicate that dyslexics are more inclined to rely on higher-level cues (such as lexical or semantic information) than low-level cues (such as acoustic or phonetic) to navigate language, which is consistent with the idea that adult dyslexics rely more on context than controls (e.g., Blomert et al., 2004; Gross-Glenn et al., 1990). Alternatively, even if the lexical frequency of words in our experiment is matched, the fragments *la* vs. *l'a* are not. The non-elided article (*la*) is more frequent than the elided one (*l'*) (frequency of 14946 vs. 8129 fppm). Moreover, the probability that the vowel /a/ follows *l'* is 8.3% but the probability of having a consonant-initial noun after *la* is 51.1% (information extracted from the *Lexique* database).

We have exposed French listeners to ambiguous stretches of natural speech, where the F0 contours of consonant initial items (lower for both mean and slope values) were contrasted with those of vowel initial items (higher for both values). This provides further evidence of the role of F0 variations in speech perception in natural setups. Moreover, our findings exhibit that F0 contour variations are detected even in optimal acoustic conditions, which contrasts previous studies including background noise (Binns & Culling, 2007; Cutler & Foss, 1977; Laures & Weismer, 1999). Furthermore, our study offers insights into the importance of considering the F0 dynamics during speech segmentation and word recognition processes. Precisely, we have provided the first evidence of the participation of both F0 slope and F0 mean values in segmentation processes for both neurotypical and adults with DD.

The present study, thus, extends existing evidence of the role of intonational cues in spoken word segmentation in French. Our findings contribute to the idea that listeners exploit acoustic features that are detected at word limits to correctly segment speech (Gow, 2001; Quené, 1992; Shatzman & McQueen, 2006). Our study provides further evidence on the particular role of slower AM of speech, which has been proposed to be key for segmenting speech, such as recognising smaller units of rimes and syllables

(Goswami et al., 2002; Jusczyk et al., 1999; Muneaux et al., 2004; Pasquini et al., 2007). We have demonstrated the role of perceiving small acoustic contrasts and AM during the process of (real) spoken speech segmentation in French, which has facilitated participants in recognising ambiguous words. Importantly, our findings seem to reveal a language-driven modulation of cues, highlighting the importance of the material in testing dyslexic participants.

## CHAPTER 8. The interplay of L1 and L2 segmentation strategies in French speech segmentation

In this study, we examined the degree of sensitivity of French learners to the acoustic cues of homophonic DPs such as /lami/. To account for the role of these cues in segmenting speech by L2 speakers, participants were presented with an AX same-different discrimination task (see Figure 21). Specifically, we tested English and Spanish speakers who use low-level cues to speech segmentation that differ from those of French speakers.

### 8.1. Introduction

Language-specific perceptual strategies are employed by listeners to determine the location of word boundaries, as evidenced by research in numerous languages (see Altenberg, 2005; Ito & Strange, 2009, for English; Cutler et al., 2007; Quené, 1992a, for Dutch; Sanders et al., 2002, for Japanese; Dupoux et al., 1997; Peperkamp et al., 2010, for French; Soto et al., 2007, for Spanish). These segmentation strategies shape the listener's perceptual system very early during language development, laying the foundation for acquiring an L1.

Several prosodic patterns have been found to guide L1 French speakers during speech segmentation. According to a body of research, the syllable plays an important role in segmentation in French (Content et al., 2001; Dumay et al., 2002; Mehler et al., 1981). Dumay et al. (1998) showed that French speakers were significantly faster at identifying *lac* when syllable onsets were aligned, as in the non-word ZUN.LAC (where /nl/ is not a possible syllable onset in French), compared to cases where *lac* was not aligned to the syllable onset, as in ZU.GLAC (where /gl/ is a possible onset in French, which can misalign syllable and word boundaries).

Even though French has no lexical stress, it exhibits phrasal prominence. The prosodic structure in French is marked by the AP (Jun & Fougeron, 2000, 2002; Welby, 2003, 2006). The intonational pattern in French follows a two-rise configuration characterised by a LHLH sequence. The final non-reduced syllable of the last word within the AP is characterised by an F0 rise and it is lengthened, but this occurs exclusively in positions that are not at the end of the utterance. Moreover, an early F0 rise is occasionally observed close to the beginning of the AP. However, other studies have demonstrated that French speakers are sensitive to simpler variations of the two-rise L1H1L2H2 pattern during

speech segmentation. Specifically, the first early rise (L1H1) and the simple F0 inflection unaccompanied by a subsequent rise seem to serve as cues to content-word onsets (Spinelli et al., 2010; Welby, 2003, 2007).

Empirical support for the use of intonational information during the segmentation of French is evident in a study conducted by Banel and Bacri (1994). The authors exposed French speakers to ambiguous phrases like /ba.gaz/ (*bagage* “luggage” or *bas gage* “low pledge”). Listeners were more likely to perceive a single word (*bagage*) when the second syllable was lengthened and two words (*bas gage*) when the first syllable was lengthened. This observation aligns with the assumption that a syllable positioned at the end of a phrase experiences lengthening and that phrase boundaries are not expected to occur in the middle of a word.

The present study aims to explore the extent to which L1 French speakers and learners of French are sensitive to fine-grained acoustic information during the segmentation of phonemically identical DPs due to elision (e.g., *l'amie* vs. *la mie*). The discrimination between these homophonic DPs in French has been previously tested through an ABX discrimination task and an identification task. Spinelli et al. (2007) showed that when French speakers are presented with homophonic phrases such as /lami/, they can correctly discriminate between the two items at 66.31% (e.g., *l'amie* vs *la mie*). In a follow-up experiment, the authors presented only one of the items, and listeners were asked to identify it among the two possible phrases. On average, native listeners were able to correctly identify the target in 75.48% of the cases. The authors argued that fine-grained acoustic details present in homophonic DPs in elision environments are robust enough to be used by L1 French speakers in the absence of higher-level cues, such as contextual information. Yet, it has not been studied whether such acoustic-phonetic information can guide French learners during L2 speech segmentation.

It is commonly assumed that segmenting an L2 presents particular challenges. For example, when individuals process an unfamiliar or artificial language, they use the patterns of their L1 and apply them to unfamiliar stimuli in order to achieve segmentation (Cutler et al., 1986; Sanders & Neville, 2000; Vroomen et al., 1998). Therefore, it seems that speech segmentation is less efficient and more effortful for L2 than L1 speakers (Carroll, 2001; Shoemaker, 2010).



Becoming a proficient L2 listener depends not only on the availability of lexical representations in the L2 (Sanders et al., 2002a; Sanders & Neville, 2003a; Tremblay, 2011), but also on the control of segmentation strategies (Dupoux et al., 2001; Peperkamp et al., 2010; Tremblay et al., 2012; Tremblay & Spinelli, 2014). It is argued that languages with similar acoustic-phonetic patterns will facilitate the transfer of segmentation strategies to apply those of the L1 in L2 processing (e.g., Kim et al., 2008; Murty et al., 2007). Under these circumstances, L1 cues seem to benefit L2 speech segmentation. However, if the patterns diverge, the use of prosodic cues may be affected by potential interferences between the languages. In such cases, L2 perception can be constrained by L1 segmentation routines, such as phonotactics (e.g., Weber, 2001) and prosody (e.g., Cutler et al., 1989; Dupoux et al., 1997). For instance, a series of studies conducted by Dupoux et al. (1997, 2001, 2008) revealed that differences in the use of prosodic patterns of L1 French speakers interfere with and impede the correct processing of L2 Spanish. There is thus still considerable uncertainty regarding how listeners use L1 and L2 cues during L2 speech segmentation. Moreover, these findings show that achieving L2 segmentation relies on the linguistic characteristics of both L1 and L2, which implies that the outcome of the process of speech segmentation differs depending on the languages involved.

English and Spanish possess different prosodic structures and vary in the acoustic cues associated with word-boundary location. For example, stress is contrastive in English (i.e. it carries lexical information, such as in “impact”: IMPact vs. imPACT; Fry, 1958). However, stress assignment in English is variable and depends on a number of factors. As discussed by Cutler & Carter (1987) and Zeng et al. (2022), L1 speakers of English tend to exhibit a trochaic bias in word segmentation, which is linked to stress patterns. Particularly, disyllabic nouns have the main stress on the first syllable (Dabouis et al., 2020; Wells, 2008). This has been evidenced in studies where disyllabic words were segmented faster and more accurately when the stress was on the initial syllable (Cutler & Norris, 1988a). Remarkably, research suggests that word stress provides English speakers with a reliable cue to locate boundaries in noun-initial positions (Mattys, 2004; McQueen et al., 1994).

Stress in English is conveyed through multiple acoustic-phonetic cues. These cues include F0, intensity, vowel quality, and duration (Altenberg, 2005; Chrabaszcz et al., 2014; Fry, 1958, 1965; Ito & Strange, 2009; Sluijter & Van Heuven, 1996; Q. Wang, 2008; Y. Zhang

& Francis, 2010). For example, stress assignment predicts vowel quality. Specifically, full vowels typically occur under stress in English (Cutler & Carter, 1987). Moreover, the F0 is aligned to the stressed syllable within a given word, but it adjusts to the overall sentence intonation, potentially resulting in the alignment of F0 to a different word (Gordon, 2011).

Like English, stress is also a contrastive feature in Spanish (e.g., Dupoux et al., 1997). For example, *BE.be* means “(s/he) drinks” whereas *be.BE* means “baby” (Sebastian Galles et al., 1992). Moreover, it is marked by multiple cues, as in English. Specifically, the combination of F0, duration, and intensity seems to indicate syllable stress to Spanish speakers (e.g., Estebas, 2008). Llisterri et al. (2003) found that an association between higher F0 values with longer syllable durations or with higher intensities guides Spanish listeners during speech segmentation, whereas a single cue does not seem to be a sufficiently robust source.

Stress has a rather predictable pattern. Stressed syllables occur at any of the final three syllables of a word (Navarro, 1966; Peperkamp et al., 2010; Sebastian Galles et al., 1992). Namely, stress on nouns tends to fall on the penultimate syllable in vowel-final nouns (e.g. *MEsa* “table”), while stress falls on the final syllable in consonant-final nouns (e.g. *aniMAL*) (Baković, 2016). However, stress in Spanish does not align with the boundary edges of words (Soto et al., 2007; Soto-Faraco et al., 2001), and the rising pattern of F0 seems to be misaligned with the accented syllable (Ortega-Llebaria & Prieto, 2011; Prieto et al., 1996).

When confronted with an L2, both English and Spanish speakers seem to exploit their L1 strategies to segment L2 speech (e.g., Chen et al., 2017; Dupoux et al., 2008, 2010; Fox et al., 1995; Kondaurova & Francis, 2008; Peperkamp et al., 2010; Peperkamp & Dupoux, 2002; Skoruppa et al., 2009; Tyler & Cutler, 2009). For instance, in an artificial language learning study, English speakers were found to rely on duration and F0 to locate word boundaries (Tyler & Cutler, 2009). These speakers identified higher F0 values as word onsets and they relied on longer durations in final position to segment speech. Moreover, L1 English speakers used duration, F0, and vowel quality when discriminating between Mandarin Chinese tones of varying durations and pitch contours for low and high vowels ([a] and [i]) (Chen et al., 2017). Using a recall task, Peperkamp et al. (2010) exposed L1 Spanish to non-words differing in stress location (*NUmi* vs. *nuMI*), with differences in either a combination of duration, F0, and intensity or a combination lacking duration.

Spanish speakers made few errors in identifying non-words that differed in stress location, which was enhanced when duration was present as a cue compared to when it was absent.

There are several studies exploring the use of cues in an L2 by L1 English speakers and several studies have focused on how stress differences influence the perception of L2 speech for Spanish speakers. However, to our knowledge, no studies have directly examined to what extent acoustic-phonetic cues used by L1 Spanish speakers, such as F0 and duration, are used during the segmentation of an L2. In this study, we explored how French learners and L1 speakers segment homophonic French phrases. Table 13 summarises the main differences in the use of prosodic cues used in the L1 of French, English, and Spanish speakers.

**Table 13.** Prosodic strategies used in the L1 groups chosen (French, English, and Spanish).

	<b>French</b>	<b>English</b>	<b>Spanish</b>
Stress contrastiveness	-	+	+
Duration	-	+	+
F0	-	+	+
Intensity	-	+	+
Vowel quality	-	+	-

## 8.2. The present study

Over the last decade, an extensive body of research has compared the use of prosodic cues during L1 and L2 speech processing through nonce words (Dupoux et al., 2008; Lin et al., 2014; Lukyanchenko et al., 2011). Yet, few studies have used natural speech and real words to investigate the processing of prosody in both L1 and L2. The present study aims to examine natural speech segmentation. To this end, we included homophonic DPs in elision environments, which are associated with different lexical representations and meanings, but they only possess variations at the acoustic level. Importantly, an elided version of the definite article is not present in any of the languages of the L2 groups in this study (English and Spanish). This means that the strategies of these L2 listeners may have been modified to adapt their L1 segmental strategies to the L2 features.

This study aimed to tap into low-level acoustic processing to examine the differences in sensitivity between natives and non-natives to the acoustic features of homophonic phrases such as /lami/. It is therefore paramount to consider possible methodological biases, as performance on perceptual tasks can be altered by a myriad of factors. Among them, several studies have pointed out that the type of task plays an important role (e.g., Dupoux et al., 1997, 2001; Levy & Strange, 2008; Pater, 2003). AX and ABX are two of the most common discrimination tasks in speech perception studies. ABX requires listeners to judge the tokens' closest similarity to presented items (i.e. whether X is closer to A or B). On the other hand, in AX tasks, listeners make same/different judgments (i.e. X is the same token as A or a different one). The first task captures the perception of phonetic categories (phonological perception) whereas the latter is designed to capture surface acoustic-phonetic perception (Han, 2009; Levy & Strange, 2008; Werker & Logan, 1985). Moreover, AX tasks pose less memory and processing demands than ABX tasks, which facilitates acoustic processing (Boomershine et al., 2008; Dupoux et al., 2008; Werker & Logan, 1985). Thus, we considered the AX discrimination task as the most suitable approach for this study to examine the exploitation of acoustic-phonetic details as segmentation cues by both French learners and native speakers.

### **8.3. Aims and hypotheses**

In the present study, we explored whether the level of discrimination to homophonic DPs achieved in an ABX task may also result from other discrimination procedures. Particularly, we aimed to implement a simpler, less memory-demanding discrimination task, shown to be more sensitive to acoustic differences between speech stimuli: the AX task. The first aim of the study was to replicate and extend the findings of Spinelli et al. (2007), who reported that phonemically identical DPs in elision environments in French can be correctly discriminated by L1 speakers (at 66.31%). Therefore, we expected that L1 French speakers would be able to discriminate between consonant- and vowel-initial items in an AX task by exploiting the acoustic-phonetic variations.

Our second aim was to investigate the extent to which French learners use acoustic-phonetic information in the absence of other cues, such as context, when segmenting homophonic DPs in French. Due to their reliance on both F0 and duration during speech segmentation, we expected that both English and Spanish speakers would be able to discriminate between items. Moreover, since the acoustic attributes of the pivotal /a/ differ between DPs, we expected differences in the discrimination abilities between

English and Spanish speakers. Notably, we expected that English speakers would exhibit greater sensitivity in discriminating between items for two main reasons. First, stress in bisyllabic words is assigned to the first syllable. This contrasts French speakers, who lack stress, and Spanish speakers, for whom bisyllabic nouns have varying stress patterns depending on whether a vowel or a consonant is in its final position. Second, vowel quality is a salient feature of word recognition in English, but it is not in French and Spanish.

Additionally, language aptitude is known to play an important role in the success of learning an L2 (Grigorenko et al., 2000). Several cognitive abilities, including phonemic coding ability, grammatical sensitivity, inductive language learning ability, and associative memory, have been identified as being crucial for L2 learning (e.g., Carroll, 1973). Therefore, as a third aim, we will examine whether the language learning abilities of L2 individuals influence their ability to perceive differences between pairs and their overall performance in the task. These abilities were assessed using the LLAMA battery test (Meara, 2005; Meara & Rogers, 2019). Given the requirements of our task, the capacity to recognise novel words (referred to as phonemic coding ability) is considered especially relevant to perceiving the acoustic features of French homophonic DPs. We expected that L2 speakers with higher scores in this ability assessment would exhibit superior discrimination performance.

## **8.4. Method**

### **8.4.1. Participants.**

A total of 53 participants were recruited for this study. These included 31 native French speakers and 22 learners of French. Both groups reported no psychological, auditory, or language deficits at the time of testing. None of the participants spoke more than one language at home, and none were involved in professional activities including intensive sports or music training.

The L1 group comprised 31 native French speakers (28 women) between the ages of 18 and 38 ( $M = 20.23$ ,  $SD = 3.5$ ). All participants had been schooled in France and were pursuing a university degree at Côte d'Azur University in France at the time of the experiment. According to their self-reported data, 58.07% of the participants had English as an L2 and indicated that they rarely used (see the questionnaire in Annex V).

The L2 group consisted of 22 learners of French (13 women) between the ages of 18 and 41 ( $M = 26.95$ ,  $SD = 6.11$ ). They were all students at the Côte d'Azur Université, in Nice, France. Eleven participants had English as their L1, and another eleven participants had Spanish as their L1, as indicated in the biographical questionnaire. Participants completed a language background questionnaire that contained relevant biographical information (see Annex VI). For the L2 learners, this information included their age of first exposure to French, the number of years of instruction in/on French, the number of months spent in a French-speaking environment, and the percentage of weekly use of French. A summary of this information is provided in Table 14.

**Table 14.** Biographic and language background information of the two groups of learners of French (Spanish, English, and the mean of both).

	English group N=11	Spanish group N=11	Mean N=22
<i>Biographical data</i>			
Mean age	28.45 (6.67)	25.45 (5.37)	26.75 (5.78)
Highest level of formal education:			
PhD	36.36%	0%	18.18%
Master	27.27%	36.36%	31.82%
Bachelor	27.27%	36.36%	31.82%
High school	9.09%	27.27%	18.18%
<i>Language background</i>			
Mean age of first instruction	9.23 (5.73)	14.91 (6.46)	12.07 (6.63)
Mean length of formal instruction in French	10.64 (3.88)	4 (2.78)	7.32 (4.73)
Mean length of residence in a French-speaking country	4.94 (5.03) <sup>(1)</sup> years	3.40 (3.47) <sup>(2)</sup> years	4.17 (4.29) years
Learned French in school	90.91%	72.73%	81.82%
Learned French in university	9.09%	27.27%	18.18%
Mean percentage of formal instruction in French	41.82%	54.09%	47.95%
Mean percentage of informal instruction in French	58.18%	45.91%	52.05%

<sup>(1)</sup> 18.18% (2 participants) lived for less than a year.

<sup>(2)</sup> 45.45% (5 part.) lived for less than a year.

Independent-sample t-tests were conducted to compare the groups on various language background variables. Results revealed no significant differences between groups for the years of residence in a French-speaking country ( $p = .42$ ) and the percentage of formal or informal education ( $ps > .2$ ). However, there were significant differences in the age of exposure to French and the duration of formal education received in French. English speakers began learning French at a younger age (M age = 9.2) than Spanish speakers (M age = 14.9) ( $t(20) = -2.18, p = .041$ ). Furthermore, English speakers had received longer formal education in French (M years = 10.6) than Spanish speakers (M years = 4) ( $t(20) = 4.61, p = .0002$ ). These differences could have repercussions on L2 proficiency, which is why a test battery was used to objectively assess the participants' language level and abilities.

#### **8.4.2. Language Measures.**

To avoid subjective assessment of L2 French proficiency, the L2 groups were administered a battery of tests to control for their proficiency in French, as well as their working memory (WM) capacity and language aptitude (see Table 16).

*Working memory (WM) capacity, phonetic implicit memory, and language aptitude* were assessed through the LLAMA tests (Meara & Rogers, 2019), comprising a test battery of four different tasks based on the standardised MLAT tests (Carroll & Sapon, 1959). The LLAMA test has been widely used to measure the language aptitude of L2 learners (Granena, 2013; Huang et al., 2021; Rogers et al., 2016; Suzuki, 2021). The four tasks in the test battery include a vocabulary learning task (B3), an implicit learning task (D3), a sound-symbol correspondence task (E3), and a grammatical inferencing task (F3).

*General vocabulary skills and French proficiency* were measured by a written lexical decision test adapted to French: The LEXical Test for Advanced Learners of FRENch (LexTALE\_FR; Brysbaert, 2013; Lemhöfer & Broersma, 2012). D-prime scores, obtained through signal detection analysis, and accuracy were measured according to the authors' analyses (Brysbaert, 2013) to ensure that results were above chance level. Accuracy was used to evaluate participants' performance in greater detail (see Table 16).

*Self-estimated proficiency skills.* Participants were asked to provide information regarding the languages they had been exposed to before the moment of testing and to indicate the percentage of formal and informal education they received for each language.

Additionally, participants rated their proficiency level for each language on a scale of 1 (very low) to 10 (very high). The mean self-reported proficiency level for French was relatively high ( $7.41 \pm 1.97$ ), with 16 participants reporting high proficiency levels (ranging between 8 and 10), 11 participants reporting intermediate proficiency levels (5-7), and 2 reporting low proficiency levels (1-4).

#### 8.4.3. Materials.

Thirty pairs of phonemically identical DPs in French (e.g., *la fiche* – *l’affiche*) were selected (*ambiguous condition*). See the Annex III for the complete stimuli list (similar to that in Spinelli et al., 2007). As a control, 30 pairs of non-phonemically identical DPs such as *le bain* (“the bath”) – *le pain* (“the bread”) were selected (*unambiguous condition*). The two members of unambiguous pairs differed only in voicing or place of articulation corresponding to French minimal pairs. For example, in *le bain* – *le pain*, both [b] and [p] are bilabial plosives but the first is voiced whereas the latter is voiceless.

All 60 DPs were recorded in carrier sentences by a female native speaker of French who was naïve to the goals of the study. To provide two tokens for each sequence, every experimental sentence was recorded twice (e.g., two productions of *la fiche*). An example of the recorded sentences is given below (3). Experimental sequences of definite article + noun (e.g., *la fiche*) were excised from the carrier sentences.

(3) a. *C’est la fiche qui me manque.* ‘It’s the poster that I’m missing.’

b. *C’est l’affiche qui me manque.* ‘It’s the sheet that I’m missing.’

#### 8.4.4. Acoustic analyses.

We performed an acoustic analysis of the recorded stimuli (e.g., *la mie* vs. *l’amie*), comparing duration, F0, and formant of the stimuli’s first two segments. The main differences between pairs of items are described below. See more detailed information about the acoustic measurements in Chapter 5.

##### *Duration*

Durational differences were found between vowel-initial (*l’amie*) and consonant-initial (*la mie*) members of the homophonic DPs. Mean duration of the first syllable [la] of vowel-initial items was longer (157.69 ms) than that of consonant-initial items (138.68 ms) ( $F(1, 29) = 4.57, p < .0001$ ). Precisely, both the onset phoneme [l] and the vowel [a] were longer in content-initial word positions compared to those in function words



(definite article). Moreover, second syllables of consonant-initial items (e.g., [fi] in [lafɪ]), were longer (213.29 ms) than vowel-initial items (186.91 ms) ( $F(1, 29) = -2.93$ ,  $p = .007$ ).

#### *Formants*

Vowel-initial items had higher values for both F1 (M = 720.08 Hz) and F2 (M = 1439.64 Hz) compared to formants of consonant-initial items (M = 708.23 Hz and M = 1409.2 Hz, respectively). However, differences were not significant ( $p = .5$  and  $p = .71$ , respectively).

F1 and F2 showed significant differences for the second vowels of DPs. Analyses showed that F1 was significantly lower (M = 454.43 Hz) for high-mid word-medial vowels in vowel-initial content words (e.g., *l'affiche*) than for low-mid word-medial vowels in both vowel-initial content words (e.g., *l'amarre*) (M = 626.67 Hz) ( $F(1, 27) = -2.66$ ,  $p = .013$ ) and consonant-initial words (e.g., *la marre*) (M = 658.2 Hz) ( $F(1, 28) = -3.36$ ,  $p = .002$ ). F2 of front word-medial vowels in vowel-initial content words (e.g., *l'amarre*) were significantly higher (M = 1598.24 Hz) than those of back word-medial vowels (e.g., *l'allocation*) (M = 1219.14 Hz) ( $F(1, 16) = 2.23$ ,  $p = .04$ ). Additionally, they were also significantly higher than those in consonant-initial content words (e.g., *la locution*) (M = 1185.49 Hz) ( $F(1, 16) = 2.42$ ,  $p = .028$ ).

#### *Fundamental frequency*

F0 of the initial phoneme [l] in elided items was higher (191.39 Hz) compared to the [l] in non-elided items (183.46 Hz). This difference was marginally significant ( $F(1, 29) = 1.97$ ,  $p = .059$ ). Similarly, the F0 of the initial [a] of elided items was higher (200.52 Hz) than the initial [a] of non-elided items (181.99 Hz) ( $F(1, 29) = 8.76$ ,  $p < .0001$ ).

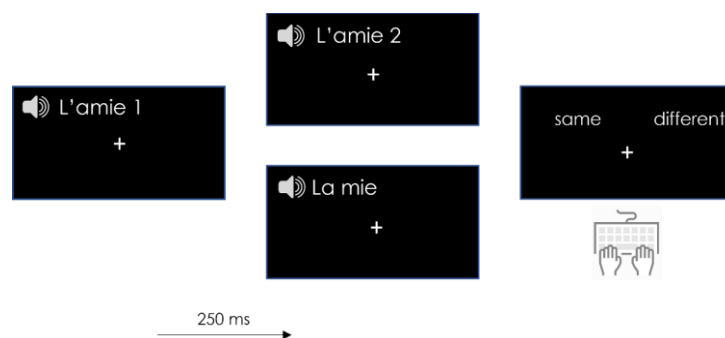
#### **8.4.5. Procedure.**

After participants had filled out the consent form and the biographical questionnaire, the task started. Participants were tested individually in a quiet room. Stimuli from both ambiguous and unambiguous sets were presented to listeners for a discrimination task using an AX paradigm. Stimuli were presented binaurally over Sennheiser HD 212Pro headphones at a comfortable listening level. Participants were informed that they would hear two sequences, one after the other, and were asked to decide whether what the speaker said in the two sequences (i.e. the content) was the same or different without

expecting an exact match, as each phrase can be inherently distinct. Participants were asked to respond quickly, but not so quickly as to sacrifice accuracy.

During the AX task, an item was presented first (A) followed by a second item (X). The X item of each trial was either another token of the same word (*same* condition) or a different word (*different* condition). For both ambiguous and unambiguous sets, participants were presented with all stimuli combinations, and the presentation of same and different conditions was counterbalanced. Stimuli presentation were divided into two blocks. For example, *l'amie1* was presented with *l'amie2* (*same, vowel1-vowel2* condition), and *la mie* with *l'amie* (*different, consonant-vowel* condition) in the first block. In the second block, *l'amie* was presented with *la mie* with (*different, vowel-consonant* condition), and *la mie1* with *la mie2* (*same, consonant1-consonant2* condition). The order of block presentation was counterbalanced as well, and the order of item presentation was randomised within each block. Stimuli were displayed only auditorily, and no feedback was given in either the training or the experimental blocks.

A pause was included between blocks. Each experimental trial consisted of one pair of stimuli separated by an inter-stimulus interval of 250 ms. Individual trials were separated by a 550 ms pause. All participants started with a training block, which contained examples of ambiguity due to elision such as *l'amer* 'the bitter one' vs. *la mère* 'the mother'. The experiment was controlled by PsychoPy Software (version 2021.2.3).



**Figure 21.** Illustration of the experimental design.

Finally, once L2 participants had completed the task, they completed the LexTALE test and the LLAMA test battery. The session lasted approximately 60 minutes for L2 participants and 20 minutes for L1 participants.

## 8.5. Data analysis

Three L1 participants and one pair of stimuli of the unambiguous set were excluded from analyses due to technical problems. Moreover, two French speakers and one English speaker were removed from the analyses due to poorer performance on the unambiguous set, which raised concerns about the reliability of their task performance. The final number of participants was 26 in the L1 group, 10 in the English group, and 11 in the Spanish group. Trials on which participants took longer than 2500 ms were discarded. This resulted in the elimination of 2.16% across trials from the analyses. The final total number of trials was 5398 in the ambiguous set and 5104 in the unambiguous set.

To examine the influence of the individual L1 languages (French, English, and Spanish), their performance on the AX discrimination task was analysed using d-prime ( $d'$ ) scores.  $D'$  is a sensitivity measure based on signal detection theory. Previous studies have shown that it is a more accurate measure for discrimination task scores given that bias of participants' responses is considered in the analysis (e.g., Macmillan & Creelman, 2005).  $D'$  scores were calculated as the distance between two distributions: signal and signal+noise. The resultant  $d'$  corresponds to the transformed z-value of the hit rates minus that of false alarm rates (*psycho*; Makowski, 2018). All analyses were performed in R (version 4.0.3).

In the present experiment, if  $d' = 0$ , participants' performance was at chance level. A  $d' > 0$  can be interpreted as participants being able to discriminate between vowel-initial and consonant-initial homophones. We followed a similar analysis in Shoemaker (2010). However, we preferred to use LMM instead of ANOVA for reasons previously stated (see Chapter 7). Mixed models were performed for the statistical analyses of  $d'$  scores. We performed LMM by Maximum Likelihood using the *lmerTest* R package (Kuznetsova et al., 2020). Only responses for the different pairs are used in the statistical analyses that follow. As fixed effects, we entered Group (French, English, Spanish) and Pair (vowel-consonant, consonant-vowel) (with interaction term) into the model. As random effects, we had intercepts for participants and items. The model's formula used was:  $dprime \sim group * pair + (1|participant) + (1|item)$ . The model's estimated coefficients for each predictor, their standard errors,  $t$  statistic, and  $p$  values are shown in Table 15. Moreover, Pearson correlations were calculated to determine whether language learning skills and WM abilities affected the discrimination sensitivity of L2 speakers.

**Table 15.** Estimated coefficients from LMM for mean  $d'$  scores across groups.

Predictor	Estimate	SE	$t$ value	$p$
(Intercept)	<b>0.4086736</b>	<b>0.0708159</b>	<b>5.771</b>	<b>&lt; .0001</b>
Group [English]	<b>0.2759009</b>	<b>0.0212904</b>	<b>12.959</b>	<b>&lt; .0001</b>
Group [Spanish]	<b>0.0598848</b>	<b>0.0198383</b>	<b>3.019</b>	<b>0.003</b>
Pair [vowel-consonant]	0.0007697	0.0132260	0.058	0.95
Group [English] × Pair [vowel-consonant]	-0.0035595	0.0254978	-0.140	0.89
Group [Spanish] × Pair [vowel-consonant]	0.0026302	0.0242039	0.109	0.91
Random effects	Variance		SD	
Item	4.068e-05		0.006378	
Participant	0.1476		0.384202	

Note: 5398 observations, 47 participants, 60 items. Bold indicates statistically significant coefficients.

## 8.6. Results

### 8.6.1. Language measures.

Results of the LLAMA and proficiency tests are presented in Table 16.

**Table 16.** LLAMA tests mean percent scores, vocabulary skills, and self-assessment of French proficiency for the two groups of French learners (Spanish, English, and the mean of both).

	English group N=11	Spanish group N=11	Mean N=22
<i>LLAMA tests</i>			
Learning new words ( $B3$ )	30%	50%	40%
Listening to new words ( $D3$ )	51.82%	57.27%	54.55%
Phoneme-grapheme associations ( $E3$ )	35.91%	45.91%	40.91%
Grammar ( $F3$ )	54.09%	50%	52.05%
Mean score	42.96%	50.8%	46.88%
<i>French proficiency (LexTALE)</i>			
$D'$ score	1.3013 (0.70)	1.3452 (0.67)	1.3232 (0.67)

Accuracy (0-1)	0.7175 (0.11)	0.7054 (0.10)	0.7114 (0.12)
Self-reported proficiency in French	7.55 (2.02) <sup>(2)</sup>	7.18 (1.94) <sup>(1)</sup>	7.36 (1.94)

<sup>(1)</sup> 63.64% (7 part.) reported a high proficiency level (>7) and 36.36% (4 part.) an intermediate proficiency level (5-7).

<sup>(2)</sup> 81.82% (9 part.) reported a high proficiency level (>7) and 18.18% (2 part.) a low proficiency level (1-4).

A one-way ANOVA revealed significant differences across tasks for English speakers. They were better at recognizing new words (M score = 51.8%) than at learning them (M score = 30%) ( $F(3, 39) = -3.08, p = .013$ ). This was also easier than learning new phoneme-grapheme associations (M score = 35.9%) ( $F(3, 39) = -2.88, p = .018$ ). Additionally, learning a new grammar (M score = 54.1%) was easier than learning new words ( $F(3, 39) = -3.58, p = .006$ ), which was also easier than learning new phoneme-grapheme associations ( $F(3, 39) = 2.66, p = .026$ ).

In contrast, few differences were observed in the scores of Spanish speakers in the LLAMA task battery. Lower scores were found in the phoneme-grapheme association task (45.9%) and higher scores were shown in the new-word recognition task (57.3%). However, results did not show any significant differences across tasks ( $ps > .3$ ), indicating similar performance.

Importantly, independent-sample t-tests showed no significant differences between the two groups for most of the LLAMA tests ( $ps > .02$ ), except for one task, the B3 task. Spanish speakers were significantly better at learning new words than English speakers (M score = 50% vs. 30%, respectively) ( $t(20) = -3.01, p = .007$ ). However, there were no significant differences between the groups in the overall LLAMA test score ( $p = .16$ ).

Moreover, despite the significant differences found between groups in terms of age of exposure to French and length of formal education in the language, scores on the French level test (LexTale) were not significantly different between groups ( $p = 2.09$ ). Overall, results indicate that the two groups were overall comparable in terms of their language skills and proficiency level in French.

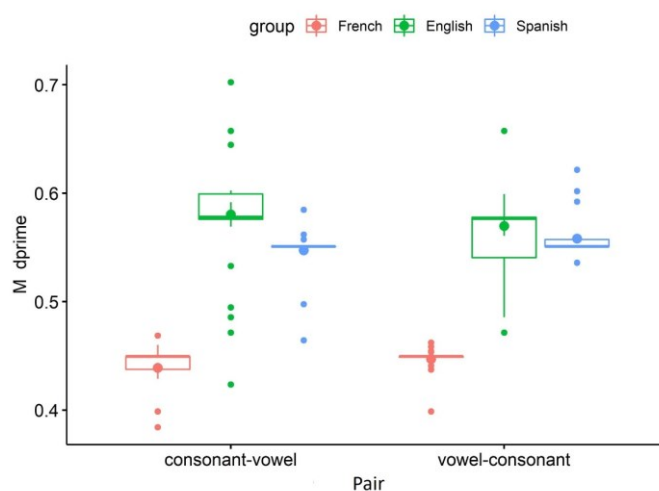
### 8.6.2. Sensitivity measures to discrimination.

Average accuracy rates in the unambiguous condition were remarkably high for all three participant groups. Specifically, the French group achieved an accuracy rate of 95.95%, the English group achieved 94.2%, and the Spanish group achieved 93.3%. However, the

mean percentage of correct responses for the same pairs was close to the chance level across the groups. The French group correctly responded at 54.59%, the English group at 56.47%, and the Spanish group at 54.71%.

Higher  $d'$  scores were found in the unambiguous condition (M L1:  $3.69 \pm 0.76$ , M L2:  $3.27 \pm 0.54$ ) compared to the ambiguous condition across groups (M L1:  $0.44 \pm 0.36$ , M L2:  $0.57 \pm 0.46$ ), which was considered an indicator of good comprehension and performance on the task. However, analyses were performed on the ambiguous set only.

An effect was observed for Group ( $p < .0001$ ), but no effects were found for Pair ( $p > .97$ ). No significant interaction was found for Group  $\times$  Pair ( $p > .98$ ). Results revealed that  $d'$  scores of French listeners were lower (M =  $0.44 \pm 0.36$ ) compared to those of English (M =  $0.58 \pm 0.59$ ) and Spanish (M =  $0.55 \pm 0.3$ ) ( $ps < .0001$ ) speakers. Interestingly, although small, significant differences were found between the L2 groups. Namely, English speakers had higher  $d'$  scores than Spanish speakers (estimate = 0.21, SE = 0.02,  $t$ -value = 9.18,  $p < .0001$ ). Mean  $d'$  scores across groups are shown in Figure 22 for consonant-vowel and vowel-consonant pairs.



**Figure 22.** Mean  $d'$  scores for consonant–vowel and vowel–consonant pairs in the discrimination task for each group: French (red), English (green), and Spanish (blue). Left boxplots show scores to consonant–vowel pairs, right boxplots to vowel–consonant pairs. Error bars indicate SDs.

Significant differences were found across groups, with learners displaying greater sensitivity to homophonic DPs in elision environments than L1 speakers. Interestingly, English speakers had the highest scores in the task. The lack of a significant interaction

Group  $\times$  Pair suggests that the listeners' discrimination performance was similar when exposed to vowel-initial or consonant-initial item presentations.

### **8.6.3. Correlation analyses: battery of language tasks and language proficiency.**

In order to explore whether language learning skills had an effect on the performance in discriminating two homophonic DPs in an L2, correlation analyses were performed between each language task (LLAMA test and LexTale) and  $d'$ -prime scores in the discrimination task for both L2 groups (English, Spanish) separately. Moreover, self-reported proficiency levels were also correlated to the  $d'$ -prime scores.

No significant correlations were found between any of the factors ( $ps > .8$  for English speakers and  $ps > .7$  for Spanish speakers). Of special interest to this study is the D3 task, which assesses the ability to recognise short stretches of speech (that were either new or repeated). The capacity to recognise speech patterns is considered an important skill in language ability (Service & Kohonen, 1995; Speciale et al., 2004). This ability helps the speaker recognise words more easily and, thus, acquire vocabulary (Rogers et al., 2016). However, no significant correlation was found between this task and  $d'$  scores for any of the groups ( $ps > .9$ ). Finally, the language background factors shown in Table 16 were individually correlated with  $d'$  scores of participants. However, none of the correlations were significant in both groups ( $ps > .7$ ).

These results suggest that the L2 listeners' sensitivity to the acoustic features present in our homophonic DPs was not affected by their WM capacity, language aptitude, proficiency in French, or their exposure to the language.

## **8.7. Discussion and conclusion**

Word and syllable boundaries may sometimes become ambiguous due to the phonological process of elision. For instance, the vowel in a definite article may be resyllabified across the word boundary to become the onset of the following syllable (e.g., *l'amie* vs. *la mie*). The misalignment between word and syllable boundaries could hinder the process of spoken word recognition since multiple lexical candidates may be consistent with the input /lami/. Duration, F1, F2, and F0 values showed significant differences between vowel-initial and consonant-initial items. Particularly, the initial syllable /la/ in vowel-initial items was longer and had a higher F0 rise compared to consonant-initial items. We, therefore, tested the sensitivity to these acoustic cues present

in homophonic items in elision environments and compared segmentation processes of L1 French speakers and French learners through an AX discrimination task.

Similar to Spinelli et al. (2007), our findings revealed that French listeners were able to discriminate between the homophonic DPs ( $d' > 0$ ). Interestingly, results from sensitivity measures ( $d'$ ) revealed better performance of L2 speakers than that of L1 speakers, which confirms our hypothesis. This is in agreement with research showing a perceptual advantage of L2 compared to L1 speakers (Chang, 2016; Chang & Mishler, 2012; Choi et al., 2019; Kim & Tremblay, 2021). For example, Kim & Tremblay (2021) explored the use of intonational cues of both Gyeongsang-Korean and Seoul-Korean learners of English to lexical stress contrasts in English. Contrary to English, Gyeongsang-Korean and Seoul-Korean have no lexical stress. However, Gyeongsang-Korean has lexical pitch accents. Through a recall task, participants were presented with four-item sequences differing either in stress placement (suprasegmental, e.g., *TRUS*ty vs. *trus*TEE) or phonemically (segmental, e.g., *taller* vs. *caller*). Moreover, stress was either marked by F0, duration, and intensity (natural condition) or by F0 only. Gyeongsang-Korean learners outperformed both Seoul-Korean learners and L1 English speakers in the suprasegmental condition when F0 was the only cue. Since F0 is a cue to pitch accents in Gyeongsang-Korean, the authors argued for a transfer of suprasegmental cues from L1 to L2 speech. They concluded that listeners are more likely to employ L1 cues to process L2 speech when cues have more relevant roles (or have greater functional weight), which is also discussed by Chang (2018).

In the present study, the differing cues to stress in the speakers' L1 may potentially explain group differences in our results. Stress is contrastive and is marked by multiple sources of information in both English (Altenberg, 2005; Chrabaszcz et al., 2014) and Spanish (Dupoux et al., 1997; Estebas, 2008). Specifically, for both English and Spanish, duration and F0 values are two of the cues associated with stress. However, these properties contrast with the French language. First, French lacks stress, and, second, duration and F0 values are cues associated with word segmentation (Peperkamp et al., 2010; Welby, 2007). The fact that the learners were more sensitive to acoustic differences than L1 French suggests that L2 speakers benefited from their L1 cues to encode L2 words. Although our experiment does not allow us to unravel differences in the weighting of each of the acoustic cues, the learners' greater sensitivity during the discrimination of



ambiguous DPs suggests a transfer of suprasegmental cues from their L1 to the L2. This is in line with Tremblay et al. (2018), who argued for a transfer of strategies from L1 to L2 during segmentation, even if the strategies served separate functions.

As predicted,  $d'$  scores reflected that English speakers were better at discriminating between DPs compared to the other groups. English speakers outperformed both French and Spanish speakers. These effects are in line with previous studies showing that stress in English is typically located in the initial syllable in disyllabic nouns (Cutler & Carter, 1987; Dabouis et al., 2020; Wells, 2008; Zeng et al., 2022). In contrast, French lacks stress but has been found word-final AP (Dupoux et al., 2001; Welby, 2003). On the other hand, in Spanish, the stress position is typically found on the penultimate syllable (Baković, 2016; Dupoux et al., 1997; Harris, 1983; Peperkamp et al., 2010) and the rising pattern of F0 seems to be misaligned with the stressed syllable (Ortega-Llebaria & Prieto, 2011; Prieto et al., 1996). Furthermore, contrary to French and Spanish speakers, vowel quality has been shown to play a crucial role in segmentation for English speakers (Cutler & Van Donselaar, 2001; Tremblay et al., 2021; Zhang & Francis, 2010). Hence, our findings indicate that language-dependent differences in suprasegmental cues seem to bias speakers' sensitivity during speech segmentation.

The lower performance of French speakers compared to L2 speakers can be explained by the particularity of intonational patterns in French. While French has no lexical stress, it displays phrasal prominence. This prominence is marked by an F0 rise and lengthening at the final non-reduced syllable of the last word within the AP in an utterance. Moreover, short phrases in isolation have found to display differing intonational patterns. While lengthening has been shown to indicate final-word edges (Banel and Bacri, 1994; Rietveld, 1980), F0 rises often mark the onset of content words (Di Cristo, 2000; Spinelli et al., 2010; Vaissière, 1997; Welby, 2003, 2007). The first syllable of vowel-initial phrases in our experiment (e.g., *l'amie*) had both longer durations and higher F0 values. Hence, it is likely that French participants did not rely as much on the acoustic cues of the DPs as they were presented in isolation. Alternatively, their lower performance may also be the result of the lexical competition between items, i.e. the use of a segmentation strategy based on higher-level information (lexical information). French speakers may recognise the words they hear, while English and Spanish speakers may perhaps focus solely on the acoustic cues. Therefore, the level of processing may differ between L1 speakers and learners in this task. It is likely that for L1 speakers there might be multiple

levels of processing involved (low and high levels), while learners may only engage in a single level of processing (low level).

It is important to note that, even though our results show scores above chance level,  $d'$  scores were rather low (the larger the  $d'$  value, the greater the difference in perception between the two items –Greenaway, 2017–). Our results show  $d'$  scores lower than 1 across groups. This is consistent with the results from Shoemaker (2014). Through an AX task, the author tested the sensitivity of native French speakers to durational differences of homophonic phrases in situations of liaison, such as in *un air* vs. *un nerf* (both produced as /œ̃.ɛʁ/). Their results revealed  $d'$  scores below 1, but they increased when the duration between items was exaggerated. The authors concluded that acoustic differences associated with the items may not be a robust enough cue in natural speech. Our results evidence the possible difficulty of L1 and L2 speakers to process acoustic differences of homophonic items in the context of elision.

We assume that, although the AX task reduces both memory and cognitive load, response choices based on item comparisons (as in ABX tasks) may be simpler. Unlike the AX task, ABX tasks involve presenting participants with two possible items. Listeners can then compare the items and focus on at least one of them to differentiate their attributes and match the X item with the (most) accurate sample (A or B). This may eliminate some of the response bias problems, as discussed in McGuire (2010), which could account for the discrepancies between our results and those of Spinelli et al. (2007).

Our material may further increase the difficulty of word segmentation since the pivotal phoneme is not a consonant as in the liaison situations in Shoemaker (2014) (e.g., /n/), but a vowel (/a/). Segment predictability in elided DPs such as /lami/ is unlikely, since the elision occurs in the article, leaving the vowel in a confusing position: either it is part of the definite article or it is the onset of the content word. It is just as common for vowels to appear at the end of syllables as at the beginning of them, and this may complicate the task of discriminating between items in elision. This conclusion is supported by the differing roles in speech processing of consonants and vowels and the idea that lexical identification is better achieved from consonants than from vowels (Caramazza et al., 2000; Cole et al., 1996; Nazzi & Cutler, 2019; Owren & Cardillo, 2006).

Our findings reflect the additional challenge that L1 and L2 French speakers may face while perceiving homophonic DPs in elision. Our findings build on the idea that even

though the segmentation of elided items in French is associated with distinctive acoustic information, these are subtle acoustic details that may not be robust enough for native or non-native speakers to fully disambiguate items in isolation (Spinelli et al., 2003, 2007). Furthermore, this suggests that, in the absence of higher-level information such as context, unintended words may still be activated by the individual's perceptual system (Shoemaker & Birdsong, 2008; Spinelli et al., 2003).

Finally, the correlation analyses revealed no significant effects between discrimination performance and scores from the language test battery in L2 groups. Although some of the tasks were more challenging for the English speakers, language aptitudes were not significantly correlated with the results, regardless of the group. The absence of significant correlations can be explained by the paradigm used in this study. Specifically, Granena (2014) included measures that drew participants' attention to language forms. Particularly, they allowed sufficient time to reflect on language correctness and structure. However, our AX task was not designed to assess language correctness, but rather it aimed at measuring the sensitivity to acoustic-phonetic differences and was done in a rather short amount of time. Therefore, the differences present in our material may be too subtle to be linked to the individual language skills.

The acquisition of non-contrastive information in L1 and L2 has not been addressed by many studies (see Darcy et al., 2007 and Shoemaker, 2010, for some examples). The present study contributed to the existing evidence on the use of non-contrastive acoustic cues associated with elision in French. Our findings provide insight into the impact of the availability of cues in continuous speech and its exploitation during segmentation by both L1 and L2 speakers. Our findings contribute to the existing evidence that both English and Spanish speakers rely on their L1 strategies to segment L2 speech (e.g., Chen et al., 2017; Dupoux et al., 2008, 2010; Fox et al., 1995; Kondaurova & Francis, 2008; Peperkamp et al., 2010; Peperkamp & Dupoux, 2002; Skoruppa et al., 2009; Tyler & Cutler, 2009). Specifically, English speakers rely on word-initial positions to locate word boundaries in their L1 (Mattys, 2004; McQueen et al., 1994), but also in L2 speech segmentation (Chen et al., 2017; Tyler & Cutler, 2009). These speakers use duration, F0, and vowel quality when discriminating between French DPs of varying durations, formants, and pitch contours (Chen et al., 2017). Furthermore, Spanish speakers seem to benefit from the use of multiple cues to signal stress in their L1 to encode L2 words (Peperkamp et al., 2010).

We have provided first-time evidence for the transfer of cues of L2 speakers when processing homophonic DPs in elision environments in French. Our study has revealed that Spanish and English learners of French are more sensitive than L1 French speakers to the nuanced acoustic features of homophonic DPs in elision. Furthermore, these findings provide evidence of the acquisition of the phonological representation of the elision phenomenon in an L2 by both English and Spanish learners.

## **CHAPTER 9. General discussion and conclusion**

The main goals of this dissertation were threefold: 1) investigate the interplay between low-level cues (acoustics) and high-level cues (context) while segmenting speech in French, 2) examine the particular role of the F0 mean and F0 trajectory as acoustic cues during speech segmentation and word recognition in French, and observe the extent to which it may be deficient for individuals with phonological processing impairments (Developmental Dyslexia), and finally 3) explore the sensitivity of French learners to acoustic cues during speech segmentation and word recognition. To accomplish this, we used ambiguous DPs in French that were homophonically identical due to elision.

These goals were approached through the experiments in Chapters 6, 7, and 8. The first section outlines the major findings of the studies. We then address the discussion on the neural correlates to the segmentation of DPs in elision environments. We then focus on the discussion of the specific role of the dynamicity of the F0 during speech segmentation. This is followed by the integration of the exploitation of fine-grained acoustic detail within the current models of word recognition. We then discuss the implications of deficient phonological processing in the detection of acoustic cues in elision environments. Next, the role of L1 and L2 strategies is discussed, followed by the final concluding remarks. At the end of this chapter, the limitations of the current studies are discussed, as well as future research directions.

### **9.1. Recapitulation of main findings**

To address the first goal of this dissertation, two studies were conducted using the EEG technique in Chapter 6. A semi-passive sentence judgement task (Experiment 1) and a word identification task (Experiment 2) were implemented bearing differences in task demands, particularly on the focused attention drawn to the items of interest. Results indicated that while task demands and differences in the object of attention do not impact the detection of acoustic-phonetic information, they affect later cognitive processes involved during segmentation, such as word retrieval and meaning disambiguation. Specifically, when attention was drawn to the sentential information (Experiment 1), context outweighed acoustic cues. However, the inhibition of possible lexical candidates seems to be less efficient for French listeners when attention is drawn to the lexical forms (Experiment 2).

The second aim of this dissertation was investigated through a word repetition task (Experiment 3) in Chapter 7, which tested the use of features characterising the F0 during the segmentation of homophonic DPs in elision environments. Specifically, the F0 mean and slope of the initial /a/ were manipulated to explore the relative importance of these F0 parameters during segmentation and word recognition processes. The results revealed a gradual importance of cues. Specifically, more targets were perceived as vowel initial by increasing the F0 of the pivotal /a/, but this effect was improved by increasing its mean, and further fostered with the increase of both cues simultaneously. The same gradual pattern of responses was found in dyslexic individuals, suggesting a similar use of cues. However, dyslexics were less sensitive to manipulations and had thus less tendency to categorise targets as vowel initial compared to controls, which reflects a reduced ability to exploit F0 cues to the same extent as typical readers.

Finally, the third goal of this dissertation was approached in Chapter 8 through an AX (same-different) discrimination task (Experiment 4) including natural productions of DPs in elision environments. We tested L1 French speakers and both English and Spanish learners of French. Results showed that French speakers had poorer performance in discriminating between items compared to both groups of learners. Moreover, English speakers showed a higher sensitivity to cues compared to the other speakers. Differences in the weight of acoustic-phonetic cues during L1 segmentation affect L2 encoding, revealing a transfer of cues from L1 to L2.

## **9.2. Neural correlates of acoustic-phonetic detection**

Acoustic analyses revealed differences between our critical DPs mainly on the onset of the phrases (/l/, /a/, and the first two adjacent phonemes: the first consonant and vowel). Particularly, /la/ was longer and had a higher mean F0 for vowel initial DPs than consonant initial ones. These differences are associated with acoustic features that are not contrastive in French and yet both ERP and behavioural results reveal that French speakers are attuned to such differences, even when the context is also available. Findings from our Experiments 1 and 2 reveal that French speakers were able to exploit fine-grained non-contrastive information from DPs (revealed by P3 ERP effects) and modulating lexical retrieval (revealed by N400 ERP effects in Experiment 2). This suggests that both elided (*l'amie*) and non-elided (*la mie*) items are perceived differently by French listeners. Even in the presence of context, French speakers are sensitive to and rely on acoustic-phonetic properties of items. The ability of French listeners to exploit

non-contrastive acoustic information aligns with the notion that coarticulation (such as variations in pitch or duration) plays a significant role in speech segmentation even in adulthood. Our findings therefore contribute to evidence showing that the absence of coarticulation is associated with word boundaries rather than within-word variations (Browman & Goldstein, 1992; Fougeron & Keating, 1997; Vaissière, 1988). Moreover, our effects further contribute to the idea that non-contrastive acoustic-phonetic information in phrases within elision may be stored in the mental lexicon, guiding speakers to segment the intended items (Do Carmo-Blanco et al., 2019).

The relatively small ERP amplitudes elicited by the critical DPs in Experiments 1 and 2 for both P3 and N400 components suggest an impact of the experimental design and task demands, which has been widely reported (e.g., Bentin et al., 1993; Erlbeck et al., 2014; Kotchoubey & Pavlov, 2019; Kutas & Federmeier, 2011; Perrin & García-Larrea, 2003). Previous research has demonstrated that ERP amplitudes are susceptible to various semantic factors, such as the degree of word expectancy within a given context (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Lau et al., 2008), regardless of whether the context consists of a single word, a phrase, or a complex utterance (see Kutas & Federmeier, 2000, 2011 for a review). Therefore, the degree of expectancy of targets within our sentences was previously tested by a separate group of participants.

One of the methodologies that serve to test word predictability in a context is the *cloze task*, which captures the probability of a word being selected as a continuation of a sentence (Taylor, 1953). A group of 30 French speakers was presented with fragments of the sentences that preceded the DPs and were asked to complete the sentences with a word that they thought was the most suitable. These speakers saw each DP member only once (i.e. they saw either the congruent or the incongruent sentences framing the DPs) and were not included in any of the experiments. The total correct predictions were divided by the total answers to calculate the cloze probability rate (Taylor, 1953). A paired t-test showed significant differences between the cloze probabilities of the contexts preceding the DPs ( $t(47) = 4.3, p < .0001$ ). Specifically, more items were correctly predicted in congruent ( $M = 0.13 \pm 0.2$ ) than in incongruent contexts ( $M = 0.004 \pm 0.01$ ) (see Table 2 in Annex I for detailed information). Results from the cloze task revealed few target choices after context onsets. Therefore, although our sentences favoured one of the two candidate words, they were overall rather low-predictable.

It has been discussed that cloze probability is influenced by the lexical properties of words, resulting in a higher likelihood of familiar words being chosen as cloze responses (Smith & Levy, 2011). Moreover, the selection of words of participants in the cloze task is likely influenced by the preceding words and context. This aligns with the assumption that exposure to a particular word triggers the spreading activation of its associated words (Neely, 1977; Oka, 1990). Cloze probability has been recently associated with (pre-)activation processes. This task has been thus argued to capture the relative activation of lexical candidates (Staub et al., 2015). If cloze probability serves as an indicator of lexical activation and captures the first word that comes to mind (as discussed by the authors), our results suggest that our critical DPs are not highly anticipated from the given sentences. Although this does not rule out the possibility that they are still expected based on the provided context, this observation may further explain the small and delayed ERP effects found in our conditions.

### **9.3. Integrating low- and high-level cues during elision**

While acoustic-phonetic information found in elided and non-elided items is shown to be salient and robust enough for native French speakers (Cordero et al., 2020; Do Carmo-Blanco et al., 2019; Spinelli et al., 2007, 2010), the role of acoustic cues during speech segmentation has been claimed to be overestimated (Mattys & Melhorn, 2007). In their study, English speakers were asked to choose between near homophonic utterances such as /plʌmpaɪ/ for “plum pie” vs. “plump eye” either within context or isolation (Mattys & Melhorn, 2007). When context was incongruent with the intended utterance, listeners strongly relied on contextual information. However, in one of the experiments, participants were asked to focus on the acoustic information of utterances, which resulted in an enhanced reliance on low-level cues compared to context. The authors argued that their findings further support the view that lexical but not sentential information modulates the processing of acoustic cues of the speech signal (Connine, 1987; Samuel, 1981).

Based on their findings, the authors proposed a redefinition of the original proposal of the hierarchy of cues theory (Mattys et al., 2005), introducing a grading between levels of processing such that the listeners’ reliance on cues would be proportional to their strength in the signal (Mattys & Melhorn, 2007). Additionally, this theory predicts a diminishing likelihood of engaging sublexical tiers as the utterance unfolds, since contextual information increasingly constrains lexical activation and selection (Mattys et al., 2005).



Moreover, the researchers discussed that scenarios in which attention is driven to word forms may not be ecological. However, our findings demonstrate that sublexical cues contribute to the process of speech segmentation even in the presence of higher-level cues. Moreover, our findings have shown the reliance on and integration of low-level information during both the processing of the signal (Experiments 1 and 2) and the activation of lexical candidates (Experiment 2). By employing French DPs in elision environments, we have shown that listeners tune to fine-grained acoustic details during the early stages of speech processing (bottom-up mechanism), regardless of the focus of attention. However, due to the nature of DPs in elision, i.e. being phonetically identical, contextual information had modulatory effects during lexical retrieval (top-down mechanism).

The hierarchy of cues theory (Mattys, 2004; Mattys et al., 2005, 2007; Mattys & Melhorn, 2007; White et al., 2012) advocates for the dominance of knowledge-based cues over sublexical cues due to their relevance to the communicative function of speech. According to this, relying on higher-level information to extract meaningful units from the input increases the likelihood of successful communication compared to relying solely on sublexical cues. However, our findings from both behavioural and electrophysiological measures evidence that acoustic-phonetic cues in speech are processed regardless of task demands. Furthermore, we have shown that sublexical cues come into play when listeners cannot rely on lexical information, even if contextual information is available. Therefore, our findings evidence that the reliance on higher-order cues does not preclude the processing of lower-order cues, such as suprasegmental information. Rather, this highlights the listeners' capacity to integrate multiple cues and selectively incorporate or adjust cues that align better with the available information. Consequently, the hierarchy of and interaction between cues can be seen as a flexible framework where cues at different levels interact at different time windows. These cues can compete or have their weights adjusted during real-time segmentation processes based on cue availability, robustness, and reliability, which can be influenced by factors such as signal quality or ambiguity.

#### **9.4. The role of F0 as a cue to speech segmentation in French**

Our findings from Experiment 3 confirm those from our previous experiments, providing further evidence of the role of the acoustic-phonetic features of DPs within elision during word recognition in French. Particularly, our findings offer additional empirical support

for the impact of fine-grained acoustic details on boundary processing (Christophe et al., 2004; Davis et al., 2002; Salverda et al., 2003). Naturally spoken utterances carry gliding patterns, characterised by continuous fluctuations in F0 within the speech segment. It is possible then that it is the combination of multiple cues –in addition to F0 variations– that contribute to a successful segmentation and word recognition, including duration and formant differences.

In order to keep coarticulatory variations present in the signal and provide a more ecological setting, the F0 characteristics alone of the pivotal vowel were manipulated while keeping other factors constant in the resynthesised conditions. The gradual enhancement found in the SR group in the perception of vowel-initial intended targets contributes to the existing evidence that the F0 serves as a cue to word boundaries in many languages (e.g., Davis et al., 2002; Llisterri et al., 2002; Wang et al., 2013; Welby & Niebuhr, 2016). In the case of French, we have provided further evidence that F0 rises mark the onset of content words (Di Cristo, 2000; Spinelli et al., 2010; Vaissière, 1997; Welby, 2003, 2007).

Additionally, our findings further contribute to this research by emphasising the relevance of the F0 dynamic parameters. F0 has typically been considered by its average frequency value within a given speech segment (e.g., Do Carmo-Blanco et al., 2019; Laures & Bunton, 2003; Laures & Weismer, 1999; Miller et al., 2010; Spinelli et al., 2007, 2010). Our findings challenge this notion and demonstrate that F0 modulations (i.e. its slope) have an impact during accurate segmentation. Particularly, compared to mean F0 values alone, the combination of both slope and mean F0 values has a greater influence on speech segmentation and facilitates word recognition. This is in line with previous studies demonstrating the benefits of the F0 trajectory (i.e. its rises and falls) as a cue during speech segmentation. The F0 trajectory has been found to resolve word-boundary ambiguities in English (Ladd & Schepman, 2003) and has been linked to word recognition in languages such as Korean (H. Kim & Tremblay, 2021) and Chinese (Wu, 2019; Zou et al., 2022). F0 fluctuations have been shown to influence the listener's perception, directing their attention towards the content words within the utterance. Consequently, the absence of these cues has been found to reduce the overall intelligibility of utterances (e.g., Binns & Culling, 2007; Laures & Weismer, 1999).

### 9.5. The exploitation of acoustic-phonetic cues to elision within spoken word recognition models

The psycholinguistic models of spoken word recognition TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and NAM/PARSYN (Goldinger, 1998; Luce, 1986; Luce et al., 2000; Luce & Pisoni, 1998) have considered the processes involved in mapping sensory information from the acoustic input to the stored lexical entries in the mental lexicon. In these models, segmentation is considered an outcome of lexical competition, wherein competition among candidate words leads to the identification of targets. Therefore, these models do not adequately explain the accurate segmentation of DPs in elision environments such as *la mie / l'amie* or *la fiche / l'affiche*.

The acoustic-phonetic cues present in these DPs have been evidenced to be processed both in isolation and within sentences. Even the Bayesian model based on Shortlist, Shortlist B (Norris & McQueen, 2008) fails to adequately account for such ambiguities since it operates based on sequences of multiple phoneme probabilities instead of sequences of discrete phonemes and does not consider the phonetic detail of speech stretches. Consequently, we suggest that these models integrate suprasegmental information in low-level stages of speech processing, enabling a modulation also in high-level stages.

Fine-grained acoustic details are accounted for in exemplar-based models of word recognition (e.g., Goldinger, 1998; Pierrehumbert, 2001), which consider that the collection of exemplars (or instances of words in memory) includes multiple-level cues including phonetic and contextual information. In addition to the stored information, the process of word recognition also incorporates the variability of the signal caused by factors such as coarticulation effects, F0, formants, and intonation patterns (Johnson, 1997, 2006). Additionally, exemplar models assume that word recognition is influenced by the frequency and recency of exemplars, with stronger recognition for exemplars that are more frequent and recently encountered. This would suggest that the identification of the intended words is (more) effective if words are frequently and recently used by individuals, increasing the variability of successful word recognition. Hence, the communicative goal of speech in these models would not be fully accomplished. Our studies further contributes to evidence suggesting that general rather than individual low-level features (such as rise at the content-word onset) may enhance word recognition (as also discussed in Spinelli et al., 2010).

## **9.6. The implications of the phonological deficits in developmental dyslexia to the detection of acoustic-phonetic information**

Experiment 3 highlights the importance of considering both task demands and materials in determining the segmentation and word recognition processes by dyslexic adults. Findings from this study revealed a gradual enhancement in the perception of vowel-initial targets for dyslexic adults, similar to the SR group. The fact that similar perceptual patterns were found in both groups can have several explanations. On the one hand, task demands impact the results. While in previous studies participants had to either make a choice among several answers or provide *yes/no*-type answers. In our study, participants were producing overt answers instead, like in (Leong & Goswami, 2014a). This is supported by differences in task results in both dyslexic children (e.g., Cheung et al., 2009) and adults (e.g., Leong & Goswami, 2014).

It is worth considering that our findings can also have been influenced by the compensatory mechanisms of dyslexic participants, who all attended university at the moment of testing. Despite poor phonological and reading skills, university students with dyslexia have been shown to develop a relatively high level of vocabulary (Cavalli et al., 2016a). They are constantly exposed to written material and most of them have likely developed adaptive reading strategies relying on oral comprehension or vocabulary skills that have allowed them to achieve a good level of reading comprehension (Callens et al., 2012, 2014; Cavalli et al., 2016; Olander et al., 2017). This hypothesis is supported by studies reporting deficits in acoustic/phonetic encoding in readers with dyslexia compared to skilled readers despite intact phonological grammar (Berent et al., 2013, 2016).

Despite similarities in the response pattern of both dyslexics and controls, dyslexics were less able to categorise targets as vowel initial items, which indicates that dyslexics show poorer processing of the F0 features. This aligns with the idea that adults with DD have been shown to have an impaired perception of speech rhythm at speech rates associated with syllable-related modulations (Goswami, 2011; J. A. Hämäläinen et al., 2012; Leong & Goswami, 2014a), but also with frequency rates at the phoneme level (Lehongre et al., 2011; Merzenich et al., 1993; Tallal et al., 1993). In the current study, participants were exposed to homophonic utterances such as /selafiʃ/, which could be segmented into either *affiche* or *fiche*. From our results, we cannot conclude whether the decoding differences between dyslexics and typical readers lay on either the phoneme or the syllable level, but they seem to be compatible with the hypothesis of a deficient slow auditory sampling in

dyslexia (Goswami, 2011), reflecting a poorer perception of the speech rise time compared to controls.

However, it is important to mention that differences in F0 variations present in our material were slightly manipulated, and slope and mean values were not exaggerated. Therefore, we think that the possible difficulties of dyslexics may also be due to the subtle differences in F0 cues. Increasing to a greater degree the acoustic parameters of resynthesised conditions may lead to different performances in dyslexic individuals, which would demonstrate the limits of reliance on the acoustic information of these individuals.

It is worth discussing that dyslexic participants seemed to display a perceptual tendency towards consonant-initial words. We have argued that our results are consistent with the neuroscientific oscillation theory, which posits that individuals with dyslexia exhibit an atypical functioning of the left auditory cortex, resulting in a faster sampling of auditory stimuli compared to typical readers (Lehongre et al., 2011, 2013). This aligns with the notion that the left auditory cortex of dyslexic individuals may have reduced responsiveness to variations at specific frequencies that are relevant for phonemic processing (30 Hz) while maintaining normal or even heightened sensitivity to higher frequencies (the anchoring-deficit hypothesis, Ahissar, 2007).

Notably, consonants have been shown to exhibit higher frequencies (higher Hz values) due to their shorter duration and higher energy content, as opposed to vowels which display lower frequencies (lower Hz values) and longer durations (Hillenbrand et al., 1995; Ramus & Mehler, 1999). Consequently, auditory stimuli may undergo differential processing depending on the underlying neuronal oscillatory activity. The consonant bias found for dyslexics in the current study seems to reflect a differing underlying processing between consonants and vowels, consistent with the oversampling of higher frequencies (i.e. consonants).

### **9.7. The role of L1 strategies during the segmentation of French speech as an L2**

In Chapter 8, both Spanish and English groups of learners display better discrimination sensitivity compared to L1 French speakers. These findings suggest that learners relied on their L1 strategies to segment speech and recognise words in the L2. Specifically, some of the acoustic cues present in the critical items (duration and F0 values) signal

stress in the learners' L1 (both English and Spanish). Notably, English listeners exhibit better performance during the discrimination of items. Contrasts between vowel-initial and consonant-initial items reside in acoustic cue variations at word onsets, i.e. in the first syllable of DPs. Hence, the enhanced sensitivity of English learners can be linked to the use of F0 rise patterns in word-initial positions for disyllabic nouns in English (e.g., Clopper, 2002; Cutler & Carter, 1987; Dabouis et al., 2020; Tyler & Cutler, 2009). These findings support evidence from the cue-weighting theory, which proposes that the weight of prosodic cues in the L1 can predict its use during L2 encoding, even if such cues serve different functions in the languages (Tremblay et al., 2016, 2017). We argue that learners were able to benefit from the cues that are relevant in their L1, which is in line with the idea that the greater the functional weight of cues in the L1, the more likely listeners will employ it to process an L2 (Chang, 2018).

Our results indicate a transfer of suprasegmental cues between L1 and L2, allowing both Spanish and English listeners to benefit from acoustic cues to locate word boundaries and discriminate between items in the L2. These findings align with the study of Tremblay et al. (2021) and suggests that for Spanish and, particularly, English speakers it is easier to learn new (L2) segmentation strategies between prosodic cues and word boundaries in French than to inhibit their L1 segmentation strategies (e.g., Tremblay & Spinelli, 2014; Weber & Cutler, 2006).

The adequate processing of the prosodic information of our pair contrasts was necessary to locate word edges between function (definite article) and content (noun) words in order to properly recognise words (*l#amie* vs. *la#mie*). However, English and Spanish speakers lack such boundary routines in their L1, as there is an absence of an elided form of the article in both languages. As our results revealed improved discrimination performance by learners compared to L1 French speakers, it suggests that beyond transferring segmentation cues from their L1 to L2, these learners effectively incorporated L1 segmentation strategies and prosodic cues into their L2 segmentation process for successful recognition of L2-specific speech units. Consequently, we argue that this cross-linguistic transfer may increase the sensitivity to prosodic information in an L2 rather than functioning exclusively as a specific cue for segmenting speech.

Our results suggest that the presence of similar segmentation patterns between the L1 and the prosodic features of an L2 can be advantageous for L2 processing. We argue that if acoustic information found in speech segments is relevant to successful communication

in the target language (such as the acoustic differences between elided and non-elided phrases), it may be assimilated and used by learners. This seems to be consistent with the findings by Tremblay et al. (2021). In their study, French learners of Korean were able to assimilate the intonational patterns related to word edges in the target language (Korean) (final H – initial L tones) compared to the L1 segmentation strategy (final High but not initial Low). Our findings provide further evidence indicating that listeners are not always at a disadvantage when processing L2 speech, but cross-linguistic transfer from their L1 can result in a non-native advantage surpassing that of L1 speakers, which is consistent with previous studies (e.g., Chang, 2018; Chang & Mishler, 2012; Choi et al., 2019; Davidson, 2011; Wiener & Goss, 2019).

Like the participants in the study by Tremblay et al. (2021), the learners in our study were all living in the country where the L2 is spoken and had daily exposure to the language. This is likely to influence the perception and/or learning of the acoustic-phonetic features in the target language. As mentioned by Tremblay et al. (2021), the context of immersion (i.e. individuals living in the country of the target language and having daily exposure to the L2) may reduce perceptual difficulties of subtle differences between the prosodic features of L1 and L2. The authors hypothesised that if the immersion experience reduces the perceptual difficulties between languages, the underlying mechanisms of acquisition and segmentation will not be different but rather vary in their “nature of information”, such that learning difficulties may be less persistent for prosodic cues than for segmentation cues. In other words and as explained by the authors, learners would “become more sensitive to fine-grained phonetic details of intonation compared to those of segments” (Tremblay et al., 2021, p. 11). Our study seems to indicate that the sensitivity of learners may still be modulated as a function of its relevance in the L1. However, our study does not investigate to what extent proficiency of or exposure to the L2 can influence sensitivity to low-level information. Nevertheless, our research does not explore the degree to which proficiency in or exposure to the L2 can impact sensitivity to low-level information. Furthermore, it remains unclear whether the processes of L1-L2 transfer and accommodation differ for languages that exhibit similar relationships between cues and word boundaries versus those with dissimilar relationships.

### 9.8. Main differences between liaison and elision: The differing roles of consonants and vowels

The present dissertation focused on the sensitivity of cues to elision processing in French. However, since few studies have focused on the perception of elision, the results of studies assessing the perception of liaison have also been presented throughout the previous chapters in order to show the influence of resyllabification during spoken word recognition, which has been studied in both L1 speakers (Dumay et al., 2002; Nguyen et al., 2007; Shoemaker, 2014; Spinelli et al., 2002, 2003; Tremblay & Spinelli, 2013; Wauquier-Gravelines & Braud, 2005; Welby, 2003, 2007) and learners (Gaskell et al., 2002; Gustafson & Bradlow, 2016; Shoemaker, 2009; Shoemaker & Birdsong, 2008; Shoemaker, 2010; Tremblay & Spinelli, 2014).

While elision and liaison are both phonological phenomena that cause a misalignment of syllable and word boundaries, an important difference between these phenomena lies in the pivotal phoneme involved in the resyllabification segment. In liaison environments, resyllabification entails pivotal consonants, while in elision contexts, it involves the /a/ vowel. Research has demonstrated that consonant and vowel information have different roles during lexical activation and word recognition processes. For instance, studies using a word reconstruction paradigm (Boelte, 1997; Cutler et al., 2000; van Ooijen, 1996) have demonstrated that vowel information has less influence on listeners compared to consonant information, which has been associated with the idea that consonants carry more weight than vowels during lexical retrieval processes. When participants were asked to transform nonwords such as “kebra” into real words, they were more prone to choose *cobra* than *zebra*, i.e. one requiring substitution of a vowel. Moreover, substitutions requiring vowels were consistently more accurate and faster than those requiring consonants (Cutler et al., 2000, for Dutch and Spanish; van Ooijen, 1996, for English).

Differences between vowel and consonant information were also explored by Nespor et al. (2003), who proposed that vowels carry structural information, while consonants play a relevant role in lexical processing. This functional asymmetry was suggested to impact language acquisition (Havy et al., 2014; Hochmann et al., 2011; Nazzi & Cutler, 2019). However, infant studies conducted across various languages have provided evidence contradicting this hypothesis, instead highlighting an early advantage for vowels. For instance, French-learning infants at 5 months old were more impaired at recognising their name when the vowel was mispronounced than when the consonant was, reflecting a



vowel bias (Bouchon et al., 2015). This bias toward vowels appears to be present for 6-month-old infants but transitions to a consonant bias between 6 and 8 months old (Nishibayashi & Nazzi, 2016). Similarly, Italian-learning infants also exhibit a vocalic bias at 6 months old, which later shifts to a consonant bias during the second half of their first year of life (Hochmann et al., 2017). Consequently, while there seems to be a vocalic advantage in early stages of life, a consonant bias seems to emerge as early as the first year of life and continues into adulthood (see Nazzi & Cutler, 2019 for a recent comprehensive review).

In addition to the functional differences between vowel and consonant in speech, it's important to note that the resyllabification processes in elision and liaison are not entirely analogous. Elision commonly takes place at the beginning of a content word, marking the boundary between a function word and a content word. Consequently, in most cases, the word on the left edge is shorter than the word on the right. In contrast, liaison can occur at the juncture of two function words, such as an adjective and a noun. Both liaison and elision involve variations in the properties of their pivotal sounds compared to their counterparts. However, the aforementioned differences along with the substantial acoustic variations identified in the pivotal sound [a] within elision in this dissertation, suggest that the impact of liaison and elision on speech perception and word recognition may not be identical.

## 9.9. Conclusion

In this dissertation, we have examined the role of acoustic-phonetic cues in the detection and disambiguation of DPs in ambiguous environments due to elision in French. The current work has investigated the perceptual capacities of native French speakers with and without phonological processing deficits and adult learners of French in discriminating differences in prosodic properties including F0. As our acoustic analyses revealed, such differences arise at the onset, i.e. in word-initial position, in resyllabified DPs within elision. The main findings from the material across experiments show that the first segment of vowel-initial DPs (e.g., *l'amie*) had both longer durations and higher F0 values.

Overall, our findings provide evidence supporting the notion that listeners exploit word onsets in French (e.g., Content et al., 2001; Cutler & Norris, 1988; Dumay et al., 1998, 2002) and rely on subtle acoustic distinctions, when available, as guidance during the

identification of words (e.g., Chang & Mishler, 2012; Davidson, 2011; Do Carmo-Blanco et al., 2019; Gaskell et al., 2002; Ivanova et al., 2023; Kim & Tremblay, 2021; Lukyanchenko et al., 2011; Qin et al., 2017; Shoemaker, 2009, 2014; Spinelli et al., 2002, 2003, 2007, 2010; Tremblay & Spinelli, 2014; Varnet, 2015; Welby, 2007; Welby & Niebuhr, 2016; Zhang & Francis, 2010). Particularly, our studies contribute to the existing evidence of the prominence of acoustic-phonetic cues present in DPs in elision environments in French. This has been previously demonstrated across multiple methodological approaches using behavioural measures (discrimination, identification, and priming tasks in Spinelli et al. (2007)) and electrophysiological measures (a modified MMN paradigm in Do Carmo-Blanco et al. (2019)).

Our first study (Chapter 6) sought to replicate and extend these findings with the implementation of two experimental designs that allowed for the exploration of the neural correlates to online speech processing in a more natural setup, in the presence of higher-level cues (context). Embedding homophonic DPs in congruent and incongruent contexts provided participants with natural environments of elision that, through the use of the ERP technique, allowed us to directly test the online processing of low-level and high-level information as they became available. Although results have to be interpreted with caution, they provide insight into the integration of the cues while processing speech.

Even though the acoustic features might be subtle, our findings indicate that the acoustic features of DPs in elision environments are processed by L1 French speakers, regardless of task demands and attentional focus. However, as speech unfolds, when informative enough sentential information is provided, listeners give more credence to high-level information than acoustic or lexical information. Nonetheless, the activation of lexical candidates seems to be less efficient when there is a lack of agreement between cues (i.e. when acoustic-phonetic information is embedded in disavouring contextual information). This suggests that both possible homophonic candidates are activated during lexical retrieval, as indicated by results from both ERP and behavioural measures. Therefore, our findings reveal that the listeners' perceptual system undergoes processes linked to bottom-up control mechanisms during the processing of ambiguity in elision environments, which are later restored by top-down mechanisms as the utterance unfolds. This highlights that even though a hierarchy of cues is present during speech processing (Mattys et al., 2005; Mattys & Melhorn, 2007; White et al., 2012), fine-grained acoustic

details have a relevant role and seem to be continuously integrated throughout the segmentation process.

Rise time and F0 have been shown to serve as an important acoustic cue for acoustic prominence (e.g., Christophe et al., 1994; Goswami et al., 2011; Greenberg, 2006; Leong et al., 2011; Rosen, 1992; Salverda et al., 2003; Welby, 2003). This dissertation contributes to the existing evidence that F0 is an important cue to word edges, also in the of elision.

In Chapter 7, we compared the perceptual capacities of acoustic-phonetic cues in elision between individuals with dyslexia and age-matched controls. We directly assessed the sensitivity of these French listeners to the intonational patterns of elision by manipulating the F0 contour of the pivotal /a/, while holding other acoustic cues in the signal constant. F0 presents itself in speech as “microscopic perturbations” (Welby, 2003 p. 222), and these fluctuations seem to guide listeners in finding word boundaries (e.g., Ivanova et al., 2023; Kim & Tremblay, 2021, 2022; Nguyen et al., 2008; Tremblay et al., 2021). Our methodology allowed us to test the impact of the F0 features on the perception of homophonic word pairs, including the individual impact of the F0 slope and mean, as well as their combined effect.

Our findings contribute to the existing evidence that F0 is an important cue to word edges, included in French elision. Particularly, our findings reveal a gradual reliance on F0 parameters for both groups of participants. Neurotypical French speakers exhibit a stronger reliance on the combination of both the mean value and slope value (trajectory) of F0 than on the mean value alone, and both are stronger than the F0 trajectory alone. These findings underscore the relevance of F0 as a cue to segment speech in French, a cue that is used even by speakers with phonological processing deficits, albeit to a lesser degree. Namely, dyslexic adults show a reduced level of categorisation of items as being vowel-initial, likely displaying a bias towards consonants. These effects reflect difficulties in processing intonation patterns, such as F0 features, aligning with the hypothesis of a deficient slow auditory sampling in dyslexia, as proposed by Goswami (2011).

Our findings from the last study (Chapter 8) show a non-native advantage for the detection and discrimination of elision in French DPs. English learners were better than Spanish learners at discriminating between elided and non-elided items, and both were

better than L1 French speakers. Determiner elisions are absent constructions in both learners' L1, suggesting that a transfer of prosodic cues from L1 to L2. Particularly, learners benefited from L1 cues during L2 processing, which is in line with previous studies showing an advantage of learners over L1 speakers (Chang, 2018; Chang & Mishler, 2012; Choi et al., 2019; Davidson, 2011; Wiener & Goss, 2019). These learners seem to have abstracted from their L1 routines a general ability to extract information from initial word-boundary cues, which they are then able to apply to new word onsets in a different language, which is consistent with the cue-weighting theory (Tremblay et al., 2016, 2017; Tremblay, Kim, et al., 2021).

### **9.10. Limitations**

Despite providing evidence of the relevance of F0 (mean and trajectory) as a cue to identify word boundaries in elision contexts during speech segmentation in French, our findings have revealed additional acoustic features that distinguish elided and non-elided items. Acoustic analyses have identified variations in duration and formant values for the phonemes comprising the onset syllable (/l/ and /a/). Additionally, the adjacent consonant displays differences in F0 and duration, and the subsequent vowel exhibits variations in F0, duration, and formant values. Based on our materials and experimental designs, disentangling the relative weights of the multiple cues associated with these items proves challenging. Consequently, we cannot definitively conclude that F0 alone is pivotal in word boundary recognition during the segmentation of ambiguous speech due to elision environments; rather, it is plausible that a combination of cues, increasing item variability and discrimination, guides listeners in segmenting intended items.

Moreover, the utterances in which DPs were embedded in Experiments 1 and 2 (Chapter 6) may have further affected the ERP and behavioural effects. The cloze task results showed lower levels of cloze probability even for congruent contexts. Moreover, the sentences used in the experiments had varying verb tense and number of elements previous to the critical item, which can also influence ERP effects.

Finally, the sample sizes in Experiments 3 and 4 represent limitations in our studies. As a result, we believe it is crucial to replicate these studies with larger sample sizes to determine whether the findings remain consistent with our current results.

### 9.11. Future research perspectives

Future studies can contribute to the knowledge of the integration of multiple cues during speech perception by exploring additional measures. For instance, an early detection of differences in our critical DPs may be also captured by earlier ERP components, such as the N1-P2 complex. As evidenced in both experiments, our results from both topographic maps and ERPs revealed an absence of this early sensory component. Prior research has indicated that the absence of this complex component (N1-P2) can be attributed to the lack of a pause before words in continuous spoken speech (Connolly et al., 1990; Connolly & Phillips, 1994; Friederici et al., 1993; Hagoort & Brown, 2000). In our study, ERPs were measured from the article onset across conditions, which means that some sentential information had already been processed by participants. Consequently, speech stretches that precede the DPs entail intonational variations elicited within the continuous speech stream, which may not evoke clear N1-P2 effects. This is in agreement with studies displaying smaller ERP amplitudes for early sensory components when elicited within continuous speech (Sanders & Neville, 2003a; Van Petten et al., 2000). Further studies that incorporate clearer onsets of DPs may be able to elucidate the extent to which the detection of acoustic-phonetic information in homophonic DPs begins even earlier and is driven by automatic cognitive processes of sensory detection.

Moreover, subsequent investigations should assess the materials employed in this thesis in sentences featuring identical sentence structure, i.e., including the same elements preceding the utterance and placing the DPs at sentence-final positions. As an illustration, *Il manque ...* (“... is missing”) can be used as a neutral context preceding all DPs (e.g., *Il manque l’amie/la mie*, “The friend/The breadcrumb is missing”).

Further studies should address the gap regarding the influence of the learners’ proficiency level in an L2, as well as the influence of their daily exposure on the perception and possibly production of L2 prosodic features. Research that encompasses groups with varying L2 proficiency levels and diverse levels of daily L2 exposure can shed light on the respective impact of these factors on the acquisition and use of pertinent prosodic features in an L2. Furthermore, the inclusion of the eye-tracking technique would offer insights into the real-time processing of such features.

The material used in this dissertation enables testing different populations to explore the reliance on and use of acoustic-phonetic cues in elision. Moreover, this material can be used to evaluate potential variations in how these cues are processed by various

populations, including children or individuals with hearing problems or cochlear implants.

## References

- Ablin, P., Cardoso, J.-F., & Gramfort, A. (2018). Faster Independent Component Analysis by Preconditioning With Hessian Approximations. *IEEE Transactions on Signal Processing*, *66*. <https://doi.org/10.1109/TSP.2018.2844203>
- Adda-Decker, M., Gendrot, C., & Nguyen, N. (2008). *Contributions du traitement automatique de la parole à l'étude des voyelles orales du français*.
- Adlard, A., & Hazan, V. (1998). Speech perception in children with specific reading difficulties (dyslexia). *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *51*(1), 153–177. <https://doi.org/10.1080/713755750>
- Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, *11*(11), 458–465. <https://doi.org/10.1016/j.tics.2007.08.015>
- Ahissar, M., Lubin, Y., Putter-Katz, H., & Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, *9*(12), 1558–1564. <https://doi.org/10.1038/nn1800>
- Ainsworth, W. (1975). *Intrinsic and Extrinsic Factors in Vowel Judgements* (G. Fant & M. A. A. Tatham, Eds.; pp. 103–113). Academic Press. <https://doi.org/10.1016/B978-0-12-248550-3.50011-8>
- Aldamen, H., & Al-Deaibes, M. (2023). Arabic emphatic consonants as produced by English speakers: An acoustic study. *Heliyon*, *9*(2), e13401. <https://doi.org/10.1016/j.heliyon.2023.e13401>
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. <https://doi.org/10.1006/jmla.1997.2558>
- Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research*, *21*(4), 325–358. <https://doi.org/10.1191/0267658305sr250oa>
- Anderson, C. (2018). *4.1 Phonemes and Contrast*. <https://ecampusontario.pressbooks.pub/essentialsoflinguistics/chapter/4-2-phonemes-and-contrast/>
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*, 233–250. [https://doi.org/10.1016/S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)
- Assmann, P. (1999). Fundamental frequency and the intelligibility of competing voices. *Proceedings of the 14th International Congress of Phonetic Sciences*, 179–182. [https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14\\_0179.pdf](https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0179.pdf)
- Baayen, H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*.
- Bahler, C. L., Coughlin, C., & Tremblay, A. (2011). Differential Contribution of Prosodic Cues in Native and Non-native Speech Segmentation. *ICPhS*.
- Baković, E. (2016). Exceptionality in Spanish Stress. *Catalan Journal of Linguistics*, *15*, 9–25. <https://doi.org/10.5565/rev/catjl.182>
- Baldeweg, T., Richardson, A., Watkins, S., Foale, C., & Gruzelier, J. (1999). Impaired auditory frequency discrimination in dyslexia detected with mismatch evoked potentials. *Annals of Neurology*, *45*(4), 495–503. [https://doi.org/10.1002/1531-8249\(199904\)45:4<495::aid-ana11>3.0.co;2-m](https://doi.org/10.1002/1531-8249(199904)45:4<495::aid-ana11>3.0.co;2-m)

- Banai, K., & Ahissar, M. (2006). Auditory Processing Deficits in Dyslexia: Task or Stimulus Related? *Cerebral Cortex*, *16*(12), 1718–1728. <https://doi.org/10.1093/cercor/bhj107>
- Banel, M.-H., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication*, *15*(1), 115–126. [https://doi.org/10.1016/0167-6393\(94\)90046-9](https://doi.org/10.1016/0167-6393(94)90046-9)
- Banel, M.-H., & Bacri, N. (1997). Reconnaissance de la parole et indices de segmentation métriques et phonotactiques. *L'Année psychologique*, *97*(1), 77–112. <https://doi.org/10.3406/psy.1997.28937>
- Bartoń, K. (2020). *r.squaredGLMM: Pseudo-R-squared for Generalized Mixed-Effect models in MuMIn: Multi-Model Inference* [Computer software]. <https://rdrr.io/cran/MuMIn/man/r.squaredGLMM.html>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). <https://doi.org/10.18637/jss.v067.i01>
- Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. *Papers in Laboratory Phonology*, *3*.
- Bentin, S., Kutas, M., & Steven, A. H. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*(2), 161–169. <https://doi.org/10.1111/j.1469-8986.1993.tb01729.x>
- Berent, I., Vaknin-Nusbaum, V., Balaban, E., & Galaburda, A. M. (2012). Dyslexia Impairs Speech Recognition but Can Spare Phonological Competence. *PLoS ONE*, *7*(9), e44875. <https://doi.org/10.1371/journal.pone.0044875>
- Berent, I., Vaknin-Nusbaum, V., Balaban, E., & Galaburda, A. M. (2013). Phonological generalizations in dyslexia: The phonological grammar may not be impaired. *Cognitive Neuropsychology*, *30*(5), 285–310. <https://doi.org/10.1080/02643294.2013.863182>
- Berent, I., Zhao, X., Balaban, E., & Galaburda, A. (2016). Phonology and phonetics dissociate in dyslexia: Evidence from adult English speakers. *Language, Cognition and Neuroscience*, *31*(9), 1178–1192. <https://doi.org/10.1080/23273798.2016.1211301>
- Best, C. (1993). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.
- Best, C. (1995). A direct realist view of crosslanguage speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171–204).
- Best, C., McRoberts, G., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *The Journal of the Acoustical Society of America*, *109*, 775–794.
- Best, C. T. (1994). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. In H. Nussbaum, J. Goodman, & D. Howard (Eds.), *The transition from speech to spoken words: The development of speech perception* (pp. 167–224). MIT Press.
- Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege* (pp. 13–34). <https://doi.org/10.1075/llt.17.07bes>
- Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America*, *122*(3), 1765–1776. <https://doi.org/10.1121/1.2751394>
- Blau, V., Atteveldt, N. van, Ekkebus, M., Goebel, R., & Blomert, L. (2009). Reduced Neural Integration of Letters and Speech Sounds Links Phonological and Reading Deficits in



- Adult Dyslexia. *Current Biology*, 19(6), 503–508. <https://doi.org/10.1016/j.cub.2009.01.065>
- Blomert, L. (2011a). The neural signature of orthographic–phonological binding in successful and failing reading development. *NeuroImage*, 57(3), 695–703. <https://doi.org/10.1016/j.neuroimage.2010.11.003>
- Blomert, L. (2011b). The neural signature of orthographic-phonological binding in successful and failing reading development. *NeuroImage*, 57(3), 695–703. <https://doi.org/10.1016/j.neuroimage.2010.11.003>
- Blomert, L., & Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech. *Brain and Language*, 89(1), 21–26. [https://doi.org/10.1016/S0093-934X\(03\)00305-5](https://doi.org/10.1016/S0093-934X(03)00305-5)
- Blomert, L., Mitterer, H., & Paffen, C. (2004). In Search of the Auditory, Phonetic, and/or Phonological Problems in Dyslexia. *Journal of Speech, Language, and Hearing Research : JSLHR*, 47, 1030–1047. [https://doi.org/10.1044/1092-4388\(2004\)077](https://doi.org/10.1044/1092-4388(2004)077)
- Boelte, J. (1997). *The role of mismatching information in spoken word recognition*. Verlag Dr. Kovac.
- Boets, B., Op de Beek, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Martini, D., Bulthé, J., Sunaert, S., Wouters, J., & Ghesquière, P. (2013). *Intact But Less Accessible Phonetic Representations in Adults with Dyslexia*. 342(6163), 1251–1254. <https://doi.org/10.1126/science.1244333>
- Bogliotti, C., Messaoud-Galusi, S., & Serniclaes, W. (2002). *Relations entre la perception catégorielle de la parole et l'apprentissage de la lecture*. 197–200.
- Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., & Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology*, 101(2), 137–155. <https://doi.org/10.1016/j.jecp.2008.03.006>
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16(6), 451–459. <https://doi.org/10.1111/j.0956-7976.2005.01556.x>
- Bonte, M., & Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Brain Research. Cognitive Brain Research*, 21, 360–376. <https://doi.org/10.1016/j.cogbrainres.2004.06.010>
- Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. In *The impact of allophony versus contrast on speech* (pp. 145–172). De Gruyter Mouton. <https://doi.org/10.1515/9783110208603.2.145>
- Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2019). Cracking the Code: The Impact of Orthographic Transparency and Morphological-Syllabic Complexity on Reading and Developmental Dyslexia. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02534>
- Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, 18(4), 587–598. <https://doi.org/10.1111/desc.12242>
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, 301(5899), Article 5899. <https://doi.org/10.1038/301419a0>
- Brèthes, H., Cavalli, E., Denis-Noël, A., Melmi, J.-B., El Ahmadi, A., Bianco, M., & Colé, P. (2022). Text Reading Fluency and Text Reading Comprehension Do Not Rely on the

- Same Abilities in University Students With and Without Dyslexia. *Frontiers in Psychology*, 13. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.866543>
- Bretz, F., Hothorn, T., & Westfall, P. (2010). Multiple Comparisons Using R. In *Chapman and Hall/CRC*. <https://doi.org/10.1201/9781420010909>
- Breznitz, Z., & Meyler, A. (2003). Speed of lower-level auditory and visual processing as a basic factor in dyslexia: Electrophysiological evidence. *Brain and Language*, 85(2), 166–184. [https://doi.org/10.1016/S0093-934X\(02\)00513-8](https://doi.org/10.1016/S0093-934X(02)00513-8)
- Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology (2006)*, 64(1), 74–95. <https://doi.org/10.1080/17470218.2010.499174>
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180. <https://doi.org/10.1159/000261913>
- Brysbaert, M. (2013). Lextale\_FR A Fast, Free, and Efficient Test to Measure Language Proficiency in French. *Psychologica Belgica*, 53, 23. <https://doi.org/10.5334/pb-53-1-23>
- Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science (New York, N.Y.)*, 304(5679), 1926–1929. <https://doi.org/10.1126/science.1099745>
- Callens, M., Tops, W., & Brysbaert, M. (2012). Cognitive profile of students who enter higher education with an indication of dyslexia. *PloS One*, 7(6), e38081. <https://doi.org/10.1371/journal.pone.0038081>
- Callens, M., Tops, W., Stevens, M., & Brysbaert, M. (2014). An exploratory factor analysis of the cognitive functioning of first-year bachelor students with dyslexia. *Annals of Dyslexia*, 64. <https://doi.org/10.1007/s11881-013-0088-6>
- Campbell, R., & Butterworth, B. (1985). Phonological Dyslexia and Dysgraphia in a Highly Literate Subject: A Developmental Case with Associated Deficits of Phonemic Processing and Awareness. *Quarterly Journal of Experimental Psychology*, 37(3). <https://doi.org/10.1080/14640748508400944>
- Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, 403(6768), 428–430. <https://doi.org/10.1038/35000206>
- Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory of foreign-language teaching. *Linguistics*, 11(112), 5–14. <https://doi.org/10.1515/ling.1973.11.112.5>
- Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test* (p. 27). Psychological Corporation.
- Carroll, S. E. (2001). *Input and Evidence: The raw material of second language acquisition* (Vol. 25). John Benjamins Publishing Company. <https://doi.org/10.1075/lald.25>
- Castles, A., & Friedmann, N. (2014). Developmental dyslexia and the phonological deficit hypothesis. *Mind & Language*, 29(3), 270–285. <https://doi.org/10.1111/mila.12050>
- Cavalli, E., Casalis, S., Ahmadi, A. E., Zira, M., Poracchia-George, F., & Colé, P. (2016a). Vocabulary skills are well developed in university students with dyslexia: Evidence from multiple case studies. *Research in Developmental Disabilities*, 51–52, 89–102. <https://doi.org/10.1016/j.ridd.2016.01.006>
- Cavalli, E., Casalis, S., Ahmadi, A. E., Zira, M., Poracchia-George, F., & Colé, P. (2016b). Vocabulary skills are well developed in university students with dyslexia: Evidence from multiple case studies. *Research in Developmental Disabilities*, 51–52, 89–102. <https://doi.org/10.1016/j.ridd.2016.01.006>

- Cavalli, E., Colé, P., Brèthes, H., Lefevre, E., Lascombe, S., & Velay, J.-L. (2019). E-book reading hinders aspects of long-text comprehension for adults with dyslexia. *Annals of Dyslexia*, *69*(2), 243–259. <https://doi.org/10.1007/s11881-019-00182-w>
- Cavalli, E., Colé, P., Leloup, G., Poracchia-George, F., Sprenger-Charolles, L., & El Ahmadi, A. (2018). Screening for Dyslexia in French-Speaking University Students: An Evaluation of the Detection Accuracy of the Alouette Test. *Journal of Learning Disabilities*, *51*(3), 268–282. <https://doi.org/10.1177/0022219417704637>
- Cavalli, E., Duncan, L. G., Elbro, C., El Ahmadi, A., & Colé, P. (2017). Phonemic—Morphemic dissociation in university students with dyslexia: An index of reading compensation? *Annals of Dyslexia*, *67*(1), 63–84. <https://doi.org/10.1007/s11881-016-0138-y>
- Chang, C. (2018). Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics*, *68*, 85–102. <https://doi.org/10.1016/j.wocn.2018.03.003>
- Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791–809. <https://doi.org/10.1017/S1366728914000261>
- Chang, C. B., & Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *The Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.4747615>
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*(1), 205–215. <https://doi.org/10.1017/s0305000900013180>
- Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PLOS ONE*, *12*(7), e0180656. <https://doi.org/10.1371/journal.pone.0180656>
- Cheung, H., Chung, K. K. H., Wong, S. W. L., McBride-Chang, C., Penney, T. B., & Ho, C. S. H. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *50*(6), 726–733. <https://doi.org/10.1111/j.1469-7610.2008.02001.x>
- Chiappe, P., Chiappe, D. L., & Siegel, L. S. (2001). Speech Perception, Lexicality, and Reading Skill. *Journal of Experimental Child Psychology*, *80*(1), 58–74. <https://doi.org/10.1006/jecp.2000.2624>
- Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*(2). <https://doi.org/10.1006/jpho.2001.0153>
- Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155–190. <https://doi.org/10.1006/jpho.2001.0131>
- Cho, T., & Mücke, D. (2021). Articulatory Measures of Prosody. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 15–28). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198832232.013.2>
- Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners. *Cognition*, *189*, 188–192. <https://doi.org/10.1016/j.cognition.2019.04.004>
- Chrabaszcz, A., & Gor, K. (2014). Context Effects in the Processing of Phonolexical Ambiguity in L2: Context Effects in Processing of L2. *Language Learning*, *64*(3), 415–455. <https://doi.org/10.1111/lang.12063>
- Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic Cues to Perception of Word Stress by English, Mandarin, and Russian Speakers. *Journal of Speech, Language,*

and *Hearing Research : JSLHR*, 57(4), 1468–1479.  
[https://doi.org/10.1044/2014\\_JSLHR-L-13-0279](https://doi.org/10.1044/2014_JSLHR-L-13-0279)

- Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, 95(3), 1570–1580. <https://doi.org/10.1121/1.408544>
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, 51(4), 523–547. <https://doi.org/10.1016/j.jml.2004.07.001>
- Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, 32(3), 274–285. <https://doi.org/10.1111/j.1469-8986.1995.tb02956.x>
- Clements, G. N. (2001). Representational economy in constraint-based phonology. In T. A. Hall (Ed.), *Distinctive feature theory* (pp. 71–146). Mouton de Gruyter. <https://www.degruyter.com/document/doi/10.1515/9783110886672.71/html?lang=en>
- Clopper, C. (2002). Frequency of Stress Patterns in English: A Computational Analysis. *Indiana University Linguistics Club Working Papers Online 2*.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences, 3rd ed* (pp. xxviii, 703). Lawrence Erlbaum Associates Publishers.
- Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press.
- Cole, R., Yan, Y., Mak, B., & Fanty, M. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *The Journal of the Acoustical Society of America*, 100, 2689–2689. <https://doi.org/10.1121/1.417028>
- Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110(1), 24–30. [https://doi.org/10.1016/S0168-5597\(98\)00033-1](https://doi.org/10.1016/S0168-5597(98)00033-1)
- Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, 26(5), 527–538. [https://doi.org/10.1016/0749-596X\(87\)90138-0](https://doi.org/10.1016/0749-596X(87)90138-0)
- Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3), 256–266. <https://doi.org/10.1162/jocn.1994.6.3.256>
- Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, 39(2), 302–318. [https://doi.org/10.1016/0093-934X\(90\)90016-A](https://doi.org/10.1016/0093-934X(90)90016-A)
- Content, A., Kearns, R., & Frauenfelder, U. (2001). Boundaries versus Onsets in Syllabic Segmentation. *Journal of Memory and Language*, 45, 177–199. <https://doi.org/10.1006/jmla.2000.2775>
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of Lexical Stress on Lexical Access in English: Evidence from Native and Non-native Listeners. *Language and Speech*, 45(3), 207–228. <https://doi.org/10.1177/00238309020450030101>
- Corbera, S., Escera, C., & Artigas, J. (2006). Impaired duration mismatch negativity in developmental dyslexia. *NeuroReport*, 17(10), 1051–1055. <https://doi.org/10.1097/01.wnr.0000221846.43126.a6>
- Cordero, M. del M., Meunier, F., Grimault, N., Pota, S., & Spinelli, E. (2020). F0 Slope and Mean: Cues to Speech Segmentation in French. *Interspeech 2020*, 1610–1614. <https://doi.org/10.21437/Interspeech.2020-2509>

- Coughlin, C. E., & Tremblay, A. (2011). The Role of Prosodic Information in L2 Speech Segmentation. *Proceedings of the 35 Th BUCLD*, 18.
- Cutler, A. (1986). Forbear is a Homophone: Lexical Prosody Does Not Constrain Lexical Access. *Language and Speech*, 29(3), 201–220. <https://doi.org/10.1177/002383098602900302>
- Cutler, A. (1996). Prosody and the Word Boundary Problem. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax. Bootstrapping From Speech To Grammar in Early Acquisition* (p. 13). Psychology Press.
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, 2(3), 133–142. [https://doi.org/10.1016/0885-2308\(87\)90004-0](https://doi.org/10.1016/0885-2308(87)90004-0)
- Cutler, A., & Foss, D. J. (1977). On the Role of Sentence Stress in Sentence Processing. *Language and Speech*, 20(1). <https://doi.org/10.1177/002383097702000101>
- Cutler, A., Howard, D., & Patterson, K. E. (1989). Misplaced stress on prosody: A reply to Black and Byng. *Cognitive Neuropsychology*, 6(1), 67–83. <https://doi.org/10.1080/02643298908253285>
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385–400. [https://doi.org/10.1016/0749-596X\(86\)90033-1](https://doi.org/10.1016/0749-596X(86)90033-1)
- Cutler, A., & Norris, D. (1988a). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113–121. <https://doi.org/10.1037/0096-1523.14.1.113>
- Cutler, A., & Norris, D. (1988b). *The role of strong syllables in segmentation for lexical access*. <https://doi.org/10.1037/0096-1523.14.1.113>
- Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, 28(5), 746–755. <https://doi.org/10.3758/bf03198409>
- Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a Homophone: Lexical Prosody and Lexical Access in Dutch. *Language and Speech*, 44(2), 171–195. <https://doi.org/10.1177/00238309010440020301>
- Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). *Dutch listeners' use of suprasegmental cues to English stress*. 4.
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34(2), 269–284. <https://doi.org/10.1016/j.wocn.2005.06.002>
- d'Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Nguyen, N., & Welby, P. (2007). Are tones aligned with articulatory events? Evidence from Italian and French. *Papers in Laboratory Phonology* 9, 9, 577–608.
- Dabouis, Q., Enguehard, G., Fournier, J.-M., & Lampitelli, N. (2020). The English “Arab Rule” without feet. *Acta Linguistica Academica*, 67(1), 121–134. <https://doi.org/10.1556/2062.2020.00009>
- Dahan, D., & Brent, M. (1999). On the discovery of novel wordlike units from utterances: An artificial-language study with implications for native-language acquisition. *Journal of Experimental Psychology. General*, 128, 165–185. <https://doi.org/10.1037/0096-3445.128.2.165>
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42(4), 317–367. <https://doi.org/10.1006/cogp.2001.0750>

- Dal Ben, R., Souza, D. de H., & Hay, J. F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition*, *49*(7), 1300–1310. <https://doi.org/10.3758/s13421-021-01163-4>
- Daniels, P. T., & Share, D. L. (2018). Writing System Variation and Its Consequences for Reading and Dyslexia. *Scientific Studies of Reading*, *22*(1), 101–116. <https://doi.org/10.1080/10888438.2017.1379082>
- Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English– L2 French acquisition. *Second Language Research*, *28*(1), 5.
- Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Plasticity in compensation for phonological variation: The case of late second language learners. *Laboratory Phonology*, *9*, 145–172.
- Davidson, L. (2011). Phonetic, Phonemic, and Phonological Factors in Cross-Language Discrimination of Phonotactic Contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 270–282. <https://doi.org/10.1037/a0020988>
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147. <https://doi.org/10.1016/j.heares.2007.01.014>
- Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 218–244. <https://doi.org/10.1037/0096-1523.28.1.218>
- de Carvalho, C. A. F., Kida, A. de S. B., Capellini, S. A., & de Avila, C. R. B. (2014). Phonological working memory and reading in students with dyslexia. *Frontiers in Psychology*, *5*, 746. <https://doi.org/10.3389/fpsyg.2014.00746>
- Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*(1), 60–74. <https://doi.org/10.1111/1469-8986.00120>
- Deacon, D., & Shelley-Tremblay, J. (2000). How automatically is meaning accessed: A review of the effects of attention on semantic processing. *Frontiers in Bioscience-Landmark*, *5*(4), Article 4. <https://doi.org/10.2741/deacon>
- Del Tufo, S. N., & Earle, F. S. (2020). Skill Profiles of College Students With a History of Developmental Language Disorder and Developmental Dyslexia. *Journal of Learning Disabilities*, *53*(3), 228–240. <https://doi.org/10.1177/0022219420904348>
- Delattre, P. (1940). Le mot est-il une unité phonétique en français ? *Le Français Moderne*, *8*(1), 47–56.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, *27*, 761–764. <https://doi.org/10.1121/1.1908020>
- Denis-Noël, A., Pattamadilok, C., Castet, É., & Colé, P. (2020). Activation time-course of phonological code in silent word recognition in adult readers with and without dyslexia. *Annals of Dyslexia*, *70*(3), 313–338. <https://doi.org/10.1007/s11881-020-00201-1>
- Di Cristo, A. (2000). Vers une modélisation de l'accentuation du français (seconde partie). *Journal of French Language Studies*, *10*(1), 27–44. <https://doi.org/10.1017/S0959269500000120>

- Di Liberto, G. M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., & Lalor, E. C. (2018). Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*, *175*, 70–79. <https://doi.org/10.1016/j.neuroimage.2018.03.072>
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*, 121–144. [https://doi.org/10.1207/s15326969eco0102\\_2](https://doi.org/10.1207/s15326969eco0102_2)
- Diehl, R. L., Kluender, K. R., Walsh, M. A., & Parker, E. M. (1991). Auditory enhancement in speech perception and phonology. In *Cognition and the symbolic processes: Applied and ecological perspectives* (pp. 59–76). Lawrence Erlbaum Associates, Inc.
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, *55*, 149–179. <https://doi.org/10.1146/annurev.psych.55.090902.142028>
- Diehl, R., Lindblom, B., & Creeger, C. (2003). *Increasing realism of auditory representations yields further insights into vowel phonetics*.
- Do Carmo-Blanco, N., Hoen, M., Pota, S., Spinelli, E., & Meunier, F. (2019). Processing of non-contrastive subphonemic features in French homophonous utterances: An MMN study. *Journal of Neurolinguistics*, *52*, 100849. <https://doi.org/10.1016/j.jneuroling.2019.05.001>
- Du, Z., & Chen, B. (2023). Uyghur speakers' acquisition of Mandarin tones. *Journal of Pidgin and Creole Languages*, *38*(Language Contact with Chinese), 62–82. <https://doi.org/10.1075/jpcl.00104.du>
- Dufor, O., Serniclaes, W., Sprenger-Charolles, L., & Démonet, J.-F. (2007). Top-down processes during auditory phoneme categorization in dyslexia: A PET study. *NeuroImage*, *34*(4), 1692–1707. <https://doi.org/10.1016/j.neuroimage.2006.10.034>
- Dumay, N., Banel, M.-H., Frauenfelder, U., & Content, A. (1998). Le rôle de la syllabe: Segmentation lexicale ou classification? *Actes Des XXIIèmes Journées d'Etudes Sur La Parole*, 33–36. <https://kar.kent.ac.uk/14934/>
- Dumay, N., Frauenfelder, U., & Content, A. (2002). The Role of the Syllable in Lexical Segmentation in French: Word-Spotting Data. *Brain and Language*, *81*, 144–161. <https://doi.org/10.1006/brln.2001.2513>
- Dunn, L. M., Thieriault-Whalen, C. M., & Dunn, L. M. (1993). *Echelles de vocabulaire en image Peabody. Adaptation française du Peabody Picture Vocabulary Test-Revised*. Richmon Hill.
- Dunn, M., & Dunn, L. M. (1981). *Peabody picture vocabulary test – Revised*. AGS, Circle Pines.
- Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A Destressing “Deafness” in French? *Journal of Memory and Language*, *36*(3), 406–421. <https://doi.org/10.1006/jmla.1996.2500>
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress ‘deafness’. *The Journal of the Acoustical Society of America*, *110*(3 Pt 1), 1606–1618. <https://doi.org/10.1121/1.1380437>
- Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress ‘deafness’ in simultaneous French–Spanish bilinguals. *Cognition*, *114*(2), 266–275. <https://doi.org/10.1016/j.cognition.2009.10.001>
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition*, *106*(2), 682–706. <https://doi.org/10.1016/j.cognition.2007.04.001>
- Durand, J., & Lyche, C. (2008). French liaison in the light of corpus data. *Journal of French Language Studies*, *18*(1), 33.

- Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, *18*(5), 341–347. <https://doi.org/10.3758/BF03211210>
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science (New York, N.Y.)*, *171*(3968), 303–306. <https://doi.org/10.1126/science.171.3968.303>
- Elbro, C. (1998). When reading is ‘readn’ or somthn. Distinctness of phonological representations of lexical items in normal and disabled readers. *Scandinavian Journal of Psychology*, *39*(3), 149–153. <https://doi.org/10.1111/1467-9450.393070>
- Elbro, C., & Arnbak, E. (1996). The role of morpheme recognition and morphological awareness in dyslexia. *Annals of Dyslexia*, *46*, 209–240. <https://doi.org/10.1007/BF02648177>
- Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia*, *44*(1), 203–226. <https://doi.org/10.1007/BF02648162>
- Erlbeck, H., Kübler, A., Kotchoubey, B., & Vesper, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400. *Frontiers in Human Neuroscience*, *8*. <https://www.frontiersin.org/articles/10.3389/fnhum.2014.00654>
- Ernestus, M., Lahey, M., Verhees, F., Baayen, R. H., & Jasa, E. (2006). Lexical frequency and voice assimilation. *J. Acoust. Soc. Am.*, *120*, 1040–1051.
- Estebas, E. (2008). The Phonological Status of English and Spanish Prenuclear F0 Peaks. *ATLANTIS*, *292*, 39–57.
- Face, T. L. (2005). F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, *3*(2 (6)), 49–65.
- Faraway, J. J. (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Second Edition* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315382722>
- Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, *2*(4), 460–493. <https://doi.org/10.3758/BF03210983>
- Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. <https://doi.org/10.1006/jmla.1999.2660>
- Flege, J. (1995). *Second language speech learning: Theory, findings and problems* (pp. 229–273).
- Flege, J. E. (1980). Phonetic Approximation in Second Language Acquisition1. *Language Learning*, *30*(1), 117–134. <https://doi.org/10.1111/j.1467-1770.1980.tb00154.x>
- Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults’ perception of /ɪ/ and /I/. *The Journal of the Acoustical Society of America*, *99*(2), 1161–1173. <https://doi.org/10.1121/1.414884>
- Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 414–420. [https://doi.org/10.1016/S0022-5371\(65\)80081-0](https://doi.org/10.1016/S0022-5371(65)80081-0)
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*(2), 109–135. <https://doi.org/10.1006/jpho.2000.0114>
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*(6), 3728–3740. <https://doi.org/10.1121/1.418332>



- Fox, R., Flege, J., & Munro, M. (1995). The perception of English and Spanish vowels by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *97*, 2540–2551. <https://doi.org/10.1121/1.411974>
- Francis, A., Baldwin, K., & Nusbaum, H. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*, 1668–1680.
- Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, *124*(2), 1234–1251. <https://doi.org/10.1121/1.2945161>
- Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. <https://doi.org/10.1037/0096-1523.28.2.349>
- Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*, *54*(3), 287–295. <https://doi.org/10.3758/BF03205263>
- Friederici, A., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing; Effect of semantic morphological and syntactic violations. *Brain Research. Cognitive Brain Research*, *1*, 183–192. [https://doi.org/10.1016/0926-6410\(93\)90026-2](https://doi.org/10.1016/0926-6410(93)90026-2)
- Friedrich, M., & Friederici, A. D. (2004). N400-like Semantic Incongruity Effect in 19-Month-Olds: Processing Known Words in Picture Contexts. *Journal of Cognitive Neuroscience*, *16*(8), 1465–1477. <https://doi.org/10.1162/0898929042304705>
- Fry, D. B. (1958). Experiments in the Perception of Stress. *Language and Speech*, *1*(2), 126–152. <https://doi.org/10.1177/002383095800100207>
- Fry, D. B. (1965). The Dependence of Stress Judgments on Vowel Formant Structure. *Phonetic Sciences*, 306–311. <https://doi.org/10.1159/000426965>
- Garner, W. R., & Hake, H. W. (1951). The amount of information in absolute judgements. *Psychological Review*, *58*(6), 446–459. <https://doi.org/10.1037/h0054482>
- Gaskell, G., & Mirković, J. (2017). *Speech Perception and Spoken Word Recognition* (1st ed.). Psychology Press. <https://www.routledge.com/Speech-Perception-and-Spoken-Word-Recognition/Gaskell-Mirkovic/p/book/9781848724402>
- Gaskell, M. G., Spinelli, E., & Meunier, F. (2002). Perception of resyllabification in French. *Memory & Cognition*, *30*(5), 798–810. <https://doi.org/10.3758/BF03196435>
- Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130. <https://doi.org/10.3389/fpsyg.2011.00130>
- Ghitza, O., & Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, *66*(1–2), 113–126. <https://doi.org/10.1159/000208934>
- Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), Article 4. <https://doi.org/10.1038/nn.3063>
- Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology*, *32*(3), 401–424. [https://doi.org/10.1016/0022-0965\(81\)90105-3](https://doi.org/10.1016/0022-0965(81)90105-3)
- Goldinger, S. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, *105*, 251–279. <https://doi.org/10.1037/0033-295X.105.2.251>

- Gordon, M. (2011). Stress Systems. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed., pp. 141–163). Wiley Blackwell. <https://doi.org/10.1002/9781444343069.ch5>
- Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a cross-linguistic theoretical framework. *Dyslexia*, 6(2), 133–151. [https://doi.org/10.1002/\(SICI\)1099-0909\(200004/06\)6:2<133::AID-DYS160>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1099-0909(200004/06)6:2<133::AID-DYS160>3.0.CO;2-A)
- Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15(1), 3–10. <https://doi.org/10.1016/j.tics.2010.10.001>
- Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, 16(1), Article 1. <https://doi.org/10.1038/nrn3836>
- Goswami, U. (2018). A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Current Directions in Psychological Science*, 27(1), 56–63. <https://doi.org/10.1177/0963721417727520>
- Goswami, U. (2019). Speech rhythm and language acquisition: An amplitude modulation phase hierarchy perspective. *Annals of the New York Academy of Sciences*, 1453(1), 67–78. <https://doi.org/10.1111/nyas.14137>
- Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, 23(8), 995–1019. <https://doi.org/10.1007/s11145-009-9186-6>
- Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J. P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*, 49(5), 1363–1376. <https://doi.org/10.1016/j.cortex.2012.05.005>
- Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, 4(1), 67–92. <https://doi.org/10.1515/lp-2013-0004>
- Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, 69(1), 1–17. <https://doi.org/10.1016/j.jml.2013.03.001>
- Goswami, U., Power, A. J., Lallier, M., & Facoetti, A. (2014). Oscillatory “temporal sampling” and developmental dyslexia: Toward an over-arching theoretical framework. *Frontiers in Human Neuroscience*, 8(904). <https://doi.org/10.3389/fnhum.2014.00904>
- Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99(16), 10911–10916. <https://doi.org/10.1073/pnas.122368599>
- Goswami, U., Wang, H.-L. S., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2011). Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish, and Chinese. *Journal of Cognitive Neuroscience*, 23(2), 325–337. <https://doi.org/10.1162/jocn.2010.21453>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, 9(3), 317–323. [https://doi.org/10.1016/0028-3932\(71\)90027-3](https://doi.org/10.1016/0028-3932(71)90027-3)
- Gow, D. W. (2001). Assimilation and Anticipation in Continuous Spoken Word Recognition. *Journal of Memory and Language*, 45(1), 133–159. <https://doi.org/10.1006/jmla.2000.2764>
- Granena, G. (2013). *Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test\** (pp. 105–130). <https://doi.org/10.1075/llt.35.04gra>

- Granena, G. (2014). Language Aptitude and Long-term Achievement in Early Childhood L2 Learners. *Applied Linguistics*, 35, 483–503. <https://doi.org/10.1093/applin/amu013>
- Greenaway, R. E. (2017). Chapter 13—ABX Discrimination Task. In L. Rogers (Ed.), *Discrimination Testing in Sensory Science* (pp. 267–288). Woodhead Publishing. <https://doi.org/10.1016/B978-0-08-101009-9.00013-7>
- Greenberg, S. (2006). A multi-tier framework for understanding spoken language. In S. Greenberg & W. Ainsworth (Eds.), *Understanding speech: An auditory perspective* (pp. 411–434). Lawrence Erlbaum Associates Publishers. [https://www.researchgate.net/publication/289385121\\_A\\_multi-tier\\_framework\\_for\\_understanding\\_spoken\\_language](https://www.researchgate.net/publication/289385121_A_multi-tier_framework_for_understanding_spoken_language)
- Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of foreign language learning ability: The canal-F theory and test. *Modern Language Journal*, 84(3), 390–405. <https://doi.org/10.1111/0026-7902.00076>
- Gross-Glenn, K., Jallad, B., Novoa, L., Helgren-Lempesis, V., & Lubs, H. A. (1990). Nonsense passage reading as a diagnostic aid in the study of adult familial dyslexia. *Reading and Writing*, 2(2), 161–173. <https://doi.org/10.1007/BF00401800>
- Gruber, M. (2003). *Dyslexics' phonological processing in relation to speech perception* [Umeå University]. <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A140310&dswid=5298>
- Gustafson, E., & Bradlow, A. R. (2016). French Speech Segmentation in Liaison Contexts by L1 and L2 Listeners. *Laboratory Phonology*, 7(1), 17. <https://doi.org/10.5334/labphon.59>
- Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F0 cues in the voiced–voiceless distinction. *Journal of Phonetics*, 9(1), 49–62. [https://doi.org/10.1016/S0095-4470\(19\)30926-X](https://doi.org/10.1016/S0095-4470(19)30926-X)
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38(11), 1531–1549. [https://doi.org/10.1016/s0028-3932\(00\)00053-1](https://doi.org/10.1016/s0028-3932(00)00053-1)
- Halle, M. (1992). Phonological features. In W. Bright (Ed.), *International encyclopedia of linguistics* (Vol. 3, pp. 207–212). Oxford University Press.
- Hämäläinen, J. A., Leppänen, P. H. T., Eklund, K., Thomson, J., Richardson, U., Guttorm, T. K., Witton, C., Poikkeus, A.-M., Goswami, U., & Lyytinen, H. (2009). Common variance in amplitude envelope perception tasks and their impact on phoneme duration perception and reading and spelling in Finnish children with reading disabilities. *Applied Psycholinguistics*, 30(3), 511–530. <https://doi.org/10.1017/S0142716409090250>
- Hämäläinen, J. A., Rupp, A., Soltész, F., Szűcs, D., & Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: An MEG study. *NeuroImage*, 59(3), 2952–2961. <https://doi.org/10.1016/j.neuroimage.2011.09.075>
- Hämäläinen, J., Salminen, H., & Leppänen, P. (2013). Basic Auditory Processing Deficits in Dyslexia: Systematic Review of the Behavioral and Event-Related Potential/Field Evidence. *Journal of Learning Disabilities*, 46. <https://doi.org/10.1177/0022219411436213>
- Hampson, M., Tokoglu, F., Sun, Z., Schafer, R. J., Skudlarski, P., Gore, J. C., & Constable, R. T. (2006). Connectivity–behavior analysis reveals that functional connectivity between left BA39 and Broca's area varies with reading ability. *NeuroImage*, 31(2), 513–519. <https://doi.org/10.1016/j.neuroimage.2005.12.040>
- Han, J.-I. (2009). Task Effects on the Perception of Phonological Contrasts by Korean Monolingual Adults. *언어*, 34(3), 737–759.

- Harnsberger, J. D. (2001). The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics*, 29(3), 303–327. <https://doi.org/10.1006/jpho.2001.0140>
- Harris, J. W. (1983). *SYLLABLE STRUCTURE AND STRESS IN SPANISH: A NONLINEAR ANALYSIS* (Vol. 7). MIT Press. <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/article/abs/syllable-structure-and-stress-in-spanish-a-nonlinear-analysis-james-w-harris-cambridge-mass-mit-press-1983-pp-176/B96A5200EE62CCF446C88852019AD047>
- Havy, M., & Nazzi, T. (2009). Better Processing of Consonantal Over Vocalic Information in Word Learning at 16 Months of Age. *Infancy*, 14(4), 439–456. <https://doi.org/10.1080/15250000902996532>
- Havy, M., Serres, J., & Nazzi, T. (2014). A consonant/vowel asymmetry in word-for processing: Evidence in childhood and in adulthood. *Language and Speech*, 57(Pt 2), 254–281. <https://doi.org/10.1177/0023830913507693>
- Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., & Shakespeare, B. (2009). Speech Perception Abilities of Adults With Dyslexia: Is There Any Evidence for a True Deficit? *Journal of Speech, Language, and Hearing Research*, 52(6), 1510–1529. [https://doi.org/10.1044/1092-4388\(2009/08-0220\)](https://doi.org/10.1044/1092-4388(2009/08-0220))
- Heidlmayr, K., Ferragne, E., & Isel, F. (2021). Neuroplasticity in the phonological system: The PMN and the N400 as markers for the perception of non-native phonemic contrasts by late second language learners. *Neuropsychologia*, 156, 107831. <https://doi.org/10.1016/j.neuropsychologia.2021.107831>
- Heiervang, E., Stevenson, J., & Hugdahl, K. (2002). Auditory processing in children with dyslexia. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 43, 931–938. <https://doi.org/10.1111/1469-7610.00097>
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5), 3099–3111. <https://doi.org/10.1121/1.411872>
- Hochmann, J.-R., Benavides-Varela, S., Flo, A., Nespors, M., & Mehler, J. (2017). Bias for Vocalic Over Consonantal Information in 6-Month-Olds. *Infancy*, 23. <https://doi.org/10.1111/inf.12203>
- Hochmann, J.-R., Benavides-Varela, S., Nespors, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science*, 14(6), 1445–1458. <https://doi.org/10.1111/j.1467-7687.2011.01089.x>
- Holliman, A. J., Sheehy, K., & Wood, C. (2012). A Cross-sectional study of prosodic sensitivity and reading difficulties. *Journal of Research in Reading*, 35(1), 32–48. <https://doi.org/10.1111/j.1467-9817.2010.01459.x>
- Holliman, A. J., Wood, C., & Sheehy, K. (2010). The contribution of sensitivity to speech rhythm and non-speech rhythm to early reading development. *Educational Psychology*, 30(3), 247–267. <https://doi.org/10.1080/01443410903560922>
- Holmqvist Olander, M., Wennås Brante, E., & Nyström, M. (2017). The Effect of Illustration on Improving Text Comprehension in Dyslexic Adults. *Dyslexia (Chichester, England)*, 23(1), 42–65. <https://doi.org/10.1002/dys.1545>
- Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, 119(5), 3059–3071. <https://doi.org/10.1121/1.2188377>

- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal. Biometrische Zeitschrift*, *50*(3), 346–363. <https://doi.org/10.1002/bimj.200810425>
- Huang, Y., Wang, M., & Rao, H. (2021). Working memory as a partial mediator of the relationship between multilingualism and foreign language aptitude. *International Journal of Multilingualism*, *0*(0), 1–18. <https://doi.org/10.1080/14790718.2021.1993860>
- Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *47*(6), 674–689. <https://doi.org/10.1016/j.cortex.2010.07.010>
- Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics*, *39*(4), 571–584. <https://doi.org/10.1016/j.wocn.2011.03.003>
- Ito, K., & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English: The Journal of the Acoustical Society of America: Vol 125, No 4. *The Journal of the Acoustical Society of America*, *125*(4), 2348–2360. <https://doi.org/10.1121/1.3082103>
- Ivanova, M., Neubert, C. R., Schmied, J., & Bendixen, A. (2023). ERP evidence for Slavic and German word stress cue sensitivity in English. *Frontiers in Psychology*, *14*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1193822>
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. [https://doi.org/10.1016/S0010-0277\(02\)00198-1](https://doi.org/10.1016/S0010-0277(02)00198-1)
- Ives, D. T., & Patterson, R. D. (2008). Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics. *The Journal of the Acoustical Society of America*, *123*(5), 2670–2679. <https://doi.org/10.1121/1.2890737>
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing*. Academic Press.
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. <https://doi.org/10.1016/j.wocn.2005.08.004>
- Joos, M. (1948). Acoustic phonetics. *Language Monographs*, *23*, 136–136.
- 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](https://doi.org/10.1007/978-94-011-4317-2_10)
- Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1). <https://doi.org/10.1515/prbs.2002.002>
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development*, *64*(3), 675–687. <https://doi.org/10.2307/1131210>
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, *39*(3), 159–207. <https://doi.org/10.1006/cogp.1999.0716>
- Katayama, J., & Polich, J. (1998). Stimulus context determines P3a and P3b. *Psychophysiology*, *35*(1), 23–33. <https://doi.org/10.1111/1469-8986.3510023>

- Kershner, J., & Micallef, J. (1992). Consonant-vowel lateralization in dyslexic children: Deficit or compensatory development? *Brain and Language*, 43(1), 66–82. [https://doi.org/10.1016/0093-934X\(92\)90021-6](https://doi.org/10.1016/0093-934X(92)90021-6)
- Kidd, G., Jr., Richards, V. M., Streeter, T., Mason, C. R., & Huang, R. (2011). Contextual effects in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, 130(6), 3926–3938. <https://doi.org/10.1121/1.3658442>
- Kim, H., & Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics*, 87, 101059. <https://doi.org/10.1016/j.wocn.2021.101059>
- Kim, H., & Tremblay, A. (2022). Intonational Cues to Segmental Contrasts in the Native Language Facilitate the Processing of Intonational Cues to Lexical Stress in the Second Language. *Frontiers in Communication*, 7. <https://www.frontiersin.org/article/10.3389/fcomm.2022.845430>
- Kim, J., Chris, D., & Cutler, A. (2008). Perceptual Tests of Rhythmic Similarity: II. Syllable Rhythm. *Language and Speech*, 51(Pt 4), 343–359. <https://doi.org/10.1177/0023830908099069>
- Kitzen, K. R. (2001). *Prosodic sensitivity, morphological ability, and reading ability in young adults with and without childhood histories of reading difficulty* (Vol. 62, p. 460). ProQuest Information & Learning.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5), 1208–1221. <https://doi.org/10.1121/1.380986>
- Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics*, 16(2), 153–169. [https://doi.org/10.1016/S0095-4470\(19\)30480-2](https://doi.org/10.1016/S0095-4470(19)30480-2)
- Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, 124(6), 3959–3971. <https://doi.org/10.1121/1.2999341>
- Koponen, E. (2001). Phonotactic and word accent cues for speech segmentation in Swedish one-year-olds. *Department of Linguistics and Phonetics, Lund University*. Fonetik 2001. <https://www.semanticscholar.org/paper/Phonotactic-and-word-accent-cues-for-speech-in-Koponen/2c080efc970085c287d017c9d819e70780095349>
- Kotchoubey, B., & Pavlov, Y. G. (2019). A Signature of Passivity? An Explorative Study of the N3 Event- Related Potential Component in Passive Oddball Tasks. *Frontiers in Neuroscience*, 13, 365. <https://doi.org/10.3389/fnins.2019.00365>
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20(4), 804–823. <https://doi.org/10.1037//0278-7393.20.4.804>
- Kuhl, P. K. (2007). Cracking the speech code: How infants learn language. *Acoustical Science and Technology*, 28(2), 71–83. <https://doi.org/10.1250/ast.28.71>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13–F21. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Kujala, T., Belitz, S., Tervaniemi, M., & Näätänen, R. (2003). Auditory sensory memory disorder in dyslexic adults as indexed by the mismatch negativity. *The European Journal of Neuroscience*, 17(6), 1323–1327. <https://doi.org/10.1046/j.1460-9568.2003.02559.x>

- Kujala, T., Halmetoja, J., Näätänen, R., Alku, P., Lyytinen, H., & Sussman, E. (2006). Speech- and sound-segmentation in dyslexia: Evidence for a multiple-level cortical impairment. *The European Journal of Neuroscience*, *24*, 2420–2427. <https://doi.org/10.1111/j.1460-9568.2006.05100.x>
- Kujala, T., Lovio, R., Lepistö, T., Laasonen, M., & Näätänen, R. (2006). Evaluation of multi-attribute auditory discrimination in dyslexia with the mismatch negativity. *Clinical Neurophysiology*, *117*(4), 885–893. <https://doi.org/10.1016/j.clinph.2006.01.002>
- Kujala, T., Myllyviita, K., Tervaniemi, M., Alho, K., Kallio, J., & Näätänen, R. (2000). Basic auditory dysfunction in dyslexia as demonstrated by brain activity measurements. *Psychophysiology*, *37*(2), 262–266. <https://doi.org/10.1111/1469-8986.3720262>
- Kujala, T., & Näätänen, R. (2001). The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neuroscience and Biobehavioral Reviews*, *25*(6), 535–543. [https://doi.org/10.1016/s0149-7634\(01\)00032-x](https://doi.org/10.1016/s0149-7634(01)00032-x)
- Kumar, S., & Schönwiesner, M. (2012). Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(39), 13348–13351. <https://doi.org/10.1523/JNEUROSCI.3812-12.2012>
- Küper, K., & Keil, M. (2009). Electrophysiology reveals semantic priming at a short SOA irrespective of depth of prime processing. *Neuroscience Letters*, *453*(2), 107–111.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. [https://doi.org/10.1016/S1364-6613\(00\)01560-6](https://doi.org/10.1016/S1364-6613(00)01560-6)
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*. <https://doi.org/10.1038/307161A0>
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205. <https://doi.org/10.1126/science.7350657>
- Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in Linear Mixed Effects Models* (3.1-3) [Computer software]. <https://CRAN.R-project.org/package=lmerTest>
- Ladd, D. R., & Schepman, A. (2003). “Sagging transitions” between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, *31*(1), 81–112. [https://doi.org/10.1016/S0095-4470\(02\)00073-6](https://doi.org/10.1016/S0095-4470(02)00073-6)
- Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages*. Blackwell Publishers.
- Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., & Carreiras, M. (2017). Amodal Atypical Neural Oscillatory Activity in Dyslexia: A Cross-Linguistic Perspective. *Clinical Psychological Science*, *5*(2), 379–401. <https://doi.org/10.1177/2167702616670119>
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O’Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J.-F., ... Schulte-Körne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, *54*(6), 686–694. <https://doi.org/10.1111/jcpp.12029>

- Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics*, 21(2), 243–262. <https://doi.org/10.1017/S0142716400002058>
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, 9(12), Article 12. <https://doi.org/10.1038/nrn2532>
- Laures, J. S., & Bunton, K. (2003). Perceptual effects of a flattened fundamental frequency at the sentence level under different listening conditions. *Journal of Communication Disorders*, 36(6), 449–464. [https://doi.org/10.1016/S0021-9924\(03\)00032-7](https://doi.org/10.1016/S0021-9924(03)00032-7)
- Laures, J. S., & Weismer, G. (1999). The effects of a flattened fundamental frequency on intelligibility at the sentence level. *Journal of Speech, Language, and Hearing Research: JSLHR*, 42(5), 1148–1156. <https://doi.org/10.1044/jslhr.4205.1148>
- Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, 52(11), 864–886. <https://doi.org/10.1016/j.specom.2010.08.014>
- Lefavrais, P. (1965). Description, définition et mesure de la dyslexie utilisation du test ‘L’Alouette.’ [Description, definition and measurement of a test for dyslexia.]. *Revue de Psychologie Appliquée*, 15, 33–34.
- Lefavrais, P. (2005). *Alouette-R: Test d’analyse de la vitesse en lecture à partir d’un texte*. les Éditions du centre de psychologie appliquée.
- Lehiste, I. (1970). *Suprasegmentals* (pp. viii, 194). Massachusetts Inst. of Technology P.
- Lehongre, K., Morillon, B., Giraud, A. L., & Ramus, F. (2013). Impaired auditory sampling in dyslexia: Further evidence from combined fMRI and EEG. *Frontiers in Human Neuroscience*, JUL, Article JUL. <https://doi.org/10.3389/fnhum.2013.00454>
- Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., & Giraud, A.-L. (2011). Altered Low-Gamma Sampling in Auditory Cortex Accounts for the Three Main Facets of Dyslexia. *Neuron*, 72(6), 1080–1090. <https://doi.org/10.1016/j.neuron.2011.11.002>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. <https://doi.org/10.3758/s13428-011-0146-0>
- Lenth, R. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, 69(1). <https://doi.org/10.18637/jss.v069.i01>
- Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.8.5) [Computer software]. <https://cran.r-project.org/web/packages/emmeans/index.html>
- Leong, V., & Goswami, U. (2014a). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research*, 308, 141–161. <https://doi.org/10.1016/j.heares.2013.07.015>
- Leong, V., & Goswami, U. (2014b). Impaired extraction of speech rhythm from temporal modulation patterns in speech in developmental dyslexia. *Frontiers in Human Neuroscience*, 8. <https://www.frontiersin.org/articles/10.3389/fnhum.2014.00096>
- Leong, V., Hämäläinen, J., Soltész, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, 64(1), 59–73. <https://doi.org/10.1016/j.jml.2010.09.003>



- Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36(1), 141–157. <https://doi.org/10.1016/j.wocn.2007.03.001>
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368. <https://doi.org/10.1037/h0044417>
- Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., & Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience*, 22(2), Article 2. <https://doi.org/10.1038/s41593-018-0308-9>
- Lin, C. Y., Wang, M., Idsardi, W. J., & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English\*. *Bilingualism: Language and Cognition*, 17(2), 316–346. <https://doi.org/10.1017/S1366728913000333>
- Lindblom, B. (1963). Spectrographic Study of Vowel Reduction. *The Journal of the Acoustical Society of America*, 35, 1773–1781. <https://doi.org/10.1121/1.2142410>
- 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 Netherlands. [https://doi.org/10.1007/978-94-009-2037-8\\_16](https://doi.org/10.1007/978-94-009-2037-8_16)
- Lindgren, S. D., De Renzi, E., & Richman, L. C. (1985). Cross-national comparisons of developmental dyslexia in Italy and the United States. *Child Development*, 56(6), 1404–1417.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94(3), 1242–1255. <https://doi.org/10.1121/1.408177>
- Lizarazu, M., Lallier, M., Molinaro, N., Bourguignon, M., Paz-Alonso, P. M., Lerma-Usabiaga, G., & Carreiras, M. (2015). Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. *Human Brain Mapping*, 36(12), 4986–5002. <https://doi.org/10.1002/hbm.22986>
- Llisterri, J., Machuca, M., de-la-Mota, C., Mota, L., Riera, M., & Rios, A. (2003). The perception of lexical stress in Spanish. In *Proceedings of the XVth International Conference of Phonetic Sciences: Vol. ISBN* (p. 2026).
- Llisterri, J., Machuca, M., de-la-Mota, C., RIERA, M., & Rios, A. (2002). *The role of F0 peaks in the identification of lexical stress in Spanish* (pp. 350–361).
- Lorenzi, C., Dumont, A., & Füllgrabe, C. (2000). Use of Temporal Envelope Cues by Children With Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research: JSLHR*, 43, 1367–1379. <https://doi.org/10.1044/jslhr.4306.1367>
- Lotto, A. J., & Holt, L. L. (2016). Chapter 16 - Speech Perception: The View from the Auditory System. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 185–194). Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00016-X>
- Luce, P. A. (1986a). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39, 155–158. <https://doi.org/10.3758/BF03212485>
- Luce, P. A. (1986b). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39(3), 155–158. <https://doi.org/10.3758/BF03212485>
- Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62(3), 615–625. <https://doi.org/10.3758/bf03212113>
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, 19(1), 1–36.

- Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press. <https://mitpress.mit.edu/9780262525855/an-introduction-to-the-event-related-potential-technique/>
- Lukyanchenko, A., Idsardi, W. J., & Jiang, N. (2011). Opening Your Ears: The Role of L1 in Processing of Nonnative Prosodic Contrasts. In G. Granena (Ed.), *Selected Proceedings of the 2010 Second Language Research Forum* (pp. 50–62). Cascadilla Proceedings Project.
- Lunden, A. (2017). Duration, vowel quality, and the rhythmic pattern of English. *Laboratory Phonology*, 8(1), Article 1. <https://doi.org/10.5334/labphon.37>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Lawrence Erlbaum Associates.
- Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia: The Importance of Serial Order. *Frontiers in Psychology*, 7, 1522. <https://doi.org/10.3389/fpsyg.2016.01522>
- Makowski, D. (2018). The psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, 3(22), 470. <https://doi.org/10.21105/joss.00470>
- Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are Speech Perception Deficits Associated with Developmental Dyslexia? *Journal of Experimental Child Psychology*, 66(2), 211–235. <https://doi.org/10.1006/jecp.1997.2383>
- Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America*, 69(2), 548–558. <https://doi.org/10.1121/1.385483>
- Marchesotti, S., Nicolle, J., Merlet, I., Arnal, L. H., Donoghue, J. P., & Giraud, A.-L. (2020). Selective enhancement of low-gamma activity by tACS improves phonemic processing and reading accuracy in dyslexia. *PLoS Biology*, 18(9), e3000833. <https://doi.org/10.1371/journal.pbio.3000833>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Marslen-Wilson, W. (1990). Activation, Competition, and Frequency in Lexical Access. In *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172).
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1), 71–102. [https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9)
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1), 29–63. [https://doi.org/10.1016/0010-0285\(78\)90018-X](https://doi.org/10.1016/0010-0285(78)90018-X)
- Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1), 1–71. [https://doi.org/10.1016/0010-0277\(80\)90015-3](https://doi.org/10.1016/0010-0277(80)90015-3)
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101(4), 653–675. <https://doi.org/10.1037/0033-295x.101.4.653>
- Martin, J., Colé, P., Leuwers, C., Casalis, S., Zorman, M., & Sprenger-Charolles, L. (2010). Reading in French-speaking adults with dyslexia. *Annals of Dyslexia*, 60(2), 238–264. <https://doi.org/10.1007/s11881-010-0043-8>

- Martin, J., Frauenfelder, U. H., & Colé, P. (2014). Morphological awareness in dyslexic university students. *Applied Psycholinguistics*, 35(6), 1213–1233. <https://doi.org/10.1017/S0142716413000167>
- Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 397–408. <https://doi.org/10.1037/0096-1523.30.2.397>
- Mattys, S. L., & Bortfeld, H. (2017). Speech segmentation. In *Speech Perception and Spoken Word Recognition*. Psychology Press.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978. <https://doi.org/10.1080/01690965.2012.705006>
- Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *The Journal of the Acoustical Society of America*, 122(1), 554–567. <https://doi.org/10.1121/1.2735105>
- Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 960–977. <https://doi.org/10.1037/0096-1523.33.4.960>
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology: General*, 134(4), 477–500. <https://doi.org/10.1037/0096-3445.134.4.477>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. <https://doi.org/10.1016/j.jml.2017.01.001>
- McArthur, G. m., & Bishop, D. v. m. (2001). Auditory perceptual processing in people with reading and oral language impairments: Current issues and recommendations. *Dyslexia*, 7(3), 150–170. <https://doi.org/10.1002/dys.200>
- McCarthy, G., Luby, M., Gore, J., & Goldman-Rakic, P. (1997). Infrequent Events Transiently Activate Human Prefrontal and Parietal Cortex as Measured by Functional MRI. *Journal of Neurophysiology*, 77(3), 1630–1634. <https://doi.org/10.1152/jn.1997.77.3.1630>
- McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica*, 45, 84–108.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McGivern, R., Berka, C., Languis, M., & Chapman, S. (1991). Detection of Deficits in Temporal Pattern Discrimination Using the Seashore Rhythm Test in Young Children with Reading Impairments. *Journal of Learning Disabilities*, 24, 58–62. <https://doi.org/10.1177/002221949102400110>
- McGuire, G. (2010). *A Brief Primer on Experimental Designs for Speech Perception Research*. 16.
- McGuire, G. L. (2007). *Phonetic category learning* [The Ohio State University]. [https://etd.ohiolink.edu/apexprod/rws\\_olink/r/1501/10?clear=10&p10\\_accession\\_num=osu1190065715](https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=osu1190065715)
- McNulty, M. A. (2003). Dyslexia and the Life Course. *Journal of Learning Disabilities*, 36(4). <https://doi.org/10.1177/00222194030360040701>
- McQueen, J. M. (1998). Segmentation of Continuous Speech Using Phonotactics. *Journal of Memory and Language*, 39(1), 21–46. <https://doi.org/10.1006/jmla.1998.2568>
- McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(3), 621–638. <https://doi.org/10.1037/0278-7393.20.3.621>

- Meara, P. (2005). *Llama, Language Aptitude Tests: The Manual*. Lognostics.
- Meara, P., & Rogers, V. (2019). *The LLAMA Tests v3*. Lognostics. [https://www.lognostics.co.uk/tools/LLAMA\\_3/index.htm](https://www.lognostics.co.uk/tools/LLAMA_3/index.htm)
- Mehler, J., Dommergues, J., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298–305. [https://doi.org/10.1016/S0022-5371\(81\)90450-3](https://doi.org/10.1016/S0022-5371(81)90450-3)
- Mengisidou, M., & Marshall, C. R. (2019). Deficient Explicit Access to Phonological Representations Explains Phonological Fluency Difficulties in Greek Children With Dyslexia and/or Developmental Language Disorder. *Frontiers in Psychology*, 10. <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00638>
- Merzenich, M. M., Schreiner, C., Jenkins, W., & Wang, X. (1993). Neural Mechanisms Underlying Temporal Integration, Segmentation, and Input Sequence Representation: Some Implications for the Origin of Learning Disabilities. *Annals of the New York Academy of Sciences*, 682(1 Temporal Info), 1.
- Meunier, C., & Espesser, R. (2011). Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics*, 39(3), 271–278. <https://doi.org/10.1016/j.wocn.2010.11.008>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97. <https://doi.org/10.1037/0033-295X.101.2.343>
- Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noise. *The Journal of the Acoustical Society of America*, 128(1), 435–443. <https://doi.org/10.1121/1.3397384>
- Moojen, S. M. P., Gonçalves, H. A., Bassôa, A., Navas, A. L., de Jou, G., & Miguel, E. S. (2020). Adults with dyslexia: How can they achieve academic success despite impairments in basic reading and writing abilities? The role of text structure sensitivity as a compensatory skill. *Annals of Dyslexia*, 70(1), 115–140. <https://doi.org/10.1007/s11881-020-00195-w>
- Mora, J. C. (2005). *Lexical Knowledge Effects on the Discrimination of Non-Native Phonemic Contrasts in Words and Non-Words by Spanish/Catalan Bilingual Learners of English*. 4.
- Mundy, I. R., & Carroll, J. M. (2012). Speech prosody and developmental dyslexia: Reduced phonological awareness in the context of intact phonological representations. *Journal of Cognitive Psychology*, 24(5), 560–581. <https://doi.org/10.1080/20445911.2012.662341>
- Muneaux, M., Ziegler, J. C., Truc, C., Thomson, J., & Goswami, U. (2004). Deficits in beat perception and dyslexia: Evidence from French. *NeuroReport*, 15(8), 1255. <https://doi.org/10.1097/01.wnr.0000127459.31232.c4>
- Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: Class and frequency effects. *Neuropsychologia*, 39(1), 91–102. [https://doi.org/10.1016/S0028-3932\(00\)00095-6](https://doi.org/10.1016/S0028-3932(00)00095-6)
- Murty, L., Otake, T., & Cutler, A. (2007). Perceptual Tests of Rhythmic Similarity: I. Mora Rhythm. *Language and Speech*, 50(1), 77–99. <https://doi.org/10.1177/00238309070500010401>
- Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329. [https://doi.org/10.1016/0001-6918\(78\)90006-9](https://doi.org/10.1016/0001-6918(78)90006-9)
- Nation, K., & Snowling, M. J. (2008). Individual Differences in Contextual Facilitation: Evidence from Dyslexia and Poor Reading Comprehension—Nation—1998—Child Development—Wiley Online Library. *Child Development*, 69(4), 996–1011. <https://doi.org/10.1111/j.1467-8624.1998.tb06157.x>

- Navarro, T. (1966). *Estudios de fonología española*. Las Américas.
- Nazzi, T., & Cutler, A. (2019). How Consonants and Vowels Shape Spoken-Language Recognition. *Annual Review of Linguistics*, 5(1), 25–47. <https://doi.org/10.1146/annurev-linguistics-011718-011919>
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85(5), 2088–2113. <https://doi.org/10.1121/1.397861>
- Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106(3), 226–254. <https://doi.org/10.1037/0096-3445.106.3.226>
- Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue E Linguaggio*, 2, 203–229.
- New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, 19(12), 1223–1227. <https://doi.org/10.1111/j.1467-9280.2008.02228.x>
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516–524. <https://doi.org/10.3758/BF03195598>
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, 101, 447–462.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48(2), 127–162. [https://doi.org/10.1016/S0010-0285\(03\)00128-2](https://doi.org/10.1016/S0010-0285(03)00128-2)
- Nguyen, A.-T. T., Ingram, J., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, 36, 158–190.
- Nguyen, N., Wauquier-Gravelines, S., Lancia, L., & Tuller, B. (2007). Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In P. Prieto (Ed.), *Segmental and Prosodic Issues in Romance Phonology* (pp. 3–23). John Benjamins. <https://hal.archives-ouvertes.fr/hal-00173545>
- Nishibayashi, L.-L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, 155, 188–203. <https://doi.org/10.1016/j.cognition.2016.07.003>
- Nobre, A. C., & McCarthy, G. (1994). Language-Related ERPs: Scalp Distributions and Modulation by Word Type and Semantic Priming. *Journal of Cognitive Neuroscience*, 6(3), 233–255. <https://doi.org/10.1162/jocn.1994.6.3.233>
- Noordenbos, M. W. (2013). *Phonological representations in dyslexia: Underspecified or overspecified?* Radboud University.
- Noordenbos, M. W., & Serniclaes, W. (2015). The Categorical Perception Deficit in Dyslexia: A Meta-Analysis. *Scientific Studies of Reading*, 19(5), 340–359. <https://doi.org/10.1080/10888438.2015.1052455>
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234. [https://doi.org/10.1016/0010-0277\(94\)90043-4](https://doi.org/10.1016/0010-0277(94)90043-4)
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. <https://doi.org/10.1037/0033-295X.115.2.357>

- Norris, D., McQueen, J. M., & Cutler, A. (2018). Commentary on “Interaction in Spoken Word Recognition Models”. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01568>
- Oganian, Y., & Ahissar, M. (2012). Poor anchoring limits dyslexics’ perceptual, memory, and reading skills. *Neuropsychologia*, 50(8), 1895–1905. <https://doi.org/10.1016/j.neuropsychologia.2012.04.014>
- Oka N. (1990). Effects of the task for prime words and the proportion of related prime-target pairs on semantic and orthographic priming. *The Japanese journal of psychology*, 61(4), 235–240. <https://doi.org/10.4992/jjpsy.61.235>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- O’Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology*, 60(2), 121–150. [https://doi.org/10.1016/S0301-0511\(02\)00045-5](https://doi.org/10.1016/S0301-0511(02)00045-5)
- Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. *Language and Speech*, 54, 73–97. <https://doi.org/10.1177/0023830910388014>
- Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *The Journal of the Acoustical Society of America*, 119(3), 1727–1739. <https://doi.org/10.1121/1.2161431>
- Pasdeloup, V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*. Institut de Phonétique d’Aix-en-Provence, Université de Provence.
- Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory Processing of Amplitude Envelope Rise Time in Adults Diagnosed With Developmental Dyslexia. *Scientific Studies of Reading*, 11(3), 259–286. <https://doi.org/10.1080/10888430701344280>
- Pater, J. (2003). The perceptual acquisition of Thai phonology by English speakers: Task and stimulus effects. *Second Language Research*, 19(3), 209–223.
- Paul, I., Bott, C., Heim, S., Wienbruch, C., & Elbert, T. R. (2006). Phonological but not auditory discrimination is impaired in dyslexia. *European Journal of Neuroscience*, 24(10), 2945–2953. <https://doi.org/10.1111/j.1460-9568.2006.05153.x>
- Paulesu, E., Demonet, J., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S., Cossu, G., Habib, M., Frith, C., & Frith, U. (2001). Dyslexia: Cultural Diversity and Biological Unity. *Science (New York, N.Y.)*, 291, 2165–2167. <https://doi.org/10.1126/science.1057179>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-Driven Computations in Speech Processing. *Science*, 298(5593), 604–607. <https://doi.org/10.1126/science.1072901>
- Peperkamp, S., & Dupoux, E. (2002). *A typological study of stress ‘deafness’* (C. Gussenhoven & N. Warner, Eds.). DE GRUYTER. <https://doi.org/10.1515/9783110197105.203>
- Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, 38(3), 422–430. <https://doi.org/10.1016/j.wocn.2010.04.001>
- Perrin, F., & García-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research*, 17(1), 36–47. [https://doi.org/10.1016/S0926-6410\(03\)00078-8](https://doi.org/10.1016/S0926-6410(03)00078-8)
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Thesis, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/16065>

- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the emergence of linguistic structure* (pp. 137–157). John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.45.08pie>
- Pierrehumbert, J., & Beckman, M. E. (1989). *Japanese tone structure*. The MIT Press. <https://www.cambridge.org/core/journals/journal-of-linguistics/article/abs/janet-b-pierrehumbert-and-mary-e-beckman-japanese-tone-structure-linguistic-inquiry-monograph-15-cambridge-ma-london-mit-press-1988-pp-xi-282/A68DD6C391EEAF8D3438ACA2F6815AF9>
- Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. <https://doi.org/10.1016/j.neuroimage.2019.05.026>
- Pisoni, D. B. (1979). On the Perception of Speech Sounds as Biologically Significant Signals. *Brain, Behavior and Evolution*, *16*(5–6), 330–350.
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, *41*(1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Polich, J. (1989). P300 from a passive auditory paradigm. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *74*(4), 312–320. [https://doi.org/10.1016/0168-5597\(89\)90061-0](https://doi.org/10.1016/0168-5597(89)90061-0)
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- Polich, J., Alexander, J., Bauer, L., Kuperman, S., Morzorati, S., O’Connor, S., Porjesz, B., Rohrbaugh, J., & Begleiter, H. (1997). P300 topography of Amplitude/Latency correlations. *Brain Topography*, *9*, 275–282. <https://doi.org/10.1007/BF01464482>
- Pons, F., & Bosch, L. (2010). Stress Pattern Preference in Spanish-Learning Infants: The Role of Syllable Weight. *Infancy*, *15*(3), 223–245. <https://doi.org/10.1111/j.1532-7078.2009.00016.x>
- Poser, W. (2000). *Phonetics and Phonology of Tone and Intonation in Japanese*. CSLI Publications.
- Post, B. (2002). *French Tonal Structures*.
- Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24*(4), 445–473. <https://doi.org/10.1006/jpho.1996.0024>
- Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*(4), 429–451. <https://doi.org/10.1006/jpho.1995.0032>
- Qin, Z., Chien, Y.-F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics*, *38*(3), 541–570. <https://doi.org/10.1017/S0142716416000321>
- Quené, H. (1992). Durational cues for word segmentation Dutch. *Journal of Phonetics*, *20*(3), 331–350. [https://doi.org/10.1016/S0095-4470\(19\)30638-2](https://doi.org/10.1016/S0095-4470(19)30638-2)
- Radeau, M., Mousty, P., & Bertelson, P. (1989). The effect of the uniqueness point in spoken-word recognition. *Psychological Research*, *51*(3), 123–128. <https://doi.org/10.1007/BF00309307>
- Ramus, F. (2014a). Neuroimaging sheds new light on the phonological deficit in dyslexia. *Trends in Cognitive Sciences*, *18*(6), 274–275. <https://doi.org/10.1016/j.tics.2014.01.009>
- Ramus, F. (2014b). Neuroimaging sheds new light on the phonological deficit in dyslexia. *Trends in Cognitive Sciences*, *18*(6), 274–275. <https://doi.org/10.1016/j.tics.2014.01.009>
- Ramus, F. (2014c). Should there really be a ‘Dyslexia debate’? *Brain*, *137*(12), 3371–3374. <https://doi.org/10.1093/brain/awu295>

- Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, 29, 104–122. <https://doi.org/10.1080/02643294.2012.677420>
- Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. J. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, 136(2), 630–645. <https://doi.org/10.1093/brain/aws356>
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, 105(1), 512–521. <https://doi.org/10.1121/1.424522>
- Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003a). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, 126(Pt 4), 841–865. <https://doi.org/10.1093/brain/awg076>
- Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003b). Theories of developmental dyslexia: Insights from a multiple case study of dyslexic adults. *Brain*, 126(4), 841–865. <https://doi.org/10.1093/brain/awg076>
- Ramus, F., & Szenkovits, G. (2008). What Phonological Deficit? *Quarterly Journal of Experimental Psychology*, 61(1), 129–141. <https://doi.org/10.1080/17470210701508822>
- Ransby, M. J., & Swanson, H. L. (2003). Reading Comprehension Skills of Young Adults with Childhood Diagnoses of Dyslexia. *Journal of Learning Disabilities*, 36(6). [https://journals.sagepub.com/doi/10.1177/00222194030360060501?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%200pubmed](https://journals.sagepub.com/doi/10.1177/00222194030360060501?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed)
- Raven, J. (2008). The Raven Progressive Matrices Tests: Their Theoretical Basis and Measurement Model. *Uses and Abuses of Intelligence: Studies Advancing Spearman and Raven's Quest for Non-Arbitrary Metrics*.
- Raven, J., & Court, J. H. (1996). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Standard progressive matrices*. Oxford Psychologists Press.
- Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, 48(2), 270–292. [https://doi.org/10.1016/0022-0965\(89\)90006-4](https://doi.org/10.1016/0022-0965(89)90006-4)
- Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, 63(4), 772–783. <https://doi.org/10.1080/17470210903104412>
- Renvall, H., & Hari, R. (2003). Diminished auditory mismatch fields in dyslexic adults. *Annals of Neurology*, 53(5), 551–557. <https://doi.org/10.1002/ana.10504>
- Rietveld, A. C. M. (1980). Word Boundaries in the French Language. *Language and Speech*, 23(3), 289–296. <https://doi.org/10.1177/002383098002300306>
- Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude: Investigating Meara's (2005) LLAMA tests. *EUROSLA Yearbook*, 16(1), 179–210. <https://doi.org/10.1075/eurosla.16.07rog>
- Rojczyk, A. (2013). Vowel Quality and Duration as a Cue to Word Stress for Non-native Listeners: Polish Listeners' Perception of Stress in English. In E. Waniek-Klimczak & L. R. Shockey (Eds.), *Teaching and Researching English Accents in Native and Non-native Speakers* (pp. 59–71). Springer. [https://doi.org/10.1007/978-3-642-24019-5\\_5](https://doi.org/10.1007/978-3-642-24019-5_5)
- Rosen, S. (1992). Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. *Philosophical Transactions: Biological Sciences*, 336(1278), 367–373.



- Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, 31(3), 509–527. [https://doi.org/10.1016/S0095-4470\(03\)00046-9](https://doi.org/10.1016/S0095-4470(03)00046-9)
- Rosen, S., & Manganari, E. (2001). Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *Journal of Speech, Language, and Hearing Research*, 44, 720–736. [https://doi.org/10.1044/1092-4388\(2001/057\)](https://doi.org/10.1044/1092-4388(2001/057))
- Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Demonet, J., Bricout, L., Billard, C., Nguyen-Morel, M.-A., Le Heuzey, M., Soares-Boucaud, I., George, F., Ziegler, J., & Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, 52. <https://doi.org/10.1037/dev0000184>
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90(1), 51–89. [https://doi.org/10.1016/S0010-0277\(03\)00139-2](https://doi.org/10.1016/S0010-0277(03)00139-2)
- Sampson, R. (2016). Sandhi phenomena. In A. Ledgeway & M. Maiden (Eds.), *The Oxford Guide to the Romance Languages* (p. 0). Oxford Academic. <https://doi.org/10.1093/acprof:oso/9780199677108.003.0040>
- Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474–494. <https://doi.org/10.1037/0096-3445.110.4.474>
- Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language*, 26(1), 36–56. [https://doi.org/10.1016/0749-596X\(87\)90061-1](https://doi.org/10.1016/0749-596X(87)90061-1)
- Sanders, L. D., & Neville, H. J. (2000). Lexical, Syntactic, and Stress-Pattern Cues for Speech Segmentation. *Journal of Speech, Language, and Hearing Research*, 43(6), 1301–1321. <https://doi.org/10.1044/jslhr.4306.1301>
- Sanders, L. D., & Neville, H. J. (2003a). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, 15(3), 228–240. [https://doi.org/10.1016/S0926-6410\(02\)00195-7](https://doi.org/10.1016/S0926-6410(02)00195-7)
- Sanders, L. D., & Neville, H. J. (2003b). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, 12(3), 214–227. [https://doi.org/10.1016/S0926-6410\(02\)00194-5](https://doi.org/10.1016/S0926-6410(02)00194-5)
- Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002a). Speech Segmentation by Native and Non-Native Speakers. *Journal of Speech, Language, and Hearing Research*, 45(3), 519–530. [https://doi.org/10.1044/1092-4388\(2002/041\)](https://doi.org/10.1044/1092-4388(2002/041))
- Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002b). Speech Segmentation by Native and Non-Native Speakers: The Use of Lexical, Syntactic, and Stress-Pattern Cues. *Journal of Speech, Language, and Hearing Research*, 45(3), 519–530. [https://doi.org/10.1044/1092-4388\(2002/041\)](https://doi.org/10.1044/1092-4388(2002/041))
- Sandhi Phenomena in the Languages of Europe. (1986). In H. K. Andersen (Ed.), *Sandhi Phenomena in the Languages of Europe* (Vol. 33). De Gruyter Mouton. <https://doi.org/10.1515/9783110858532>
- Schraeyen, K., Elst, W. V. der, Geudens, A., Ghesquière, P., & Sandra, D. (2019). Short-term memory problems for phonemes' serial order in adults with dyslexia: Evidence from a different analysis of the Nonword repetition task. *Applied Psycholinguistics*, 40(3), 613–644. <https://doi.org/10.1017/S0142716418000759>
- Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1999). Pre-attentive processing of auditory patterns in dyslexic human subjects. *Neuroscience Letters*, 276(1), 41–44. [https://doi.org/10.1016/S0304-3940\(99\)00785-5](https://doi.org/10.1016/S0304-3940(99)00785-5)

- Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (2001). Speech perception deficit in dyslexic adults as measured by mismatch negativity (MMN). *International Journal of Psychophysiology*, *40*(1), 77–87. [https://doi.org/10.1016/S0167-8760\(00\)00152-5](https://doi.org/10.1016/S0167-8760(00)00152-5)
- Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician*, *34*(4), 216–221. <https://doi.org/10.1080/00031305.1980.10483031>
- Sebastian Galles, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In *Twenty-First Century Psycholinguistics: Four Cornerstones* (p. (pp. 279-292)).
- Sebastian Galles, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, *31*, 18–32. [https://doi.org/10.1016/0749-596X\(92\)90003-G](https://doi.org/10.1016/0749-596X(92)90003-G)
- Serniclaes, W., Heghe, S. V., Mousty, P., Carré, R., & Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, *87*(4), 336–361. <https://doi.org/10.1016/j.jecp.2004.02.001>
- Serniclaes, W., & Sprenger-Charolles, L. (2015). *Reading impairment: From behavior to brain*.
- Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research: JSLHR*, *44*(2), 384–399. [https://doi.org/10.1044/1092-4388\(2001/032\)](https://doi.org/10.1044/1092-4388(2001/032))
- Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, *16*(2), 155–172. <https://doi.org/10.1017/S0142716400007062>
- Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, *68*(1), 1–16. <https://doi.org/10.3758/BF03193651>
- Shatzman, K., & McQueen, J. (2007). The modulation of lexical competition by segment duration. *Psychonomic Bulletin & Review*, *13*, 966–971. <https://doi.org/10.3758/BF03213910>
- Shaywitz, S. E., Shaywitz, B. A., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., Pugh, K. R., Holahan, J. M., Marchione, K. E., Fletcher, J. M., Lyon, G. R., & Gore, J. C. (2003). Neural systems for compensation and persistence: Young adult outcome of childhood reading disability. *Biological Psychiatry*, *54*(1), 25–33. [https://doi.org/10.1016/S0006-3223\(02\)01836-X](https://doi.org/10.1016/S0006-3223(02)01836-X)
- Shaywitz, S., Shaywitz, B., Pugh, K., Fulbright, R., Constable, R., Mencl, W., Shankweiler, D., Liberman, A., Skudlarski, P., Fletcher, J., Katz, L., Marchione, K., Lacadie, C., Gatenby, C., & Gore, J. (1998). Functional Disruption in the Organization of the Brain for Reading in Dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 2636–2641. <https://doi.org/10.1073/pnas.95.5.2636>
- Shibata, T., & Shibata, R. (1990). Akusento wa doo'ongo o donoteido benbetsu shiuruno ka?: Nihongo, eigo, chuugokugo no baai [How much can accent distinguish homophones?: Cases of Japanese, English and Chinese]. *Keiryoo Kokugo-Gaku [Mathematical Linguistics]*, *17*, 317–327.
- Shoemaker, E. (2009). *Acoustic Cues to Speech Segmentation in Spoken French: Native and Nonnative Strategies* [University of Texas at Austin]. <https://tel.archives-ouvertes.fr/tel-00676964>
- Shoemaker, E. (2014). Durational cues to word recognition in spoken French. *Applied Psycholinguistics*, *35*(2), 243–273. <https://doi.org/10.1017/S0142716412000380>

- Shoemaker, E., & Birdsong, D. (2008). La résolution de la liaison en français par des locuteurs natifs et non-natifs. *Acquisition et interaction en langue étrangère*, 27, Article 27. <https://doi.org/10.4000/aile.4023>
- Shoemaker, E. M. (2010). The exploitation of fine phonetic detail in the processing of L2 French. In B. VanPatten & J. Jegerski (Eds.), *Language Acquisition and Language Disorders* (Vol. 53, pp. 259–280). John Benjamins Publishing Company. <https://doi.org/10.1075/lald.53.11sho>
- Shulver, K., & Badcock, N. (2021). Chasing the Anchor: A Systematic Review and Meta-Analysis of Perceptual Anchoring Deficits in Developmental Dyslexia. *Journal of Speech, Language, and Hearing Research: JSLHR*, 64, 1–14. [https://doi.org/10.1044/2021\\_JSLHR-20-00533](https://doi.org/10.1044/2021_JSLHR-20-00533)
- Simons, R. F., Graham, F. K., Miles, M. A., & Chen, X. (2001). On the relationship of P3a and the Novelty-P3. *Biological Psychology*, 56(3), 207–218. [https://doi.org/10.1016/S0301-0511\(01\)00078-3](https://doi.org/10.1016/S0301-0511(01)00078-3)
- Siok, W. T., Niu, Z., Jin, Z., Perfetti, C. A., & Tan, L. H. (2008). A structural–functional basis for dyslexia in the cortex of Chinese readers. *Proceedings of the National Academy of Sciences*, 105(14), 5561–5566. <https://doi.org/10.1073/pnas.0801750105>
- Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, 12, 914–919. <https://doi.org/10.1111/j.1467-7687.2009.00835.x>
- Sluijter, A., & Van Heuven, V. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *ICSLP* (Vol. 2, p. 633 vol.2). <https://doi.org/10.1109/ICSLP.1996.607440>
- Smith, N. J., & Levy, R. (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 1637–1642. <https://www.semanticscholar.org/paper/Cloze-but-no-cigar%3A-The-complex-relationship-cloze%2C-Smith-Levy/9ab79d8daaf5e76a1e8c470b321f992339de6431>
- Smits, R., Warner, N., McQueen, J., & Cutler, A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *The Journal of the Acoustical Society of America*, 113, 563–574. <https://doi.org/10.1121/1.1525287>
- Snowling, M. J. (1995). Phonological processing and developmental dyslexia. *Journal of Research in Reading*, 18(2), 132–138. <https://doi.org/10.1111/j.1467-9817.1995.tb00079.x>
- Snowling, M. J. (2000). *Dyslexia* (2nd ed., pp. xiv, 253). Blackwell Publishing.
- Snowling, M. J. (2001). From language to reading and dyslexia1. *Dyslexia*, 7(1), 37–46. <https://doi.org/10.1002/dys.185>
- Snyder, J. S., & Weintraub, D. M. (2013). Loss and persistence of implicit memory for sound: Evidence from auditory stream segregation context effects. *Attention, Perception, & Psychophysics*, 75(5), 1059–1074. <https://doi.org/10.3758/s13414-013-0460-y>
- So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53(2), 273–293. <https://doi.org/10.1177/0023830909357156>
- So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36(2), 195–221. <https://doi.org/10.1017/S0272263114000047>

- Soroli, E., Szenkovits, G., & Ramus, F. (2010). Exploring dyslexics' phonological deficit III: Foreign speech perception and production. *Dyslexia*, *16*(4), 318–340. <https://doi.org/10.1002/dys.415>
- Soto, J., Rodriguez-Fornells, A., & Sebastian Galles, N. (2007). Stress Placement and word segmentation by Spanish speakers. *Psicológica: Revista de Metodología y Psicología Experimental*, *ISSN 0211-2159*, *Vol. 28*, *Nº 2*, 2007, *Pags. 167-176*, 28.
- Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and Suprasegmental Mismatch in Lexical Access. *Journal of Memory and Language*, *45*(3), 412–432. <https://doi.org/10.1006/jmla.2000.2783>
- Speciale, G., Ellis, N., & Bywater, T. (2004). *Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition*. *25*(2), 293–321. <https://doi.org/10.1017/S0142716404001146>
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*(2), 343–358. <https://doi.org/10.1111/1469-8986.3820343>
- Spinelli, E., Cutle, A., & McQueen, J. M. (2002). Resolution of liaison for lexical access in French. *Revue française de linguistique appliquée*, *VII*(1), 83–96. <https://doi.org/10.3917/rfla.071.0083>
- Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception, & Psychophysics*, *72*(3), 775–787. <https://doi.org/10.3758/APP.72.3.775>
- Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, *48*(2), 233–254. [https://doi.org/10.1016/S0749-596X\(02\)00513-2](https://doi.org/10.1016/S0749-596X(02)00513-2)
- Spinelli, E., Welby, P., & Schaegis, A.-L. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision. *Language and Cognitive Processes*, *22*(6), 828–859. <https://doi.org/10.1080/01690960601076472>
- Spivey, M., & Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*, 281–284.
- Sprenger-Charolles, L., & Colé, P. (2013). *Lecture et dyslexie: Approche cognitive* (2nd ed.). Dunod.
- Sprenger-Charolles, L., Colé, P., Piquard, A., & Leloup, G. (2005). Batterie informatisée d'évaluation diagnostique des troubles spécifiques d'apprentissage de la lecture (Traduction de 'French Normative Data on Reading and Related Skills from EVALEC, a New Computerized Battery of Tests' ERAP, 2005). *Revue Européenne de Psychologie Appliquée*.
- Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17. <https://doi.org/10.1016/j.jml.2015.02.004>
- Stein, J. (2018). What is Developmental Dyslexia? *Brain Sciences*, *8*(2), Article 2. <https://doi.org/10.3390/brainsci8020026>
- Steinhauer, K., & Connolly, J. F. (2008). CHAPTER 9—Event-Related Potentials in the Study of Language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the Neuroscience of Language* (pp. 91–104). Elsevier. <https://doi.org/10.1016/B978-0-08-045352-1.00009-4>
- Steinhauer, K., Royle, P., Drury, J. E., & Fromont, L. A. (2017). The priming of priming: Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neuroscience Letters*, *651*, 192–197. <https://doi.org/10.1016/j.neulet.2017.05.007>

- Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17(1), 3–45. [https://doi.org/10.1016/S0095-4470\(19\)31520-7](https://doi.org/10.1016/S0095-4470(19)31520-7)
- Stilp, C. (2018). Short-term, not long-term, average spectra of preceding sentences bias consonant categorization. *The Journal of the Acoustical Society of America*, 144(3\_Supplement), 1797. <https://doi.org/10.1121/1.5067927>
- Stilp, C. (2020). Acoustic context effects in speech perception. *WIREs Cognitive Science*, 11(1), e1517. <https://doi.org/10.1002/wcs.1517>
- Sussman, E. S. (2017). Auditory Scene Analysis: An Attention Perspective. *Journal of Speech, Language, and Hearing Research: JSLHR*, 60(10), 2989–3000. [https://doi.org/10.1044/2017\\_JSLHR-H-17-0041](https://doi.org/10.1044/2017_JSLHR-H-17-0041)
- Suzuki, Y. (2021). Probing the construct validity of llama\_d as a measure of implicit learning aptitude: Incidental instructions, confidence ratings, and reaction time. *Studies in Second Language Acquisition*, 43(3), 663–676. <https://doi.org/10.1017/S0272263120000704>
- Swanson, H. L., & Hsieh, C.-J. (2009). Reading Disabilities in Adults: A Selective Meta-Analysis of the Literature. *Review of Educational Research*, 79(4), 1362–1390.
- Szenkovits, G., Darma, Q., Darcy, I., & Ramus, F. (2016). Exploring dyslexics' phonological deficit II: Phonological grammar. *First Language*, 36(3), 316–337. <https://doi.org/10.1177/0142723716648841>
- Szenkovits, G., & Ramus, F. (2005a). Exploring dyslexics' phonological deficit I: Lexical vs sub-lexical and input vs output processes. *Dyslexia*, 11(4), 253–268. <https://doi.org/10.1002/dys.308>
- Szenkovits, G., & Ramus, F. (2005b). Exploring dyslexics' phonological deficit I: Lexical vs sub-lexical and input vs output processes. *Dyslexia*, 11(4), 253–268. <https://doi.org/10.1002/dys.308>
- Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, 22(3), 259–282. [https://doi.org/10.1016/0010-0277\(86\)90017-X](https://doi.org/10.1016/0010-0277(86)90017-X)
- Talcott, J. B., Witton, C., McLean, M. F., Hansen, P. C., Rees, A., Green, G. G. R., & Stein, J. F. (2000). Dynamic sensory sensitivity and children's word decoding skills. *Proceedings of the National Academy of Sciences*, 97(6), 2952–2957. <https://doi.org/10.1073/pnas.040546597>
- Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, 9(2), 182–198. [https://doi.org/10.1016/0093-934X\(80\)90139-X](https://doi.org/10.1016/0093-934X(80)90139-X)
- Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological Basis of Speech: A Case for the Preeminence of Temporal Processing. *Annals of the New York Academy of Sciences*, 682(1 Temporal Info), 27–47. <https://doi.org/10.1111/j.1749-6632.1993.tb22957.x>
- Taylor, W. L. (1953). “Cloze Procedure”: A New Tool for Measuring Readability. *Journalism Quarterly*, 30(4), 415–433. <https://doi.org/10.1177/107769905303000401>
- Temple, E., Poldrack, R. A., Protopapas, A., Nagarajan, S., Salz, T., Tallal, P., & Merzenich, M. M. (2000). Disruption of the neural response to rapid acoustic stimuli in dyslexia: Evidence from functional. *Proceedings of the National Academy of Sciences*, 97(25), 13907–13912. <https://doi.org/doi.org/10.1073/pnas.240461697>
- Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding Words and Rules in a Speech Stream: Functional Differences between Vowels and Consonants. *Psychological Science*, 19(2), 137–144.
- Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception & Psychophysics*, 70(8), 1515–1525. <https://doi.org/10.3758/PP.70.8.1515>

- Tremblay, A. (2011). Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism: Language and Cognition*, *14*, 257–279. <https://doi.org/10.1017/S1366728910000271>
- Tremblay, A., Broersma, M., & Coughlin, C. (2017). The functional weight of a prosodic cue in the native language predicts the learning of speech segmentation in a second language. *Bilingualism: Language and Cognition*, *21*, 1–13. <https://doi.org/10.1017/S136672891700030X>
- Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the Native Language on the Learning of Fundamental Frequency in Second-Language Speech Segmentation. *Frontiers in Psychology*, *7*, 985. <https://doi.org/10.3389/fpsyg.2016.00985>
- Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021a). Dutch listeners' perception of English lexical stress: A cue-weighting approach. *The Journal of the Acoustical Society of America*, *149*(6), 3703–3714. <https://doi.org/10.1121/10.0005086>
- Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021b). Dutch listeners' perception of English lexical stress: A cue-weighting approach. *The Journal of the Acoustical Society of America*, *149*(6), 3703–3714. <https://doi.org/10.1121/10.0005086>
- Tremblay, A., Cho, T., Kim, S., & Shin, S. (2019). Phonetic and phonological effects of tonal information in the segmentation of Korean speech: An artificial-language segmentation study. *Applied Psycholinguistics*, *40*(05), 1221–1240. <https://doi.org/10.1017/S0142716419000237>
- Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Laboratory Phonology*, *3*(2), 385–423. <https://doi.org/10.1515/lp-2012-0018>
- Tremblay, A., Kim, S., Shin, S., & Cho, T. (2021). Re-examining the effect of phonological similarity between the native- and second-language intonational systems in second-language speech segmentation. *Bilingualism: Language and Cognition*, *24*(2), 401–413. <https://doi.org/10.1017/S136672892000053X>
- Tremblay, A., & Spinelli, E. (2013). Segmenting Liaison-Initial Words: The role of predictive dependencies. *Language and Cognitive Processes*, *28*, 1093–1113. <https://doi.org/10.1080/01690965.2012.689306>
- Tremblay, A., & Spinelli, E. (2014). English Listeners' Use of Distributional and Acoustic-Phonetic Cues to Liaison in French: Evidence from Eye Movements. *Language and Speech*. <https://doi.org/10.1177/0023830913504569>
- Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367–376. <https://doi.org/10.1121/1.3129127>
- Vaissière, J. (1983). Language-Independent Prosodic Features. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (Vol. 14, pp. 53–65). Springer Verlag. [https://doi.org/10.1007/978-3-642-69103-4\\_5](https://doi.org/10.1007/978-3-642-69103-4_5)
- Vaissière, J. (1988). Prediction of velum movement from phonological specifications. *Phonetica*, *45*(2–4), 122–139. <https://doi.org/10.1159/000261822>
- Vaissière, J. (1997). Langues, prosodies et syntaxe: Prosodie et syntaxe. *Traitement Automatique des Langues*, *38*(1), 53–82.
- van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the Role of Lexical stress in Lexical Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 251–273. <https://doi.org/10.1080/02724980343000927>
- Van Hirtum, T., Ghesquière, P., & Wouters, J. (2019). Atypical neural processing of rise time by adults with dyslexia. *Cortex*, *113*, 128–140. <https://doi.org/10.1016/j.cortex.2018.12.006>

- van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, 24, 573–583. <https://doi.org/10.3758/BF03201084>
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18(4), 380–393. <https://doi.org/10.3758/BF03197127>
- Van Petten, C., Rubin, S., Parks, M., Plante, E., & Coulson, S. (2000). Time course of word identification and semantic integration in spoken language. *The Journal of the Acoustical Society of America*, 108(5), 2643–2643. <https://doi.org/10.1121/1.4743849>
- Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Golestani, N., Wouters, J., & Ghesquiere, P. (2010). Adults with dyslexia are impaired in categorizing speech and nonspeech sounds on the basis of temporal cues. *Proceedings of the National Academy of Sciences*, 107(23), 10389–10394. <https://doi.org/10.1073/pnas.0912858107>
- Varnet, L. (2015). *Identification des indices acoustiques utilisés lors de la compréhension de la parole dégradée* [Phdthesis, Université Claude Bernard - Lyon I]. <https://theses.hal.science/tel-01266326>
- Vasic, N., Lohr, C., Steinbrink, C., Martin, C., & Wolf, R. C. (2008). Neural correlates of working memory performance in adolescents and young adults with dyslexia. *Neuropsychologia*, 46(2), 640–648. <https://doi.org/10.1016/j.neuropsychologia.2007.09.002>
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2–40. <https://doi.org/10.1046/j.0021-9630.2003.00305.x>
- Verbaten, M. N., Huyben, M. A., & Kemner, C. (1997). Processing capacity and the frontal P3. *International Journal of Psychophysiology*, 25(3), 237–248. [https://doi.org/10.1016/S0167-8760\(96\)00748-9](https://doi.org/10.1016/S0167-8760(96)00748-9)
- Virtala, P., Talola, S., Partanen, E., & Kujala, T. (2020). Poor neural and perceptual phoneme discrimination during acoustic variation in dyslexia. *Scientific Reports*, 10(1), 8646. <https://doi.org/10.1038/s41598-020-65490-3>
- Voeten, C. C. (2022). *permutest: Permutation Tests for Time Series Data (2.5)* [Computer software]. <https://CRAN.R-project.org/package=permutest>
- Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., & Maj, M. (2007). The cortical generators of P3a and P3b: A LORETA study. *Brain Research Bulletin*, 73(4), 220–230. <https://doi.org/10.1016/j.brainresbull.2007.03.003>
- Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21(1), 98–108. <https://doi.org/10.1037/0096-1523.21.1.98>
- Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The Roles of Word Stress and Vowel Harmony in Speech Segmentation. *Journal of Memory and Language*, 38(2), 133–149. <https://doi.org/10.1006/jmla.1997.2548>
- Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, 67(6), 939–950. <https://doi.org/10.3758/BF03193621>
- Wang, H.-L. S., Huss, M., Hämäläinen, J. A., & Goswami, U. (2012). Basic auditory processing and developmental dyslexia in Chinese. *Reading and Writing*, 25(2), 509–536. <https://doi.org/10.1007/s11145-010-9284-5>

- Wang, J., Shu, H., Zhang, L., Liu, Z., & Zhang, Y. (2013). The roles of fundamental frequency contours and sentence context in Mandarin Chinese speech intelligibility. *The Journal of the Acoustical Society of America*, *134*, EL91-7. <https://doi.org/10.1121/1.4811159>
- Wang, Q. (2008). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study* [Thesis]. <https://dspace.library.uvic.ca/handle/1828/1282>
- Warner, N., Otake, T., & Arai, T. (2010). Intonational Structure as a Word-boundary Cue in Tokyo Japanese. *Language and Speech*, *53*(1), 107–131. <https://doi.org/10.1177/0023830909351235>
- Wauquier-Gravelines, S., & Braud, V. (2005). Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages*, *158*, 53–65.
- Weber, A. (2001). *Language-specific listening: The case of phonetic sequences* [University of Nijmegen Nijmegen, The Netherlands]. <https://doi.org/10.17617/2.68255>
- Weber, A., Braun, B., & Crocker, M. W. (2006). Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents. *Language and Speech*, *49*(3), 367–392. <https://doi.org/10.1177/00238309060490030301>
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. [https://doi.org/10.1016/S0749-596X\(03\)00105-0](https://doi.org/10.1016/S0749-596X(03)00105-0)
- Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, *119*(1), 597–607. <https://doi.org/10.1121/1.2141003>
- Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, *34*(3), 343–371. <https://doi.org/10.1016/j.wocn.2005.09.001>
- Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, *49*(1), 28–48. <https://doi.org/10.1016/j.specom.2006.10.005>
- Welby, P., & Niebuhr, O. (2016, May). The influence of F0 discontinuity on intonational cues to word segmentation: A preliminary investigation. *Speech Prosody*. <https://hal.archives-ouvertes.fr/hal-01452826>
- Welby, P. S. (2003). *The Slaying of Lady Mondegreen, being a Study of French Tonal Association and Alignment and their Role in Speech Segmentation*. The Ohio State University.
- Wells, J. C. (2008). *Longman pronunciation dictionary* (3rd ed.). Pearson/Longman.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. <https://doi.org/10.3758/BF03207136>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation Cues in Conversational Speech: Robust Semantics and Fragile Phonotactics. *Frontiers in Psychology*, *3*, 375. <https://doi.org/10.3389/fpsyg.2012.00375>
- Wiener, S., & Goss, S. (2019). Second and third language learners' sensitivity to Japanese pitch accent is additive: An information-based model of pitch perception. *Studies in Second Language Acquisition*, *41*(4), 897–910. <https://doi.org/10.1017/S0272263119000068>
- Wilcox, R. (2017). *Introduction to Robust Estimation and Hypothesis Testing—4th Edition: Vol. Statistical Modeling and Decision Science* (R. Wilcox, Ed.; 4th ed.). Academic Press.



<https://www.elsevier.com/books/introduction-to-robust-estimation-and-hypothesis-testing/wilcox/978-0-12-804733-0>

- Witton, C., Swoboda, K., Shapiro, L. R., & Talcott, J. B. (2020). Auditory frequency discrimination in developmental dyslexia: A meta-analysis. *Dyslexia*, 26(1), 36–51. <https://doi.org/10.1002/dys.1645>
- Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing*, 15, 179–206. <https://doi.org/10.1023/A:1013880723925>
- Wright, B. A., Bowen, R. W., & Zecker, S. G. (2000). Nonlinguistic perceptual deficits associated with reading and language disorders. *Current Opinion in Neurobiology*, 10(4), 482–486. [https://doi.org/10.1016/s0959-4388\(00\)00119-7](https://doi.org/10.1016/s0959-4388(00)00119-7)
- Wu, M. (2019). Effect of F0 contour on perception of Mandarin Chinese speech against masking. *PLoS ONE*, 14(1), e0209976. <https://doi.org/10.1371/journal.pone.0209976>
- Xia, Z., Hoeft, F., Zhang, L., & Shu, H. (2016). Neuroanatomical anomalies of dyslexia: Disambiguating the effects of disorder, performance, and maturation. *Neuropsychologia*, 81, 68–78. <https://doi.org/10.1016/j.neuropsychologia.2015.12.003>
- Zeng, Z., Liu, L., Tuninetti, Varghese, P., Tsao, F.-M., & Mattock, K. (2022). English and Mandarin native speakers' cue-weighting of lexical stress. Results from MMN and LDN. *Brain and Language*, 232. <https://doi.org/doi:10.1016/j.bandl.2022.105151>
- Zhang, J., Meng, Y., Wu, C., & Yuan, Z. (2023). Spoken word recognition across language boundary: ERP evidence of prosodic transfer driven by pitch. *Brain Sciences*, 13(202). <https://doi.org/10.3390/brainsci13020202>
- Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, 38(2), 260–271. <https://doi.org/10.1016/j.wocn.2009.11.002>
- Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review*, 5(4), 683–689. <https://doi.org/10.3758/BF03208845>
- Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, 48(4), 779–793. [https://doi.org/10.1016/S0749-596X\(03\)00006-8](https://doi.org/10.1016/S0749-596X(03)00006-8)
- Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, 12(5), 732–745. <https://doi.org/10.1111/j.1467-7687.2009.00817.x>
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology*, 86(3), 169–193. [https://doi.org/10.1016/S0022-0965\(03\)00139-5](https://doi.org/10.1016/S0022-0965(03)00139-5)
- Ziegler, J., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, 131, 3–29. <https://doi.org/10.1037/0033-2909.131.1.3>
- Zora, H., Schwarz, I.-C., & Heldner, M. (2015). Neural correlates of lexical stress: Mismatch negativity reflects fundamental frequency and intensity. *NeuroReport*, 26(13), 791–796. <https://doi.org/10.1097/WNR.0000000000000426>
- Zou, T., Caspers, J., & Chen, Y. (2022). Perception of Different Tone Contrasts at Sub-Lexical and Lexical Levels by Dutch Learners of Mandarin Chinese. *Frontiers in Psychology*, 13. Scopus. <https://doi.org/10.3389/fpsyg.2022.891756>
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32(1), 25–64. [https://doi.org/10.1016/0010-0277\(89\)90013-9](https://doi.org/10.1016/0010-0277(89)90013-9)

## **Annexes**

### ANNEX I. Stimuli list of Experiments 1 and 2 in Chapter 6

The written target displayed in Experiment 2 is shown under *Target presented*. Number of words, durations in seconds and speech rate of each sentence are given for congruent and incongruent conditions in Table 1. Written word frequency and spoken word frequency were extracted from the published corpus *Lexique* (New et al., 2001, 2004). DPs' genre, duration in seconds, position within sentences, and cloze probability are given in Table 2. Cloze Probability was calculated from responses of 30 participants (naïve to the study) who were asked to complete the sentences up until the DP of interest. They were asked to freely provide a response that would best complete each sentence.

**Table 1.** Complete list of sentences used in the semi-passive sentence judgement task (Experiment 1) and the word identification task (Experiment 2).

CONGRUENT CONDITION									
<i>Vowel initial (N=24)</i>					<i>Consonant initial (N=24)</i>				
Sentence	Target presented	Nb. of words	Sentence length (s)	Speech rate (words/s)	Sentence	Target presented	Nb. of words	Sentence length (s)	Speech rate (words/s)
Le publicitaire trouvait l'accroche trop simple	accroche	7	3.349	2.0902	Le pianiste a raté la croche dans le troisième mouvement de sa sonate	croche	13	5.447	2.3866
Le théâtre national a choisi l'affiche du spectacle	affiche	9	4.04	2.2277	La secrétaire médicale a perdu la fiche du patient	fiche	9	3.928	2.2912
L'informaticien fait une notice sur l'ajout de carte mémoire	ajout	11	4.378	2.5126	La grand-mère fait un bisou sur la joue du petit enfant	joue	12	3.483	3.4453

L'incendie a déclenché l'alarme de l'université	alarme	9	3.813	2.3603		La tragédie a provoqué la larme du public	larme	8	3.55	2.2535
Les pêcheurs tiennent l'amarre fermement	amarre	6	2.552	2.3511		Les nénuphars couvrent la mare du château	mare	7	2.937	2.3834
Le collégien a rencontré l'amie de Paul	amie	8	3.141	2.547		Les édentés apprécient la mie de pain	mie	7	2.645	2.6465
Le jeune photographe réparera l'appareil très bientôt	appareil	8	4.079	1.9613		La voisine dépannée rendra la pareille à ses amis	pareille	9	3.933	2.2883
Le général De Gaulle a prononcé l'appel du 18 juin	appel	11	4.1	2.6829		Le jardinier a cherché la pelle partout	pelle	7	2.645	2.6465
Le guérisseur pratique l'apposition des mains	apposition	7	2.967	2.3593		L'amiral est surpris par la position du navire	position	9	3.648	2.4671
Le taureau a abîmé les gradins de l'arène de Séville	arène	11	3.967	2.7729		Le royaume célèbre la reine défunte	reine	6	3.063	1.9589
Le chauffeur a raté l'arrêt pour aller plus vite	arrêt	10	3.233	3.0931		Le plongeur admire la raie du lagon	raie	7	2.883	2.428

Le brideur pose l'atout fièrement	atout	6	2.536	2.3659		Le sirop traite la toux grasse	toux	6	2.3	2.6087
Ma mère a recousu l'attache de mon cartable	attache	9	3.311	2.7182		Ma mère a nettoyé la tâche sur ma chemise	tâche	9	3.167	2.8418
Les objectifs préconisent l'atteinte du but	atteinte	7	3.362	2.0821		Le coiffeur change la teinte de ses cheveux	teinte	8	2.957	2.7054
Le candidat a supporté l'attente dans le bureau	attente	9	3.369	2.6714		Les campeurs montent la tente très vite	tente	7	2.547	2.7483
L'orateur demande l'attention de son public	attention	8	3.447	2.3209		L'infirmière prend la tension de son patient	tension	8	2.895	2.7634
Cette femme à barbe représente l'attraction principale du cirque	attraction	10	4.853	2.0606		Ce cheval de ferme fournit la traction nécessaire pour labourer	traction	10	4.652	2.1496
Le grammairien a étudié l'attribut du sujet et du verbe	attribut	11	4.093	2.6875		L'indien rejoint la tribu voisine avec défiance	tribu	8	3.654	2.1894
Le cirque embauche l'avaleur de sabre	avaleur	7	2.971	2.3561		Le marchand connaît la valeur du travail	valeur	7	2.897	2.4163

Les économies montrent l'avarice du vieil homme	avarice	8	3.238	2.4707		Le bas de contention évite la varice de la femme enceinte	varice	11	4.205	2.6159
Le maire a rebaptisé l'avenue des Champs-Élysées	avenue	9	3.474	2.5907		Le festival annonce la venue de l'acteur	venue	8	3.216	2.4876
L'odeur nauséabonde provoque l'aversion de cette femme	aversion	9	3.81	2.3622		Le juge doute de la version des faits	version	8	2.443	3.2747
Cet homme quitte le pays contre l'avis du juge	avis	10	3.65	2.7397		Le chroniqueur a raconté la vie de la starlette	vie	9	3.549	2.5359
L'ail donne l'haleine fétide	haleine	6	2.312	2.5952		La lessive rend la laine plus douce	laine	7	2.652	2.6395
<b>INCONGRUENT CONDITION</b>										
<i>Vowel initial (N=24)</i>						<i>Consonant initial (N=24)</i>				
<b>Sentence</b>	<b>Target presented</b>	<b>Nb. of words</b>	<b>Sentence length (s)</b>	<b>Speech rate (words/s)</b>		<b>Sentence</b>	<b>Target presented</b>	<b>Nb. of words</b>	<b>Sentence length (s)</b>	<b>Speech rate (words/s)</b>
Ce musicien jouait l'accroche trop vite	accroche	7	2.877	2.4331		Le journaliste a raté la croche dans la première page de son journal	croche	13	4.937	2.6332

Le comptable rempli l'affiche de ses employés	affiche	8	3.342	2.3938		Le réalisateur admire la fiche de son film	fiche	8	3.571	2.2403
Le bisou claque sur l'ajout de l'enfant	ajout	9	2.457	3.663		L'addition entraîne la joue d'unités	joue	7	2.563	2.7312
L'oignon lui donne l'alarme à l'œil	alarme	9	2.541	3.5419		Le pompier lui montre la larme d'incendie	larme	8	2.973	2.6909
Les canards envahissent l'amarre du jardin	amarre	7	2.886	2.4255		Les marins attrapent la mare du bateau	mare	7	2.55	2.7451
Le boulanger a découpé l'ami de pain	amie	8	2.812	2.845		La famille apprécie la mie du marié	mie	7	2.698	2.5945
Le boxeur rend l'appareil à son adversaire	appareil	8	3.435	2.329		La mutuelle rembourse la pareille dentaire de l'enfant	pareille	9	3.635	2.4759
La tarte se sert à l'appel, pas le gâteau	appel	10	2.989	3.3456		Le soldat manque à la pelle du matin	pelle	8	2.736	2.924

Le satellite donne l'apposition du bateau	apposition	7	3.12	2.2436		Le magnétiseur a soigné par la position des mains	position	9	3.611	2.4924
Le prince a ajusté la couronne de l'arène d'Angleterre	arène	11	4.222	2.6054		Le gladiateur imagine la reine de Rome	reine	7	3.172	2.2068
Le poissonnier a découpé l'arrêt pour faire des filets	arrêt	10	3.546	2.8201		Le bus rate la raie de l'église	raie	8	2.689	2.9751
Le médecin de campagne écoute l'atout du bébé	atout	9	3.319	2.7117		Le joueur de carreaux garde la toux pour la fin	toux	10	3.101	3.2248
La lessive attaque l'attache de gras	attache	7	2.699	2.5936		Cette valise a la tâche cassée	tâche	6	2.417	2.4824
Les pigments permettent l'atteinte des textiles	atteinte	7	3.143	2.2272		Le bilan évalue la teinte des objectifs	teinte	7	3.101	2.2573
Le boy scout a monté l'attente dans la montagne	attente	9	3.365	2.6746		Les patients trouvent la tente trop longue	tente	7	2.839	2.4657



Cette corde d'escalade supporte de l'attention permanente	attention	9	4.328	2.0795		Cet exercice de maths demande de la tension soutenue	tension	9	3.872	2.3244
L'ancienne Citroën est l'attraction préférée du garagiste	attraction	9	4.39	2.0501		Le grand huit est la traction préférée des adolescents	traction	9	3.657	2.461
L'ethnologue a étudié l'attribut des indiens iroquois	attribut	9	4.009	2.2449		Cet adjectif est la tribu du sujet dans la phrase	tribu	10	3.603	2.7755
L'orfèvre estime l'avaleur des bijoux	avaleur	7	2.739	2.5557		Ce fakir connaît la valeur de sabre	valeur	7	2.861	2.4467
Le chirurgien opère l'avarice de la vieille dame	avarice	9	3.659	2.4597		Le sage des montagnes craint la varice des hommes de la ville	varice	12	4.046	2.9659
Le président a annoncé l'avenue du chef d'état	avenue	10	3.688	2.7115		Le chauffeur remonte la venue de Verdun	venue	7	3.113	2.2486
Le cinéophile averti préfère l'aversion française	aversion	7	4.028	1.7378		L'écolier a de la version pour les maths	version	9	2.912	3.0907

Le magasin propose un slogan contre l'avis cher	avis	9	3.728	2.4142		L'huissier a demandé la vie d'imposition	vie	8	3.063	2.6118
La grand-mère tricote l'haleine de mouton	haleine	8	2.828	2.8289		Le dentifrice rend la laine plus fraîche	laine	7	3.3	2.1212

**Table 2.** Measures of homophonic DPs within sentences (genre, DPs duration in seconds, DPs' position within sentences, and cloze probability).

CONGRUENT CONDITION										
<i>Vowel initial</i>					<i>Consonant initial</i>					
Target presented	Genre	DP length (s)	Position in sentence (s)	Cloze probability	Target presented	Genre	DP length (s)	Position in sentence (s)	Cloze probability	
accroche	f	0.672	1.6494	0.0313	croche	f	0.806	1.5737	0	
affiche	f	0.608	2.4444	0	fiche	f	0.584	2.5404	0.0938	
ajout	f	0.468	2.6725	0	joue	f	0.403	1.97	0.5938	
alarme	f	0.691	1.8795	0.1563	laine	f	0.596	1.1909	0	
amarre	m	0.548	1.2863	0	larme	f	0.613	1.8795	0.0625	
amie	f	0.451	1.9484	0.1563	mare	f	0.551	1.5433	0.1875	
appareil	m	0.696	2.4743	0.5938	mie	f	0.4	1.6653	0	
appel	f	0.661	2.3623	0.0938	pareille	f	0.788	2.1935	0.2188	
apposition	f	0.874	1.6254	0	pelle	f	0.559	1.5167	0.2188	
arène	f	0.621	2.431	0.4688	position	f	0.87	1.8729	0	
arrêt	f	0.487	1.4869	0.2500	raie	f	0.539	1.6124	0	
atout	m	0.427	1.3844	0	reine	f	0.595	1.6673	0.1250	
attache	m	0.706	1.4287	0	tâche	f	0.609	1.3492	0	
atteinte	f	0.644	1.925	0	teinte	f	0.651	1.3295	0	

attente	f	0.739	1.6761	0		tension	f	0.691	1.1506	0.1563
attention	f	0.786	1.541	0.2188		tente	f	0.728	1.0936	0.6875
attraction	m	0.778	2.3434	0		toux	f	0.417	1.2253	0.7188
attribut	f	0.604	1.8344	0		traction	f	0.711	2.1726	0
avaleur	f	0.598	1.4655	0		tribu	f	0.602	1.0989	0.3125
avarice	f	0.776	1.5697	0		valeur	f	0.675	1.2466	0.0313
avenue	f	0.678	1.8543	0		varice	f	0.794	2.1699	0
aversion	f	0.736	2.2122	0		venue	f	0.698	1.6298	0
avis	f	0.408	2.5012	0.0313		version	f	0.588	1.3203	0
haleine	f	0.648	0.9726	0.5313		vie	f	0.409	1.9159	0.1250
<b>INCONGRUENT CONDITION</b>										
<i>Vowel initial</i>						<i>Consonant initial</i>				
<b>Target presented</b>	<b>Genre</b>	<b>Sequence length (s)</b>	<b>Position in sentence (s)</b>	<b>Cloze probability</b>		<b>Target presented</b>	<b>Genre</b>	<b>Sequence length (s)</b>	<b>Position in sentence (s)</b>	<b>Cloze probability</b>
accroche	f	0.76	1.2657	0		croche	f	0.735	2.0156	0
affiche	f	0.622	1.5667	0		fiche	f	0.54	1.3377	0
ajout	f	0.358	1.3737	0		joue	f	0.381	1.6971	0
alarme	f	0.614	1.1889	0		laine	f	0.624	1.4125	0
amarre	m	0.603	1.6135	0		larme	f	0.61	1.3377	0
amie	f	0.4	1.8694	0		mare	f	0.592	1.2895	0
appareil	m	0.669	1.3346	0		mie	f	0.388	1.5738	0.0313
appel	f	0.677	1.4319	0		pareille	f	0.763	1.6837	0
apposition	f	0.918	1.5138	0		pelle	f	0.645	1.5167	0
arène	f	0.65	2.6785	0		position	f	0.877	2.2713	0
arrêt	f	0.501	1.8149	0		raie	f	0.475	1.212	0
atout	m	0.475	2.0309	0		reine	f	0.513	1.9184	0
attache	m	0.648	1.4698	0		tâche	f	0.57	1.3037	0
atteinte	f	0.662	1.4537	0		teinte	f	0.644	1.3385	0

attente	f	0.691	1.5353	0		tension	f	0.609	2.5258	0
attention	f	0.721	2.643	0		tente	f	0.622	1.315	0
attraction	m	0.725	1.8757	0		toux	f	0.441	1.8888	0
attribut	f	0.666	1.8056	0		traction	f	0.832	1.1359	0.0625
avaleur	f	0.615	1.4151	0		tribu	f	0.57	1.4329	0
avarice	f	0.725	1.6767	0		valeur	f	0.683	1.3211	0
avenue	f	0.695	1.8543	0		varice	f	0.787	1.9081	0
aversion	f	0.638	2.4649	0		venue	f	0.648	1.6746	0.0625
avis	f	0.399	2.7463	0.0313		version	f	0.798	1.1776	0
haleine	f	0.572	1.5739	0		vie	f	0.536	1.5334	0

**Table 3.** Acoustic measures of onset /l/ and /a/ of homophonic DPs within sentences. Duration in milliseconds and F0 mean of /l/. Duration, F0 mean, and formants 1 and 2 mean values.

CONGRUENT CONDITION													
<i>Consonant initial</i>							<i>Vowel initial</i>						
item	<i>/l/</i>		<i>/a/</i>				item	<i>/l/</i>		<i>/a/</i>			
	Duration (ms)	F0 mean (Hz)	Duration (ms)	F0 mean (Hz)	F1 mean (Hz)	F2 mean (Hz)		Duration (ms)	F0 mean (Hz)	Duration (ms)	F0 mean (Hz)	F1 mean (Hz)	F2 mean (Hz)
accroche	59.5	154.83	87	177.75	699.1	1916.6	croche	41.5	180.37	113.9	163.49	625.5	1916.4
affiche	87.8	171.78	100.5	188.21	658.3	1961.6	fiche	61.6	169.43	74.7	151.6	607.6	1984.3
ajout	110.9	163.74	97.8	183.07	629.2	1973.6	joue	58.4	160.84	105.3	157.25	593.6	1860.5
alarme	88.4	175.62	96.7	197.71	762.3	1906.7	larme	58.3	177.28	98.9	156.56	717.2	1844.4
amarre	96.4	150.07	80.7	181.43	738.2	1842.9	mare	80.6	177.55	99.9	170.48	755.5	1667
amie	92.5	174.54	84.4	183.67	653.9	2066.3	mie	82.3	171.58	90.3	151.67	614.1	2037.3
appareil	89.9	179.63	75.3	198.38	741.1	1769.5	pareille	63.7	167.25	76.6	156.61	755.5	1766.5
appel	91.2	161	75.4	175.87	680.1	1847.6	pelle	64.9	184.93	71.6	171.67	714.2	1788

<b>apposition</b>	80.1	151.67	64	172.79	652.5	1820.9	<b>position</b>	67.6	170.87	63.9	172.87	698.9	1751.6
<b>arène</b>	114.9	156.66	90.7	181.93	842.3	1626.7	<b>reine</b>	105.8	161.22	118.1	161.56	793.3	1720.4
<b>arrêt</b>	80.8	158.1	136.5	175.5	784.8	1718.3	<b>raie</b>	149	170.7	133.4	160.46	779.3	1706
<b>atout</b>	67.4	164.96	83	177.65	603.5	1961.6	<b>toux</b>	80.3	178.7	90.6	160.93	562.9	1874.2
<b>attache</b>	78.6	155.48	85.4	177.36	657.2	1986	<b>tâche</b>	64.8	176.4	82.4	156.04	600.8	1948.7
<b>atteinte</b>	72.5	171.1	96.3	183.84	643.9	1842.4	<b>teinte</b>	97.4	167.92	80.1	161.58	691.4	1747.6
<b>attente</b>	95.6	159.65	93.6	191.56	701.2	1937	<b>tente</b>	84.4	186.81	96.4	167.57	698	1908.6
<b>attention</b>	93.3	164.84	87.8	198.61	692.9	1730.2	<b>tension</b>	42.9	162.01	84.7	157.94	661.6	1852.7
<b>attraction</b>	94.3	183.62	82.3	198.88	681.2	1853.6	<b>traction</b>	45.6	180.8	79.3	163.49	610.2	1898
<b>attribut</b>	78.7	162.08	82.4	182.56	641.8	1950.4	<b>tribu</b>	60.6	191.61	79.5	164.68	602.8	1872
<b>avaleur</b>	107.9	192.42	80.3	203.4	749.3	1787.9	<b>valeur</b>	110.6	176.63	83.7	159	680	1814.8
<b>avarice</b>	45.8	168.74	93.1	185.96	690.6	1812.2	<b>varice</b>	67.5	151.35	86.1	151.16	670.6	1790.8
<b>avenue</b>	93.6	160.92	87.8	188.53	669.8	1853.5	<b>venue</b>	89.8	163.54	117.1	153.42	652.2	1814.8
<b>aversion</b>	77.6	161.5	91.9	185.4	651.8	1884.1	<b>version</b>	60.8	149.07	68.8	142.83	726.7	1859.5
<b>avis</b>	99.1	167.69	99.4	174.43	583.2	1954	<b>vie</b>	69.7	199.2	117.7	178.99	644.4	1967.1
<b>haleine</b>	80	154.9	94.9	168.62	670	2044.2	<b>laine</b>	57.7	185.61	112.9	159.56	663.7	2007.8

**INCONGRUENT CONDITION**

<i>Consonant initial</i>							<i>Vowel initial</i>						
<b>item</b>	<i>/l/</i>		<i>/a/</i>				<b>item</b>	<i>/l/</i>		<i>/a/</i>			
	<b>Duration (ms)</b>	<b>F0 mean (Hz)</b>	<b>Duration (ms)</b>	<b>F0 mean (Hz)</b>	<b>F1 mean (Hz)</b>	<b>F2 mean (Hz)</b>		<b>Duration (ms)</b>	<b>F0 mean (Hz)</b>	<b>Duration (ms)</b>	<b>F0 mean (Hz)</b>	<b>F1 mean (Hz)</b>	<b>F2 mean (Hz)</b>
<b>accroche</b>	102.9	159.86	80.6	178.04	667.5	1822.9	<b>croche</b>	62.1	176.63	96.8	159.42	602.2	1924.1
<b>affiche</b>	53.4	172.2	85	187.41	673.6	1946.6	<b>fiche</b>	34	172.3	90.5	161.52	574.2	1976.7
<b>ajout</b>	82.9	163.62	102.1	172.42	623.6	1933.1	<b>joue</b>	108.2	162.2	107.3	174.41	629.2	1856.7
<b>alarme</b>	120.7	156.69	99.7	175.87	709.7	1952.5	<b>larme</b>	88.2	151.7	92.8	150.6	707.4	1955
<b>amarre</b>	112.6	154.87	91.6	181.03	788.9	1756.3	<b>mare</b>	147.7	152.39	99.3	170.96	767.4	1687.5
<b>amie</b>	72.6	161.9	97	178.79	690.7	1984.7	<b>mie</b>	91.4	187.49	76.5	157.87	638	1977.4
<b>appareil</b>	103.2	173.91	67.7	199.55	758.1	1732.7	<b>pareille</b>	74.9	168.87	81.6	155.74	740.6	1747.5

<b>appel</b>	84.4	164.04	83.4	190.11	728.6	1925	<b>pelle</b>	76.2	161.72	71.2	158.56	691.5	1882.2
<b>apposition</b>	88	156.61	65.7	182.59	686.7	1891.8	<b>position</b>	78.5	152.48	58.2	153.66	689.4	1633.9
<b>arène</b>	53.4	167.19	88.1	186.46	834.1	1653.6	<b>reine</b>	47.6	163.92	108.7	161.57	795	1727.1
<b>arrêt</b>	80.9	162.48	122.9	180.54	830.5	1764.1	<b>raie</b>	107.2	164.03	137.2	160.16	777.8	1648.7
<b>atout</b>	90.1	166.41	88.1	185.91	636.6	1967.4	<b>toux</b>	80.3	178.7	90.6	160.93	562.9	1874.2
<b>attache</b>	70.9	165.62	86.8	176.28	652	1904.3	<b>tâche</b>	89	158.82	89.3	155.93	631.5	1968
<b>atteinte</b>	101.1	160.81	100	187.4	654.4	1910	<b>teinte</b>	122.7	157.85	90.2	155.88	665.2	1941.3
<b>attente</b>	87.2	161.58	92.1	191.73	702.7	1937.7	<b>tente</b>	91.4	179.55	100.6	154.46	652.2	1877.3
<b>attention</b>	56.4	164.83	90.8	186.68	651.5	1905.2	<b>tension</b>	39.5	170.9	77.7	152.86	563.8	1857.4
<b>attraction</b>	62.3	178.81	77.8	189.27	688.2	1878.3	<b>traction</b>	78.4	155.66	84.8	152.94	716.9	1784.4
<b>attribut</b>	112.6	165.15	81.1	191.09	638.5	1951.8	<b>tribu</b>	55.3	201.15	74	178.78	603.2	1831.7
<b>avaleur</b>	106.8	189.63	78.3	195.43	723.9	1807.6	<b>valeur</b>	118.8	167.89	79.8	158.32	680	1798.3
<b>avarice</b>	91.4	183.7	98.1	189.6	744.2	1742.3	<b>varice</b>	79.5	173.19	97.2	164.61	733.5	1723.3
<b>avenue</b>	92.6	157.8	96.3	190.02	690.5	1800.7	<b>venue</b>	91.7	165.71	111.1	155.91	674	1798.3
<b>aversion</b>	81.9	160.49	86.4	185.25	654.6	1898.5	<b>version</b>	75.8	150.1	67.2	142.67	729.8	1856.2
<b>avis</b>	101.6	167.97	91.2	184.79	653.8	1901.7	<b>vie</b>	69.9	197.04	102.7	179.57	641.6	1950.8
<b>haleine</b>	76.8	155.1	93.8	168.54	671.4	2045.2	<b>laine</b>	57.7	185.61	112.9	159.56	663.7	2007.8

**Table 4.** Acoustic measures of the second vowel of homophonic DPs within sentences. Duration in milliseconds, F0 mean, and mean formants 1 and 2.

CONGRUENT CONTEXT											
<i>vowel initial</i>						<i>consonant initial</i>					
item	vowel	Duration (ms)	F0 mean (Hz)	F1 mean (Hz)	F2 mean (Hz)	item	vowel	Duration (ms)	F0 mean (Hz)	F1 mean (Hz)	F2 mean (Hz)
accroche	o	171	183.82	528.3	1355.5	croche	o	155.6	181.27	547.6	1434.8
affiche	i	108.8	218.43	348.6	2433.6	fiche	i	98.7	186.02	329.4	2464.5

ajout	u	76.9	194.54	343.4	1225.6		joue	u	67.6	192.8	336.2	1318.3
alarme	a	152.5	201.62	858.6	1699.8		laine	ε	151.3	188.13	592.5	2206.7
amarre	a	143.4	188.55	825.1	1619.9		larme	a	127.8	189.69	846.2	1698.3
amie	i	133.2	205.9	336.4	2516.5		mare	a	151.8	208.31	886.4	1462.1
appareil	a	97.8	199.8	795.4	1717.6		mie	i	79.4	199.27	315.4	1574.7
appel	ε	161.3	191.78	562.1	2175.2		pareille	a	91.9	198.46	793.8	1604.5
apposition	o	66.8	196.88	391.9	1671.8		pelle	ε	129.3	187.93	567.2	2156.5
arène	ε	131.7	204.96	690.1	2094.1		position	o	71.2	218.34	402.2	1505.7
arrêt	ε	139.2	192.95	640.3	2096.6		raie	ε	121.3	188.06	601.4	2093.6
atout	u	100.2	190.74	351.7	1085.4		reine	ε	125.3	191.54	688.7	1976.6
attache	a	188.1	185.61	694.4	1913.5		tâche	a	162.2	192.96	691.9	1900.2
atteinte	a	237.4	203.64	583.1	1509		teinte	a	238.7	197.53	646.2	1505.4
attente	a	294.7	206.63	639.4	1561.3		tension	a	102.7	191.43	552.5	1468.8
attention	a	110.7	198.52	356.9	919		tente	a	232.4	195.03	611.5	1555.3
attraction	a	60.1	210.32	711.3	1877.5		toux	u	61.1	214.11	356.3	1257.4
attribut	i	67.4	200.54	361.6	2559		tribu	i	53.5	199.7	363	2364.4
avaleur	a	87.3	194.02	639	1864.5		valeur	a	75.6	190.44	648.6	1816.4
avarice	a	113	186.14	746.8	1652.8		varice	a	87.4	180.48	748	1601.2
avenue	e	105.1	191.14	409.2	1942.8		venue	e	80.2	179.08	419.2	1894.8
aversion	e	82.9	189.7	602.8	2041.5		version	e	67.9	170.14	548.6	2099.5
avis	i	80.4	212.14	321.2	2365.1		vie	i	92.5	205.68	339.2	2442
haleine	ε	152.9	174.63	525.2	2268.7		laine	ε	151.3	188.13	592.5	2206.7
<b>INCONGRUENT CONTEXT</b>												
<i>vowel initial</i>						<i>consonant initial</i>						
<b>item</b>	<b>vowel</b>	<b>Duration (ms)</b>	<b>F0 mean (Hz)</b>	<b>F1 mean (Hz)</b>	<b>F2 mean (Hz)</b>		<b>item</b>	<b>vowel</b>	<b>Duration (ms)</b>	<b>F0 mean (Hz)</b>	<b>F1 mean (Hz)</b>	<b>F2 mean (Hz)</b>
accroche	o	156.9	193.62	557.2	1298.5		croche	o	167.8	187.79	575.8	1284.9
affiche	i	115.7	215.79	447	2388.3		fiche	i	77.7	194.84	323.6	2407.3

ajout	u	54.5	201.73	354.5	1355		joue	u	39	196.47	322	1350.9
alarme	a	90.1	197.2	818.9	1687.4		laine	ε	146.5	194.01	559.9	2185.8
amarre	a	200.4	185.56	874.4	1631.1		larme	a	104.3	195.29	807.7	1581.9
amie	i	91	202.3	289.4	1237.9		mare	a	114.8	205.28	879.7	1492.4
appareil	a	99.1	210.12	814.3	1690.3		mie	i	72.3	197.84	327.6	868.6
appel	ε	151	195.59	575.7	2201.3		pareille	a	111.1	196.98	782.1	1597.7
apposition	o	82.4	206.02	412	1513		pelle	ε	128.4	183.64	575.6	2250.4
arène	ε	154.8	204.14	677.9	2154.4		position	o	72.9	199.18	381.3	1460.4
arrêt	ε	173.8	193.06	609.4	2142.5		raie	ε	117.4	190.18	649.9	2027
atout	u	81.5	215.25	339.4	1000.4		reine	ε	119.8	191.4	696.9	1977.4
attache	a	140.5	202.26	664.9	1998		tâche	a	133.9	193.77	716	1864.3
atteinte	a	293.6	209.84	477.5	1289.2		teinte	a	231.1	208.84	431.4	1270.1
attente	a	290.6	206.68	626.3	1517.5		tension	a	107.3	192.89	259.8	953.7
attention	a	100.8	183.38	466.8	1184		tente	a	233.5	179.28	602.7	1591.2
attraction	a	50.4	200.6	725	1870.6		toux	a	60.2	208.58	743.8	1790.1
attribut	i	61.3	216.12	386.8	2471.4		tribu	i	57.5	215.11	532.3	2268.2
avaleur	a	78.2	188.67	623	1870		valeur	a	83.4	189.89	643.2	1807.7
avarice	a	107.4	181.6	796.7	1728.3		varice	a	105.8	197.51	766.1	1655.4
avenue	e	104.1	186.29	409.5	1899.3		venue	e	87.8	182.81	414.9	1937.4
aversion	e	73.7	189.05	618.8	2045		version	e	62.3	169.82	544.9	2104.4
avis	i	68.6	204.14	309.4	2425.4		vie	i	177.6	208.68	287.6	2464.5
haleine	ε	147.3	183.5	568.6	2280.1		laine	ε	146.5	194.01	559.9	2185.8

**Table 5.** Acoustic measures of the second consonant of homophonic DPs within sentences (duration in milliseconds and F0 mean).

CONGRUENT CONDITION							
<i>vowel initial</i>				<i>consonant initial</i>			
item	consonant	Duration (ms)	F0 mean	item	consonant	Duration (ms)	F0 mean



			(Hz)				(Hz)
accroche	k	92.6	180.1	croche	k	93.6	152.88
accroche	κ	90.8	175.39	croche	κ	98.5	170.1
affiche	f	138.5	191.69	fiche	f	166.6	137.26
ajout	ʒ	153.3	180.42	joue	ʒ	141.7	159.68
alarme	l	74.3	199.98	larme	l	83.6	160.14
amarre	m	102.3	186.46	mare	m	104.8	178.92
amie	m	111.2	193.49	mie	m	121	157.5
appareil	p	115.4	209.02	pareille	p	124.9	170.04
appel	p	165.1	194.68	pelle	p	144.9	163.57
apposition	p	128.8	177.71	position	p	139.7	262.3
arene	κ	96.7	176.77	reine	κ	86	159.46
arrêt	κ	85.9	169.35	raie	κ	104.5	160.71
atout	t	146.8	169.88	toux	t	155.8	182.46
attache	t	131.3	188.02	tâche	t	139.9	170.15
atteinte	t	133.7	208.56	teinte	t	136	153.97
attente	t	124	202.44	tente	t	137.8	171.39
attention	t	100.3	212.52	tension	t	122.3	172.74
attraction	t	79.4	203.04	traction	t	84.3	170.4
attraction	κ	79.8	229.08	traction	κ	90.5	197.34
attribut	t	99.4	186.85	tribu	t	89.1	179.57
attribut	κ	86.9	217.1	tribu	κ	88.1	483.97
avaleur	v	76.8	191.55	valeur	v	105	160.37
avarice	v	88.4	179.22	varice	v	103.5	158.07
avenue	v	117	186.46	venue	v	117.9	150.5
aversion	v	60.8	181.96	version	v	84.7	140.96
avis	v	99.8	180.57	vie	v	99.3	169.89
haleine	l	80.9	170.55	laine	l	80.1	161.52

INCONGRUENT CONDITION							
<i>vowel initial</i>				<i>consonant initial</i>			
item	consonant	Duration (ms)	F0 mean (Hz)	item	consonant	Duration (ms)	F0 mean (Hz)
accroche	k	94	175.01	croche	k	98.3	151.77
accroche	ʁ	119.3	91	croche	ʁ	115.7	194.44
affiche	f	141	188.91	fiche	f	157.8	146.81
ajout	ʒ	88.5	175.13	joue	ʒ	98.6	168.22
alarme	l	70.3	190.22	larme	l	97.9	154.02
amarre	m	98.4	185.17	mare	m	109	177.96
amie	m	113.2	191.46	mie	m	118.1	166.57
appareil	p	116.8	217.75	pareille	p	117.4	182.97
appel	p	150.8	202.65	pelle	p	161.2	283.37
apposition	p	113.1	203.07	position	p	140.6	164.02
arene	ʁ	98.6	181.7	reine	ʁ	93.3	159.58
arrêt	ʁ	74.1	177.64	raie	ʁ	82	157.38
atout	t	173.6	199.38	toux	t	159	136.33
attache	t	146.3	188.15	tâche	t	135	161.83
atteinte	t	124.1	220.45	teinte	t	140.7	173.43
attente	t	128.2	201.06	tente	t	138.6	166.32
attention	t	101.3	193.72	tension	t	124.1	187.26
attraction	t	79.4	188.29	traction	t	120.5	147.81
attraction	ʁ	83.3	184.67	traction	ʁ	82.7	98.94
attribut	t	93.7	196.09	tribu	t	91.7	168.12
attribut	ʁ	89.7	268.16	tribu	ʁ	93.6	198.62
avaleur	v	86	183.31	valeur	v	95.2	159.53
avarice	v	96.5	179.81	varice	v	94.2	161.42
avenue	v	119.2	188.23	venue	v	117.5	153.95

aversion	v	78.2	183.76	version	v	84.4	141.72
avis	v	104.7	182.22	vie	v	113.9	168.02
haleine	l	100.4	175.95	laine	l	91.3	164.05

### ANNEX II. Stimuli list of Experiment 3 in Chapter 7

Genre, duration in seconds, written and spoken frequencies, and number of syllables of each item is given in Table 6.

**Table 6.** Complete list of stimuli used in the verbal repetition task.

<i>Vowel initial (N=30)</i>						<i>Consonant initial (N=30)</i>					
Item	Genre	Duration	Written frequency	Spoken frequency	Nb. of syllables	Item	Genre	Duration	Written frequency	Spoken frequency	Nb. of syllables
l'accroche	f	0.609	0.2	0.45	2	la croche	f	0.657	0.34	0.07	1
l'affiche	f	0.625	8.38	5.38	2	la fiche	f	0.676	7.57	7.45	1
l'ajout	m	0.503	0.2	0.73	2	la joue	f	0.477	46.35	25.57	1
l'alarme	f	0.679	6.35	16.71	2	la larme	f	0.656	10.81	5.15	1
l'amarre	f	0.621	1.42	0.58	2	la marre	f	0.592	9.86	3.66	1
l'amie	f	0.423	54.32	113.54	2	la mie	f	0.531	5.68	1.39	1
l'appareil	m	0.732	35.88	44.2	3	la pareille	f	0.717	4.19	3	2
l'appel	m	0.594	56.69	80.88	2	la pelle	f	0.616	11.35	8.75	1
l'apposition	f	0.839	0.14	0.01	4	la position	f	0.838	53.85	55.24	3
l'arène	f	0.698	4.12	2.25	2	la reine	f	0.672	30	56.26	1
l'arrêt	m	0.506	46.82	46.8	2	la raie	f	0.543	7.5	1.71	1
l'atout	m	0.482	2.84	3.66	2	la toux	f	0.531	12.23	4.94	1
l'attache	f	0.617	3.99	1.82	2	la tâche	f	0.6	33.92	12.61	1
l'atteinte	f	0.628	5.95	1.83	2	la teinte	f	0.613	7.43	0.6	1
l'attente	f	0.605	61.35	22.77	2	la tente	f	0.755	19.12	14.4	1
l'attention	f	0.745	119.66	156.9	3	la tension	f	0.752	14.93	20.05	2
l'attraction	f	0.82	5.74	4.96	3	la traction	f	0.812	7.43	1.06	2

l'attribut	m	0.699	1.69	0.08	3	la tribu	f	0.711	13.58	7.96	2
l'avaleur	m	0.584	0.41	0.17	3	la valeur	f	0.721	40.74	32.48	2
l'avarice	f	0.702	3.18	1.18	3	la varice	f	0.803	0	0.01	2
l'avenue	f	0.758	40.81	8.19	3	la venue	f	0.593	11.49	8.34	2
l'aversion	f	0.737	2.57	0.97	3	la version	f	0.735	11.01	19.1	2
l'avis	m	0.416	65.14	139.22	2	la vie	f	0.518	835.47	986.59	1
l'haleine	f	0.67	21.82	7.55	2	la laine	f	0.618	34.86	6.27	1
l'affiliation	f	1.054	0.14	0.26	4	la filiation	f	1.006	1.89	0.23	3
l'allocation	f	0.783	0.41	1.35	4	la location	f	0.819	5	5.29	3
l'allocation	f	0.902	3.11	0.31	4	la locution	f	0.901	0.41	0.04	3
l'annotation	f	0.834	0.14	0.19	4	la notation	f	0.848	0.54	0.19	3
l'apesanteur	f	0.931	0.81	0.93	4	la pesanteur	f	0.922	6.69	1.12	3
l'approbation	f	0.923	9.19	2.55	4	la probation	f	0.949	0.14	1.39	3

### ANNEX III. Stimuli list of Experiment 4 in Chapter 8

Durations were measured with Praat using an automatic script. A manual inspection validated such measures. Further lexico-grammatical information of items (genre, written and spoken frequency, and number of syllables) was taken from the French database *Lexique* (New et al., 2001, 2004). Genre, number of syllables and both written and spoken frequencies are provided for A tokens below in Table 7. Genre, written and spoken frequencies, and number of syllables, are the same for both A and X items. Therefore, only duration measures for X tokens are shown in Table 8.

**Table 7.** Complete list of stimuli used in the AX task. Genre, duration in seconds, written and spoken frequencies, and number of syllables of each item for A items.

<i>Vowel initial (N=30)</i>						<i>Consonant initial (N=30)</i>					
Item	Genre	Duration	Written frequency	Spoken frequency	Nb. of syllables	Item	Genre	Duration	Written frequency	Spoken frequency	Nb. of syllables
l'accroche	f	0.609	0.2	0.45	2	la croche	f	0.657	0.34	0.07	1
l'affiche	f	0.625	8.38	5.38	2	la fiche	f	0.676	7.57	7.45	1

l'ajout	m	0.503	0.2	0.73	2		la joue	f	0.477	46.35	25.57	1
l'alarme	f	0.679	6.35	16.71	2		la larme	f	0.656	10.81	5.15	1
l'amarre	f	0.621	1.42	0.58	2		la marre	f	0.592	9.86	3.66	1
l'amie	f	0.423	54.32	113.54	2		la mie	f	0.531	5.68	1.39	1
l'appareil	m	0.732	35.88	44.2	3		la pareille	f	0.717	4.19	3	2
l'appel	m	0.594	56.69	80.88	2		la pelle	f	0.616	11.35	8.75	1
l'apposition	f	0.839	0.14	0.01	4		la position	f	0.838	53.85	55.24	3
l'arène	f	0.698	4.12	2.25	2		la reine	f	0.672	30	56.26	1
l'arrêt	m	0.506	46.82	46.8	2		la raie	f	0.543	7.5	1.71	1
l'atout	m	0.482	2.84	3.66	2		la toux	f	0.531	12.23	4.94	1
l'attache	f	0.617	3.99	1.82	2		la tâche	f	0.6	33.92	12.61	1
l'atteinte	f	0.628	5.95	1.83	2		la teinte	f	0.613	7.43	0.6	1
l'attente	f	0.605	61.35	22.77	2		la tente	f	0.755	19.12	14.4	1
l'attention	f	0.745	119.66	156.9	3		la tension	f	0.752	14.93	20.05	2
l'attraction	f	0.82	5.74	4.96	3		la traction	f	0.812	7.43	1.06	2
l'attribut	m	0.699	1.69	0.08	3		la tribu	f	0.711	13.58	7.96	2
l'aveur	m	0.584	0.41	0.17	3		la valeur	f	0.721	40.74	32.48	2
l'avarice	f	0.702	3.18	1.18	3		la varice	f	0.803	0	0.01	2
l'avenue	f	0.758	40.81	8.19	3		la venue	f	0.593	11.49	8.34	2
l'aversion	f	0.737	2.57	0.97	3		la version	f	0.735	11.01	19.1	2
l'avis	m	0.416	65.14	139.22	2		la vie	f	0.518	835.47	986.59	1
l'haleine	f	0.67	21.82	7.55	2		la laine	f	0.618	34.86	6.27	1
l'affiliation	f	1.054	0.14	0.26	4		la filiation	f	1.006	1.89	0.23	3
l'allocation	f	0.783	0.41	1.35	4		la location	f	0.819	5	5.29	3
l'allocation	f	0.902	3.11	0.31	4		la locution	f	0.901	0.41	0.04	3
l'annotation	f	0.834	0.14	0.19	4		la notation	f	0.848	0.54	0.19	3
l'apesanteur	f	0.931	0.81	0.93	4		la pesanteur	f	0.922	6.69	1.12	3
l'approbation	f	0.923	9.19	2.55	4		la probation	f	0.949	0.14	1.39	3
<i>Distractors (N=58)</i>												
la belote	f	0.69	3.99	0.26	2		la pelote	f	0.626	4.19	0.94	2
la bile	f	0.67	4.66	2.59	1		la pile	f	0.609	0.07	0	1

la boule	f	0.668	38.31	19.29	1		la poule	f	0.706	16.69	23.5	1
la cachette	f	0.776	14.59	11.21	2		la gachette	#N/D	0.763	#N/D	#N/D	#N/D
la classe	f	0.649	90.74	70.46	1		la glace	f	0.569	76.01	58.09	1
la crasse	f	0.652	13.11	3.16	1		la grâce	f	0.715	49.66	32.08	1
la fente	f	0.688	10.54	3.61	1		la vente	f	0.623	12.97	20.93	1
la foi	f	0.566	76.49	54.06	1		la voix	f	0.552	612.7	130.83	1
la marine	f	0.854	11.76	4.73	2		la narine	f	0.826	5.14	1.11	2
la motion	f	0.71	0.68	2.87	2		la notion	f	0.745	10.61	4.99	2
le bain	m	0.601	43.11	50.52	1		le pain	m	0.611	99.32	62.81	1
le champ	m	0.648	51.76	38.05	1		le sang	m	0.595	205.2	304.3	1
le baquet	m	0.689	3.65	0.46	2		le paquet	m	0.712	62.77	36.9	2
le billard	m	0.702	11.35	7.56	2		le pillard	m	0.761	0.47	0.15	2
le blanc	m	0.614	59.73	27.56	1		le plan	m	0.572	67.84	119.54	1
le bonnet	m	0.602	14.66	6.62	2		le poney	m	0.582	0.47	3.7	2
le bottin	m	0.712	0.68	1.05	2		le potin	m	0.703	1.55	0.61	2
le boulet	m	0.608	3.78	1.99	2		le poulet	m	0.602	14.53	32.33	2
le carreau	m	0.605	13.51	4.34	2		le garrot	m	0.688	2.3	1.55	2
le cocher	m	0.679	4.12	2.1	2		le gaucher	m	0.702	0.14	0.8	2
le cou	m	0.542	112.7	43.71	1		le goût	m	0.501	124.8	50.51	1
le cri	m	0.567	71.55	18.79	1		le gris	m	0.677	25.74	2.83	1
le crin	m	0.648	3.92	0.32	1		le grain	m	0.562	24.26	10.74	1
le cube	m	0.663	5.74	1.58	1		le tube	m	0.698	11.35	12.16	1
le faisceau	m	0.708	6.82	1.27	2		le vaisseau	m	0.739	7.23	67.11	2
le faon	m	0.591	0.54	0.25	1		le paon	m	0.581	3.85	0.6	1
le filage	m	0.725	0.14	0.06	2		le village	m	0.649	118.24	87.6	2
le vent	m	0.491	207.64	71.5	1		le banc	m	0.521	48.31	8.96	1

**Table 8.** Complete list of X tokens used in the AX task with their respective durations in seconds.

<i>Vowel initial (N=30)</i>	<i>Consonant initial (N=30)</i>		<i>Distractors (N=58)</i>
-----------------------------	---------------------------------	--	---------------------------

Item	Duration (s)	Item	Duration (s)		Item	Duration (s)	Item	Duration (s)
l'accroche	0.617	la croche	0.663		la belote	0.672	la pelote	0.669
l'affiche	0.624	la fiche	0.659		la bile	0.690	la pile	0.674
l'ajout	0.502	la joue	0.492		la boule	0.715	la poule	0.743
l'alarme	0.682	la larme	0.65		la cachette	0.745	la gachette	0.862
l'amarre	0.603	la marre	0.598		la classe	0.693	la glace	0.634
l'amie	0.419	la mie	0.541		la crasse	0.710	la grâce	0.740
l'appareil	0.71	la pareille	0.718		la fente	0.745	la vente	0.645
l'appel	0.601	la pelle	0.616		la foi	0.596	la voix	0.573
l'apposition	0.838	la position	0.838		la marine	0.883	la narine	0.835
l'arène	0.698	la reine	0.672		la motion	0.787	la notion	0.767
l'arrêt	0.508	la raie	0.541		le bain	0.666	le pain	0.634
l'atout	0.479	la toux	0.529		le champ	0.603	le sang	0.651
l'attache	0.609	la tâche	0.614		le baquet	0.715	le paquet	0.686
l'atteinte	0.615	la teinte	0.614		le billard	0.737	le pillard	0.773
l'attente	0.608	la tente	0.759		le blanc	0.623	le plan	0.594
l'attention	0.748	la tension	0.761		le bonnet	0.669	le poney	0.671
l'attraction	0.821	la traction	0.816		le bottin	0.772	le potin	0.709
l'attribut	0.699	la tribu	0.713		le boulet	0.650	le poulet	0.614
l'avaleur	0.576	la valeur	0.714		le carreau	0.633	le garrot	0.732
l'avarice	0.702	la varice	0.813		le cocher	0.692	le gaucher	0.698
l'avenue	0.751	la venue	0.589		le cou	0.575	le goût	0.533
l'aversion	0.737	la version	0.735		le cri	0.517	le gris	0.685
l'avis	0.438	la vie	0.518		le crin	0.734	le grain	0.569
l'haleine	0.67	la laine	0.622		le cube	0.653	le tube	0.709
l'affiliation	1.056	la filiation	0.994		le faisceau	0.765	le vaisseau	0.783
l'allocation	0.788	la location	0.819		le faon	0.643	le paon	0.612
l'allocation	0.889	la locution	0.909		le filage	0.795	le village	0.744
l'annotation	0.827	la notation	0.857		le vent	0.536	le banc	0.566
l'apesanteur	0.945	la pesanteur	0.928					
l'approbation	0.923	la probation	0.946					

**Table 9.** Vowel properties of items (height, backness, F0, and F1 and F2) for first (/a/) and second vowels.

<i>vowel initial</i>							<i>consonant initial</i>						
item	vowel	vowel height	vowel backness	F0 (Hz)	F1 (Hz)	F2 (Hz)	item	vowel	vowel height	vowel backness	F0 (Hz)	F1 (Hz)	F2 (Hz)
accroche	a	high	back	201.3	698.3	1831.1	croche	a	high	back	177.9	721.7	1547.4
	o	high	back	228	668.2	1391.5		o	high	back	226.9	673.8	1243.5
affiche	a	high	front	218	683.9	1231.1	fiche	a	high	front	184.6	710	1573
	i	high	front	255.8	303.8	2235.9		i	high	front	257.3	301.3	2372.8
affiliation	a	high	front	215.4	640.1	1131.8	filiation	a	high	front	185.8	711.9	1059.4
	i	high	front	242.5	337.4	2171.2		i	high	front	243	356.9	2159.7
ajout	a	high	back	190.5	683	1420.4	joue	a	high	back	186.1	671.4	951.5
	u	high	back	227.2	373.2	963.8		u	high	back	237.3	433	786.7
alarme	a	low	front	196.4	791.9	1699.1	larme	a	low	front	178.9	743.5	1752.1
	a	low	front	235	878.8	1405.7		a	low	front	210.8	933.8	1487.4
allocation	a	high	back	216.3	754.4	1855.7	location	a	high	back	189.1	682.6	1079.1
	o	high	back	220.9	465.5	1159.4		o	high	back	216.4	449.2	1173.3
allocution	a	high	back	209	739.6	1873.3	locution	a	high	back	169.8	697.1	1879.1
	o	high	back	246.3	488.5	1239.9		o	high	back	207.8	433.1	1166.4
amarre	a	low	front	186.9	794.3	1237.4	marre	a	low	front	178.6	744.9	1759.4
	a	low	front	209.4	765	1301.2		a	low	front	208.4	783.1	1021.1
amie	a	high	front	202.2	655.4	1779.3	mie	a	high	front	180	710.7	1841.7
	i	high	front	257.4	345.6	779.9		i	high	front	254.1	401.8	2581.7
annotation	a	high	back	190.8	675	1356.5	notation	a	high	back	177	668.8	751.4
	o	high	back	229.5	458.4	1391.6		o	high	back	214	438.7	1367.7
apesanteur	a	high	front	211.1	754.6	1297.8	pesanteur	a	high	front	180.5	689.9	1686
	e	high	front	231.9	449.2	1554.1		e	high	front	221.2	438.5	1576.4



appareil	a	low	front	183.4	742.9	1503.3		pareille	a	low	front	178.5	746.6	1463.5
	a	low	front	202.2	962.6	974			a	low	front	203.2	940.2	998.7
appel	a	high	front	201.1	689	1136.2		pelle	a	high	front	187.3	716.9	1315.9
	e	high	front	221.6	655.3	1369.1			e	high	front	230	664.1	2042.8
apposition	a	high	back	192.9	653.1	1743.2		position	a	high	back	183.2	697.9	1755.1
	o	high	back	226.5	443.4	1380.2			o	high	back	224.2	445.9	1356.3
approbation	a	high	back	210.7	749.4	1640.6		probation	a	high	back	186.9	738.1	1683.2
	o	high	back	247.4	498.8	931.4			o	high	back	243.5	488.6	907.6
arène	a	high	front	204.2	1000.1	1603.3		reine	a	high	front	181.3	898.9	1170.9
	e	high	front	226.9	719.9	1313.5			e	high	front	214.9	665.6	1461.4
arrêt	a	high	front	187.6	905.4	1358.7		raie	a	high	front	179.9	787.2	898.5
	e	high	front	210.1	490.6	1212.8			e	high	front	223.1	448.7	1994.1
atout	a	high	back	206.6	640.7	1603.8		toux	a	high	back	190.2	717.7	1916.8
	u	high	back	243.5	272.7	552.8			u	high	back	256.2	342.9	915.1
attache	a	low	front	196.4	741.2	1393.9		tâche	a	low	front	180.1	714.1	1457.5
	a	low	front	214.8	735	1309			a	low	front	224.6	748.3	1800.3
atteinte	a	low	front	180.8	590.6	991		teinte	a	low	front	187.5	746	1734.2
	a	low	front	207.5	565.9	1156.2			a	low	front	219.8	592.5	1248.8
attente	a	low	front	204.8	694.5	988.1		tente	a	low	front	180.2	710.6	1579.2
	a	low	front	226.9	565.7	963.7			a	low	front	225.5	588.2	962.8
attention	a	low	front	213.7	682.2	1380.4		tension	a	low	front	186.1	679.3	1529.9
	a	low	front	244	700.8	1451.8			a	low	front	207.5	551.1	997.1
attraction	a	low	front	204.5	791.7	947.5		traction	a	low	front	176.4	697.7	1593
	a	low	front	223.2	793.9	1067.1			a	low	front	212.2	779.2	1081.7
attribut	a	high	front	198.2	658.6	1223.3		tribu	a	high	front	193.7	635.9	1403.5
	i	high	front	218.2	387.6	2420.5			i	high	front	246.6	373.4	2279.5
avaleur	a	low	front	185.4	715.1	1443.8		valeur	a	low	front	174	727.8	1016.5
	a	low	front	198.6	177.2	761.6			a	low	front	200	776.9	982.6

avarice	a	low	front	204.4	744.4	1057.8		varice	a	low	front	174.1	602.9	773.8
	a	low	front	209.7	859	1621.9			a	low	front	199.2	908.6	1172.6
avenue	a	high	back	210.2	662.6	1576.5		venue	a	high	back	189.6	622.7	1173
	e	high	back	216.7	442.4	1823.5			e	high	back	210.5	428.8	1793.9
aversion	a	high	front	202.4	676.8	1350.9		version	a	high	front	184.4	690.2	952.5
	e	high	front	218.1	633.8	1583.3			e	high	front	203.3	607.9	1399.4
avis	a	high	front	198.7	622.9	1942.6		vie	a	high	front	179.1	633.1	1231.2
	i	high	front	238.6	322.1	2326.6			i	high	front	268.4	327	2450.2
haleine	a	high	front	191.8	770.6	1590.8		laine	a	high	front	178.7	730.7	1747.8
	e	high	front	208.6	628.4	2068.3			e	high	front	202.5	719.8	1320.5

**Table 10.** F0 and duration (in milliseconds) of the initial /l/ and /a/ of DPs.

	<i>vowel initial</i>				<i>consonant initial</i>				
	/l/		/a/			/l/		/a/	
	<b>F0 (Hz)</b>	<b>duration (ms)</b>	<b>F0 (Hz)</b>	<b>duration (ms)</b>		<b>F0 (Hz)</b>	<b>duration (ms)</b>	<b>F0 (Hz)</b>	<b>duration (ms)</b>
accroche	192.8	72.8	199.3	62.8	croche	185.1	69.4	178	70.7
affiche	203.6	72.1	215	74.3	fiche	190.2	54	183.9	94.6
affilication	205.8	66.7	214.6	57.7	filiation	189.7	79.3	186.2	66.7
ajout	186.3	74.4	192.3	113.3	joue	189.9	66.69	186.8	89.43
alarme	187.7	88.1	197.5	88.1	larme	188.1	49.77	180	91.5
allocation	193.2	70.6	201.3	78.9	location	194.7	45	189.9	91.96
allocution	201.4	67.7	209.7	80.8	locution	180.5	68.6	170.5	87.2
amarre	187.4	77.3	188.5	78.2	marre	184.3	58.9	179.3	73.4
amie	193.7	63.5	203.2	88	mie	189.1	65	182.2	97.4
annotation	190.9	56	191.8	74	notation	185.6	47	177.1	70
apesanteur	201.1	61	210.9	61	pesanteur	190.42	47	181.5	58

appareil	180.2	64.8	182.7	61.8	pareille	183.6	60.8	178.7	57
appel	196.9	65.3	199.5	69.6	pelle	196.6	50	189.5	69.6
apposition	187.3	56	193.8	58	position	187.4	40.8	183.4	55.9
approbation	198.7	61	209	63.3	probation	191.9	63.6	187.2	47.4
arène	196.8	81.9	203.5	104.1	reine	188.5	77	181.3	100.6
arrêt	177.2	72.8	184.6	124.3	raie	185.4	63.2	180	118.5
atout	188.4	66.4	202.4	89.6	toux	194.6	66.8	191.3	70.6
attache	190.9	71.5	195	80.3	tâche	190.4	57.7	181.7	67.7
atteinte	174.3	66	176.7	70.2	teinte	195.1	72.4	188.1	78.4
attente	198.6	55	203.2	80.4	tente	186.4	45	180.7	81.1
attention	201.2	67	212.3	58.5	tension	192.1	42.1	186.8	74.9
attraction	191.4	66.6	203.6	86.7	traction	181.5	56.3	177.2	57.6
attribut	186.4	62.1	196.9	62	tribu	199.4	58	193.9	74.7
avaleur	176.6	62.9	183.4	74	valeur	176.4	55.6	174	76.78
avarice	194.7	57	203.6	77.1	varice	182.9	73	174.6	88.34
avenue	193.9	79.9	205.4	90.8	venue	194.8	44.35	189.5	85.56
aversion	191.9	74.2	200.1	72	version	184.7	48.5	184.1	87.57
avis	184	80	196.9	83.6	vie	187.3	77.36	180.4	86.3
haleine	185.4	83.8	191.7	89.2	laine	188.5	66.8	180.6	92.19

**Table 11.** Acoustic measures of the first consonant of DP nouns (duration in milliseconds and F0 (Hz)).

<i>vowel initial</i>				<i>consonant initial</i>			
item	consonant	Duration (ms)	F0 (Hz)	item	consonant	Duration (ms)	Mean (Hz)
accroche	kʁ	171.4	217.89	croche	kʁ	183.3	192.83
affiche	f	157	228.81	fiche	f	186.6	185.31

affiliation	f	108.4	202.19		filiation	f	113.7	170.81
ajout	з	110.5	194.96		joue	з	130.4	187.2
alarme	l	93.8	213.9		larme	l	89.2	192.67
allocation	l	72.8	231.55		location	l	72.3	197.47
allocution	l	69.2	225.39		locution	l	82.2	182.67
amarre	m	98.4	199.25		marre	m	104.6	191.68
amie	m	96.4	230.45		mie	m	110.3	197.81
annotation	n	72.9	209.77		notation	n	63.7	195.47
apesanteur	p	137.8	220.58		pesanteur	p	117.6	185.32
appareil	p	90.4	205.8		pareille	p	117.9	169.26
appel	p	110.2	224.78		pelle	p	149.4	222.89
apposition	p	109.1	196.09		position	p	126.7	191.57
approbation	pв	168	224.9		probation	pв	181.4	201.07
arène	к	94.7	212.51		reine	к	109.2	189.37
arrêt	к	82	196.27		raie	к	128.3	189.32
atout	t	133	248.25		toux	t	191.8	199.35
attache	t	111.9	224.81		tâche	t	150.5	202.31
atteinte	t	82.6	217.59		teinte	t	144.7	201.28
attente	t	103.3	232.58		tente	t	168.6	200.61
attention	t	97.1	246.53		tension	t	111.4	186.65
attraction	tв	132.8	218		traction	tв	155	175.12
attribut	tв	106.5	205.5		tribu	tв	113.7	191.6
avaleur	v	64.8	191.78		valeur	v	100.4	174.73
avarice	v	99.9	200.06		varice	v	109.2	179.03
avenue	v	101	204.4		venue	v	120	186.2
aversion	v	63.3	196.78		version	v	83.8	177.4
avis	v	83.8	200.28		vie	v	136.3	184.81

haine	1	103.5	198.26		laine	1	93.7	186.32
-------	---	-------	--------	--	-------	---	------	--------

## QUESTIONNAIRE CONFIDENTIEL



N° du participant :

Créneau horaire du participant :

Questionnaire confidentiel

1. Quelle âge avez-vous ? .....
2. Quelle est votre langue maternelle ?  
.....
3. Quelle(s) autre(s) langue(s) parlez-vous de façon fluide ?  
.....
4. Souffrez-vous ou avez-vous souffert des troubles neurologiques et/ou psychiatriques?  
(épilepsie, trauma crânien, dépression...)  
.....
5. Combien d'heures avez-vous dormi la semaine dernière ? (+ / - que d'habitude)  
.....
6. Comment vous sentez-vous ? (+ / - bien que d'habitude)  
.....
7. Quelle quantité de café / thé avez-vous bu depuis ce matin ?  
.....
8. Avez-vous pris des substances telles que de l'alcool ou médicaments dans les dernières 24h ?  
.....
9. Avez-vous des déficiences visuelles qui ne peuvent pas être corrigées par des lunettes/lentilles ?  
.....
10. Avez-vous des pertes ou déficiences auditives ?  
.....

Indiquez votre préférence manuelle, droite, gauche ou les deux, pour chacune des activités ci-dessous :

1	<b>Ecrire</b>	Droite	Gche
2	<b>Dessiner</b>	Droite	Gche
3	<b>Coudre (main tenant l'aiguille)</b>	Droite	Gche
4	<b>Tenir une paire de ciseaux</b>	Droite	Gche
5	<b>Se brosser les dents</b>	Droite	Gche
6	<b>Tenir un couteau</b>	Droite	Gche
7	<b>Tenir un balai (main supérieure)</b>	Droite	Gche
8	<b>Tenir une cuillère</b>	Droite	Gche
9	<b>Allumer une allumette (main tenant l'allumette)</b>	Droite	Gche
10	<b>Ouvrir une boîte (main tenant le couvercle)</b>	Droite	Gche

**QUESTIONNAIRE BIOGRAPHIQUE : LOCUTEUR NATIF**

Ce questionnaire porte sur vos expériences linguistiques tout au long de votre vie. Toutes les réponses sont confidentielles et votre anonymat sera strictement respecté. Votre nom et votre adresse électronique peuvent être utiles au cas où il nous faudra de plus amples informations. S'il y a des questions auxquelles vous préférez ne pas répondre, vous pouvez les omettre.

Nom et prénom : \_\_\_\_\_

Adresse e-mail : \_\_\_\_\_

Date de naissance : \_\_\_\_\_

Lieu de naissance : \_\_\_\_\_

Genre : M F Autre

1) Quelles langues parlez-vous et depuis combien de temps ?

---



---



---



---



---

2) Combien de temps avez-vous passé dans un pays étranger ? Veuillez indiquer le(s) pays et l'année et temps total de séjour.

---



---



---

3) Combien de temps avez-vous passé dans aux pays francophones ? Veuillez indiquer le(s) pays et l'année et temps total de séjour.

---



---



---

**MERCI DE VOTRE PARTICIPATION !**

## ANNEX VI

### BIOGRAPHICAL QUESTIONNAIRE: NON-NATIVE

This questionnaire is about your language experience over the course of your lifetime. Please feel free to elaborate when you think it may be helpful to our study. If there are any questions you would rather not answer, please feel free to ignore them.

All responses are confidential. Your name, phone number, and email address will be helpful in case we need additional information but will not be shared with anyone.

Name and surname: \_\_\_\_\_

Date of birth: \_\_\_\_\_

Place of birth: \_\_\_\_\_

Email: \_\_\_\_\_

Gender: M F Other

1) What is the highest level of formal education you have completed?

\_\_\_\_\_

2) List all the languages you speak in order of acquisition (native language first) and the age at which you started learning that language (age of exposure).

Language 1. \_\_\_\_\_ Age of exposure   0  

Language 2. \_\_\_\_\_ Age of exposure \_\_\_\_\_

Language 3. \_\_\_\_\_ Age of exposure \_\_\_\_\_

Language 4. \_\_\_\_\_ Age of exposure \_\_\_\_\_

Language 5. \_\_\_\_\_ Age of exposure \_\_\_\_\_

3) In your learning of languages, what percentage do you think you learned through **formal instruction** (e.g., school, language courses, books, etc.) and what percentage do you think you learned in more **informal settings** (e.g., interactions with other people, watching TV or movies, etc.)? (For each language, your percentages should add up to 100%).

Language 1. Formal: \_\_\_\_\_ Informal: \_\_\_\_\_

Language 2. Formal: \_\_\_\_\_ Informal: \_\_\_\_\_

Language 3. Formal: \_\_\_\_\_ Informal: \_\_\_\_\_

Language 4. Formal: \_\_\_\_\_ Informal: \_\_\_\_\_

Language 5. Formal: \_\_\_\_\_ Informal: \_\_\_\_\_



4) Please indicate the approximate periods during which you formally studied **French**. Circle "school" or "university" as appropriate.

At school/university/another type of course, I studied French from \_\_\_\_\_ to \_\_\_\_\_.

At school/university/another type of course, I studied French from \_\_\_\_\_ to \_\_\_\_\_

At school/university/another type of course, I studied French from \_\_\_\_\_ to \_\_\_\_\_

At school/university/another type of course, I studied French from \_\_\_\_\_ to \_\_\_\_\_

TOTAL = \_\_\_\_\_ years \_\_\_\_\_ months

5) How long have you been living in the France or in a French-speaking country? (Please indicate the country).