THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse - Jean Jaurès

Présentée et soutenue par :
Noémie te Rietmolen
le Mercredi 27 février, 2019

Titre :

Neural signature of metrical stress processing in French

École doctorale et discipline ou spécialité :
ED CLESCO : Sciences du langage

Unité de recherche :
URI Octogone-Lordat (EA 4156), CerCo (CNRS UMR 5549)

Directeur/trice(s) de Thèse :
Corine Astésano, MCF-HDR, URI Octogone-Lordat, Université Toulouse - Jean Jaurès
Simon Thorpe, DR, CerCo, Université Toulouse - Paul Sabatier

Jury :
Mireille Besson, DR, Laboratoire de Neurosciences Cognitives, AMU (Rapporteure)
Noël Nguyen, PR, Laboratoire Parole et Langage, AMU (Rapporteur)
Radouane El Yagoubi, MCF, CLLE-LTC, Université Toulouse - Jean Jaurès (Examinateur)
Hélène Lœvenbruck, CR, Lab. de Psychologie et NeuroCognition, Université Grenoble-Alpes (Examinatrice)
Maren Schmidt-Kassow, Assistant Prof., Institute of Medical Psychology, Frankfurt (Examinatrice)

Contents

List of Figures

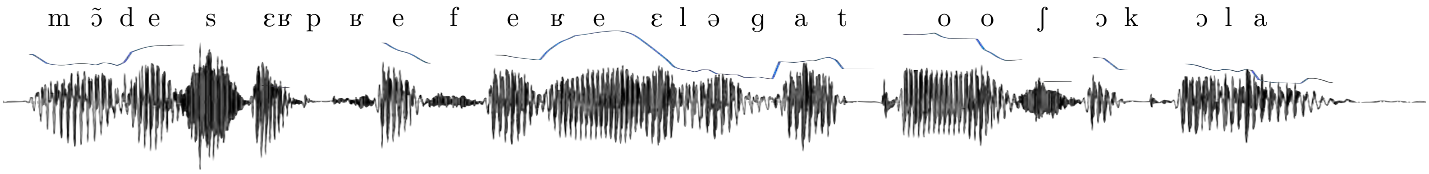
List of Tables

List of Acronyms

f_0	fundamental frequency
ABH	Attentional Bounce Hypothesis
AP	Accentual Phrase
DAT	Dynamic Attending Theory
EEG	Electroencephanlography
EMG	Electromyography
ERP	Event Related Potential
FA	Final Accent
IA	Initial Accent
ICA	Independent Component Analaysis
IP	Intonation Phrase
ISI	Inter-stimulus interval
ITI	Inter-trial interval
LD	Lexical Decision
MMN	MisMatch Negativity
MSS	Metrical Segmentation Strategy
PMN	Phonological Mapping Negativity
PWC	Possible Word Constraint
SOSH	Syllable Onset Segmentation Strategy

Part I

INTRODUCTION



Speech is one of the most informationally rich sounds we encounter on a daily basis. From a speech signal, the listener can extract extra-linguistic or para-linguistic information about for instance the speaker's identity, health or emotional state, and s/he can extract the segmental information (e.g. phonemes such as vowel and consonants) and supra-segmental information (i.e. bigger than the individual segments) necessary to recognize words. An acoustic representation of speech (the French spoken sentence "Mon dessert préféré est le gâteau au chocolat", *My favorite desert is chocolate cake*) is depicted above. The acoustic signal shows sound amplitude and fundamental frequency (f_0) to vary over time, and reflects some of the richness contained in the speech stream.

The information in speech is not presented in a robotic, staccato nor continuous manner, but organized in a 'musical arrangement' which structures and groups the speech flow, such that it can be perceived in coherent and bite-sized chunks. In this musical arrangement of speech—better known as **prosody**—intonation, accentuation and rhythm, self-organize to form a hierarchical prosodic framework, with, at its *heart*, **metrical stress**.

Metrically stressed syllables may be seen as the *pillars* guiding listeners through their analysis of an utterance. Metrical stress serves as the anchor point of the intonation contour, and it is the regulating force behind rhythm.

Moreover, modulations in the acoustic characteristics of metrical stress inform listeners on the depth of the prosodic hierarchy. Additionally, metrical stress underlies the abstract representation of the word—the smallest informational unit—such that it readily contributes to lexical, word-level processing. Indeed, metrical stress provides listeners with syllables that are perceptually stable; the listener can cling onto those syllables in difficult or noisy listening environments wherein the speech signal may be corrupted. Also, stressed syllables often *mark* the *boundaries* of a word, which, as we will discuss later in this work, are not typically evident from the acoustic signal. Finally, metrical stress interplays with **attention**, concurrently through its acoustic prominence at the **surface level** and through its **underlying** temporal predictability. That is, metrical stress harnesses attention from bottom-up through its attention-grasping acoustic salience, as well as top-down guides attention by providing a metrical framework, which the listener can use to bounce attention from one stressed syllable to the next. Given all these features, clearly, metrical stress forms the *backbone* of speech processing.

Amidst human languages, French, the language under investigation in the current dissertation, may be seen as eccentric, because the language, allegedly, has no stress (Rossi, 1980). The reasons for this peculiar, but, to date, accepted traditional view, will be discussed—and argued against—throughout this dissertation. Important at this point, is to note that the traditional view had two major consequences for speech processing in French. Firstly, the language was difficult to position in advanced theoretical models, wherein metrical stress plays a central role in the understanding of how listeners analyze speech. Secondly, alleged with no stress, French accentual phenomena, and particularly their role in word-level processing, have attracted little scientific interest. This means that, even if there were lexical or metrical stress in French, and even if it were to contribute to lexical processing, this function is not easily recognized, either because it is not investigated at all, or because the functions are attributed to domains other than the word.

Here, we question the view of French as a "language without stress". We align to two metrical models on French accentuation, which propose stress to be encoded in cognitive templates underlying the abstract representation of the word (Di Cristo, 1999, 2000; Astésano et al., 2007; Astésano & Bertrand, 2016; Astésano, 2016, 2017), and allow us to envision a role for French accentuation in word-level processing. In our interdisciplinary investigation of metrical stress processing in French, we take a functional, yet metrically rooted, approach. We use the method of Event-Related Potentials (ERP), which provides us with a highly sensitive and temporally precise measure such that we may determine whether there is metrical stress in French, and to what extent metrical stress aids the listener in speech comprehension.

Dissertation overview

▸ In Part II, we lay out our assumptions on a biologically plausible mechanism which accounts for the role of metrical stress in the analysis of speech. This part is constructed as follows:

In Chapter 1: *Defining metrical stress*, we will present stress as an abstract and cognitive entity which is specified in the mental lexicon and part of the lexical entry. We will motivate our functional approach and show that, while the interplay between accentuation, intonation and rhythm may, at times, make it difficult to recognize a proper lexical stress from the acoustic speech signal, stress can more readily be observed through its role in lexical processing. Additionally, we will present three theoretical frameworks that have a central place in the current work and help explain the functional role of stress in speech comprehension by relaying metrical structure to attentional processes:

- the Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988; Cutler, 1990),
- the Attentional Bounce Hypothesis (ABH; Pitt & Samuel, 1990) and,
- the Dynamic Attending Theory (DAT; Large & Jones, 1999).

Finally, we discuss in more detail why the existence of stress is questioned in French, and present the models which allow us to imagine a representation and functional role for metrical stress in French.

In order to appreciate the contribution of French accentuation to the process of speech comprehension, it will be necessary to better understand the challenges a speech system faces when confronted with an acoustic speech signal. Therefore, in Chapter 2: *Speech processing*, we will look closer as to what speech processing precisely entails. That is, we will discuss the processes involved in the perception of speech and explain how, in these processes, the listener uses metrical stress to confront two well-known challenges in the analysis of speech:

1. the challenge of persistent variability in the surface realization of linguistic information, and

2. the challenge in segmenting a continuous speech stream wherein the boundaries between linguistic units are not immediately obvious.

It will be concluded that listeners rely on the metrical and rhythmic regularities in speech to process utterances over multiple time-windows in parallel. Such a strategy to speech processing is presumed to rely on the interactions between prosody and attention, and brings us to the next chapter, Chapter 3: *Neural alignment to speech rhythm*.

In this Chapter, we describe a mechanism which provides a biologically plausible account on how the relationship between metrical stress and attention facilitates speech comprehension. In the same chapter, we will present the method used in this dissertation: the method of Event-Related Potentials (ERP). It is explained how, with this measure, we may observe whether French listeners have metrical expectations when processing speech, whether they expect stressed syllables at the level of the word, and whether speech processing is hindered when expectations are not met.

We will discuss four ERP-components relevant in the current work the MMN, PMN, N325 and N400. These components have in common that they all reflect a mismatch between a prediction based on long-term memory representations or established phonological representations and the violation in an experimental setting. The components allow for inferences on the time course (i.e. processing stage) and obstructed linguistic process (e.g. word recognition, access to meaning) that could result from presenting French listeners with stress or metrical information that does not agree with their anticipation, making the components—and the method of ERP in general—exceptionally well-suited for the current investigation.

▸ In Part III, we present our research contribution. This part is constructed as follows:

We will start, in Chapter 4: *Research Questions and Hypotheses*, by outlining the questions that motivated the current work. We explain the experimental approach that was set out to address the questions, and defend our hypotheses from which predictions were derived.

Chapter 5: *Experimental Strategy* presents the methods used in the thesis. More specifically, we will describe how we created our stimuli, modulating the surface realization of stressed syllables, while maintaining the natural sound of the speech signal. Additionally, we will motivate and explain our choices in the analysis of our data.

Finally, in Chapter 6: *Studies*, we present our studies wherein we investigated three main properties of metrical stress in French:

1. The phonological representation of French metrical stress
2. The function of metrical stress in word recognition
3. The interaction of French stress with lexico-semantic processing

Each investigation will be preceded by an introduction, such that the reader is well informed on the specific research question motivating each individual study. Also, the results of each investigation will be thoroughly discussed, as they often-times motivated the next investigation, and, also in this Chapter, the findings revealed by our data will be related to each other.

▸ Finally, in Part IV, we will present our general conclusion regarding metrical stress processing during speech comprehension French. We will argue for metrical stress, as well as the domain of the word, to be given a more prominent place in the descriptions of French prosody, and end our conclusion by discussing possible avenues for future studies. Indeed, if, as we will argue, metrical stress is at the foundation of a biologically plausible speech decoding system, and if, as we hope to have *shown* in the current dissertation, metrical stress facilitates processing *throughout* comprehension, acknowledging its existence is but a first step to myriads of enticing research prospectives.

Part II

THEORETICAL BACKGROUND

1 Defining metrical stress

Accentuation has been referred to with a range of different terms; accent, stress, prominence, salience, emphasis, force, but the list goes on. The choice of term may be merely stylistic; the author may have a personal preference for one term over the other, or use a range of them to avoid monotonous repetition. But the choice may also be deliberate and convey information that disambiguates the accent's phonological status, prosodic domain and linguistic function.[1]

A prime example of the confusion surrounding these different terms, is the distinction between 'stress' and 'accent' (Fox, 2000; van der Hulst, 2014). Some authors may use the term 'stress' to denote word level prominence and use 'accent' to exclusively refer to sentence level accentuation (e.g. Cutler & Foss, 1977). But for others, the two terms distinguish between the accent's representation in the brain and its phonetic realization in speech (Bolinger, 1958; Jassem & Gibbon, 1980; Abercrombie, 1976; Laver & John, 1994, in Fox 2000).[2] That is, for these authors, the terms 'accent' and 'stress' both refer to word level prominence, but they distinguish its phonological form as it underlies the cognitive representation of words, from the phonetic properties with which it surfaces in speech.

Indeed, while lexical stress belongs to the word and serves a number of functions in lexical processing, its surface realization is often co-determined by factors outside of the word (i.e. structural, pragmatic or rhythmic factors). These post-lexical factors may elevate the accent's pitch movement, increase its length or loudness, and they may also reduce its phonetic correlates, but,

---

[1] In the present dissertation, the terms shall be used synonymously to refer to abstract, word level accentuation unless otherwise specified.

[2] To make matters worse, while for Abercrombie (1976) and Laver & John (1994) 'accent' refers to the underlying representation and 'stress' to the surface level, for Bolinger (1958) and Jassem & Gibbon (1980) it is the other way around (Fox, 2000).

as will be argued in the current chapter, such modulations at the surface level do not change the accent's underlying lexical identity. The stress is still a property of the word and still contributes to lexical processing during speech comprehension.

1.1 Lexical stress

Lexical stress is held to be phonologically encoded and attached to the cognitive representation of the word (e.g. Eulitz & Lahiri, 2004; Cutler, 2010).[3] This means that a word's underlying stress template is specified in the mental lexicon and part of the lexical entry. An accent's place in the mental lexicon, and its role in lexical access, is most apparent when accentuation is **lexically distinctive**, i.e. when word meaning changes depending on which syllable carries the stress. In English, for instance, the noun 'suspect' differs form the verb '[to] suspect' in the position of its lexical stress. Identifying which of the two words is being referred to is, therefore, mandatorily guided by their underlying stress patterns. Thus, in these minimal pairs, the stress patterns must be encoded in the mental lexicon where they can act as gateway to the lexical representations. And so, when the accentual template is the decisive factor in word recognition, there is no doubt the accent belongs to the word.

However, an accent does not need to be lexically distinctive to have the word as domain. For example, in English, all content words (i.e. nouns, verbs, adjectives and adverbs) contain at least one lexical stress, and while most of them will not change meaning when stress is misplaced, such a misstep is likely to compromise word recognition (e.g. Cutler & Clifton, 1984, see also Magne et al. 2007 for a study wherein missplaced stess hindered word processing in French).

In fact, several stress properties make lexical accents excellent candidates to contribute to word processing. First, stressed syllables are perceptually stable. That is, stress enhances the syllable's segmental structure such that

---

[3] The current work assumes the abstractionist theory (e.g. Eulitz & Lahiri, 2004; Cutler, 2010) according to which phonological information is stored in templates which are registered in long-term memory and phonetically underspecified (i.e. fine-grained phonetic detail is omitted). But note that this is not the only view on the representation of linguistic information, with others holding encountered speech segment to be encoded in detailed, short-term memory traces called *exemplars* (e.g. Pierrehumbert, 2001). We will return to these two theories in section 2.2.1.

it is perceptually more robust against noise and more informative about the word's identity (e.g. Cutler & Foss, 1977; Bond & Garnes, 1980, see also Eisner & McQueen 2018 for a recent review). This means that in situations where the speech signal may be corrupted (e.g. in a noisy bar or on a windy beach), prosodic information often survives, which is less likely for the more vulnerable segmental information (e.g. Mattys et al., 2005).

Second, stress is acoustically salient. This means that the stressed syllables in an utterance, automatically draw in attention by means of their phonetic prominence (e.g. higher pitch, longer duration, increased intensity). Attention then boosts further processing of the word and facilitates, for example, lexical access, semantic retrieval and even post-lexical processes such as lexico-semantic integration (e.g. Cutler & Clifton, 1984; Pitt & Samuel, 1990; Rothermich et al., 2010, 2012; Rothermich & Kotz, 2013).

Third, lexical accents signal word boundaries (Hyman, 1977). In continuous speech there are no spaces reliably and unambiguously marking the boundaries between words, giving rise to the so-called **segmentation problem**. As we will discuss in more detail in chapter 2, listeners may rely on stressed syllables to detect the boundaries between individual words. Indeed, lexical accents are often located word-initial (e.g. Hyman, 1977; Cutler & Carter, 1987; Vroomen & de Gelder, 1995), and can, therefore, indicate the onsets of words and cue listeners on when to initiate lexical access.

Finally, in stress based languages, such as English or Dutch, the accents underlie the languages' metrical beat (e.g. Pike, 1945; Abercrombie, 1976, *cf.* Cutler & Norris 1988; Vroomen & de Gelder 1995). In these languages, stressed syllables form a metrical framework which guides the listener's attention so as to facilitate speech processing (e.g. Lehiste, 1973; Pitt & Samuel, 1990; Large & Jones, 1999, see also Quené & Port 2005).

Note that these properties to stress mean that stress interplays with attention during speech comprehension. Lexical accents both attract attention from bottom-up due to their acoustic salience, as well as a priori harness attention top-down through their predictability. We will return to all of the functions of word level stress in section 1.4, where we will additionally present three theoretical frameworks that help explain the facilitatory role of lexical stress in speech comprehension: the Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988; Cutler, 1990), the Attentional Bounce Hypothesis (ABH; Pitt & Samuel, 1990) and the Dynamic Attending Theory (DAT; Large & Jones, 1999).

But first, it will be necessary to better understand the difference between the phonological identity of stress and its surface realization. That is, while stress may be realized with a cocktail of phonetic parameters (e.g. f_0, duration, intensity), these are separate from its abstract and cognitive representation. We will see that several post-lexical influences (structural, pragmatic and rhythmic) may modulate the surface form of a lexical accent. Recall from the Introduction that, indeed, accentuation is only part of the prosodic organization which shapes the phonological form of an utterance, the others being intonation and rhythm. The interaction between accentuation, intonation and rhythm modulates the surface form of lexical stress which can create a problem in languages wherein stress is less straightforwardly a property of the word. As we will discuss in section 1.5, French, for instance, presents an (allegedly) syllable-timed language wherein stress is not lexically distinctive, and often overlaps with intonation. This has led some scholars to question the phonological status of the French accent. However, it will be argued here that these outside influences which co-determine the surface manifestation of accentuation do not modulate its lexical representation nor its lexical role in speech comprehension.

1.2 Higher level prominence

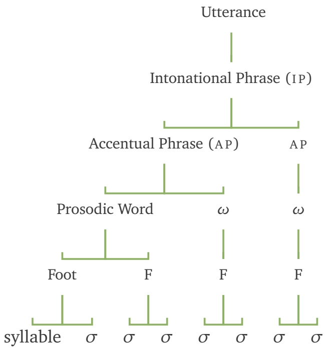


Figure 1.1: Representation of the prosodic hierarchical structure according to the **Strict Layer Hypothesis**, wherein each constituent in the hierarchy, exclusively dominates the level below it (Beckman & Pierrehumbert, 1986).[4]

Prosody is best described as the coordinating principle behind the *hierarchical* framework for speech (Beckman, 1996). That is, the (structural) relationship between accentuation, intonation and rhythm allows for the otherwise continuous speech stream to be broken down into bite-size chunks that can then be reassembled into meaningful and increasingly larger constituents. More concretely, prosodic organization enables speech utterances to be decomposed into hierarchically nested domains that group syllables into words, words into phrases, and phrases into the utterance (see figure 1.1 for a schematic representation).

Precisely how many constituents are involved in this prosodic structure and how they should be defined remains a widely debated topic in prosodic phonology (see table 1.1 for an overview) and is beyond the scope of the

---

[4] While the Strict Layer Hypothesis is generally upheld, there have been accounts of recursion in prosodic structure as well as observations of prosodic domains dominating constituents lower down the hierarchy (e.g. Shattuck-Hufnagel & Turk, 1996). For instance, Ladd (1986) proposes that an Intonational Phrase can directly dominate another IP, so that Intonational Phrases can be nested recursively, and, recursive prosodic words or prosodic phrases have also been suggested (e.g. Selkirk, 1996; Gussenhoven, 2004).

present work (but see e.g. Shattuck-Hufnagel & Turk, 1996, for an extensive review). It is, therefore, not the intent to provide a full description of the different views on prosodic hierarchy and how they relate to each other, but rather to broadly introduce the different layers of constituency with the purpose of showing how these constituents impose their own sets of phonological rules to the speech utterance.

Roughly, the largest domain in the hierarchy, the **Utterance** (U), corresponds to a full sentence. It is often demarcated by non-hesitation pauses and dominates over at least one Intonational Phrase (Hayes, 1989, in Shattuck-Hufnagel & Turk 1996). The **Intonational Phrase** (IP) usually corresponds to a clause (e.g. Féry, 2016) and is the domain of the intonation contour. A syllabic lengthening will cue its right boundary, optionally assisted by a movement in f_0 (Shattuck-Hufnagel & Turk, 1996). The Intonational Phrase typically holds one or more **Phonological** or **Accentual Phrases** (AP) which correspond roughly to syntactic phrases (e.g. noun phrase, verb phrase, adjective phrase; Frota, 2012) and are cued by pre-boundary lengthening of the last syllable and initial strengthening of the first syllable of the phrase.

As mentioned above, these layers interact with the surface realization of a lexical stress, resulting in accents of different degrees of relative salience. These accents then belong to the level above (and outside) of the lexical do-

Table 1.1: The hierarchy of prosodic constituents has been approached from three different perspectives: A **syntax-based approach** (left) in which prosodic constituents account for the combined morpho-syntactic rules and phonological constraints (e.g. Nespor & Vogel, 1983; Hayes, 1989; Selkirk, 1986). An **intonation-based approach** (middle) that attempts to describe the phonological rules underlying the intonation of languages(e.g. Beckman & Pierrehumbert, 1986; Jun, 2007). And a **prominence-based approach** (right), wherein levels of constituency correspond to levels of prominence which can apply to the word (in lexical stress), the phrase or the sentence (e.g. Beckman & Edwards, 1990, 1994). (Table adapted from Shattuck-Hufnagel & Turk, 1996)

Syntax-based		Intonation-based	Prominence-based
Utterance	(Utterance)		
Intonational Phrase	Intonational Phrase	Full Intonational Phrase (IP)	Nuclear Accent
Phonological Phrase	Major Phrase	Intermed. Intonational Phrase (ip)	(Pre-Nuclear) Accent
	Minor Phrase	Accentual Phrase (AP)	
Clitic Group			
Prosodic Word	Prosodic Word		(Lexical) Stress
Foot	Foot		
Syllable	Syllable		
	Mora		

main, and do not directly contribute to word processing,[5] but instead serve post-lexical functions in the structuration of the speech stream. For instance, they may mark higher level constituent boundaries, or serve in pragmatic functions such as contrast or emphasis (e.g. "I called Mary not Peter", or "I was super relieved"). Indeed, we all know, intuitively, that the form of an utterance never resembles a sequence of isolated words, but is, instead, shaped by an overlaying prosodic structure, which lifts certain (lexically) accented syllables to different degrees of 'higher level' prominence.

Such post-lexical accentuation thus serves functions in expressing, for instance, the information structure (e.g. focus or topic) or information status (e.g. given compared to new information) of a message (Gordon, 2014, see also Fox 2000 and van der Hulst 2014). For example, in the English sentence "John was the suspect", it is possible for both John or suspect to carry a prominent accent, depending on which of the two is in focus (e.g. "Who was the suspect?" → "John was the suspect", "Was John the victim?" → "John was the suspect"). On the other hand, it is also possible for was to carry an accent if the speaker had been asked "Is John the suspect?", and he *was* but he no longer *is*.

Thus, higher level accents convey pragmatic information and are post-lexical in nature, i.e. they are not a property of the word. The accents, however, do typically land on the lexical stresses, such that the phonetic parameters of the lexical stress merge with the acoustic realization of the post-lexical accent (Gordon, 2014; van der Hulst, 2014). When a higher level accent temporally aligns with a lexical stress, authors will often speak of a promotion for the syllable carrying word level stress. However, the term 'promotion' may be misleading in this context. While it is true that the syllable received additional prominence from a constituent that is higher up in the prosodic hierarchy, the original prominence is not "moved forward" or "moved upward", or, indeed, moved at all. It is still there, but the syllable carrying the lexical stress, received additional post-lexical prominence. That is, while the stressed syllable now functions in post-lexical structuration, it *still also* flags the lexical word. Post-lexical prominence is thus more integrated and culminative, than it is promotional.

Nevertheless, it can be problematic to distill lexical marking from processes outside of the word; word stress and higher level accentuation are often inextricably intertwined (see Gordon, 2014, for a discussion). Recall that the status of French accentuation provides a good example of this diffi-

---

[5] Although we will discover later, that they may do so *indirectly*. Because phrase boundaries coincide with the boundary of a word (recall the Strict Layer Hypothesis in figure 1.1), they can indirectly contribute to processes such as word segmentation and lexical access.

culty. Traditionally, accentuation holds a (per definition) post-lexical status in this language, i.e. the language would have no word level stress and accentuation would apply only to the phrase (see section 1.5 for a discussion). One of the reasons for this view of French stress, is the temporal alignment of the French primary accent with the intonation contour. As van der Hulst remarks in a chapter on the study of lexical stress:

> "[Post-lexical] effects can [be] dramatic when the claim is made that an alleged word stress is not present at all and that the impression of word stress is caused by the fact that final syllables of words (typically when occurring in phrasal final position) carry an intonation pitch accent, not because they have stress, but simply because they are phrase final. This is one way of analyzing the 'final stress' in French."
>
> *van der Hulst (2014, p. 24)*

Indeed. In the current work, we will attempt to dissociate the French accent from post-lexical effects and determine whether there is a possibility for lexical stress in French. Gordon (2014) discusses several methods to address the challenge in teasing apart word-level stress from post-lexical influences. For example, a common strategy is to record or present words either in isolation (i.e. without phrasal context), or in a carrier phrase wherein the target words are positioned away from phrase boundaries and out of focal position. Unfortunately, however, these methods do have their shortcomings.

For instance, isolating a word may remove it from phrasal context, but the word remains an utterance. Because the word is also an utterance, its lexical stress is necessarily 'contaminated' by an utterance level (i.e. post-lexical) intonation contour marking both left and right boundaries. In order to avoid this type of post-lexical boundary marking, researchers will attempt to embed words in a sentence. However, embedding a target word in a carrier phrase while avoiding post-lexical marking, remains a difficult task. Considering the depth of prosodic constituency, the lower level phrasal constituents, just above the word but outside the lexical domain (see figure 1.1), may contain as few as one word. This means that even when words are embedded into a sentence, the lexical stress will still often be at or near the edge of a phrase where it may also cue higher constituent boundaries. Moreover, in experimental settings, the target words are often the only changing variable in an otherwise static sentence. Therefore, the target word is likely to attract implicit (and post-lexical) focus. Consider for instance a carrier sentence like:

> "John has [target-word] tattooed on his arm, not on his chest".

Here, the target word is:

1. embedded into a sentence (i.e. not presented in isolation)
2. away from constituent boundaries (i.e. in the middle of AP "John has [. . .] tattooed")
3. and out of (explicit) focus (which is on the contrastively stressed "arm" versus "chest")

However, because, in the experimental setting, the otherwise static sentence varies only on the target word, the word is likely to draw attention. The word is implicitly focused (Gordon, 2014).

Finally, there are attempts to tease apart lexical stress from phrasal prominence by relying on their respective unique surface level realizations. That is, if for instance word level stress were always realized by, say, increased intensity, while post-lexical prominence were consistently flagged by a movement in pitch, then they can straightforwardly be distinguished by means of acoustic analysis.

Unfortunately, however, any categorization of prosodic phenomena based only on their (surface-level) acoustic realization appears not to be possible. That is, while, indeed, word stress is often realized with a cocktail of phonetic parameters (including duration, f_0 and intensity), and post-lexical prominence is primarily cued by the intonation parameter (pitch; Gordon, 2014; van der Hulst, 2014), word stress may be primarily cued by pitch fluctuation, while, conversely, phrasal prominence may also be marked by duration or intensity. In fact, as mentioned before, phrase boundaries are frequently signaled with syllabic strengthening at their onset, and additional lengthening at their offset, and therefore not marked only by pitch. Furthermore, factors such as the duration of syllables can be co-determined by, for example, word size or speaking rate and phonemes vary in terms of intensity, duration or f_0 due to so-called micro-prosodic variation (e.g. Di Cristo, 1976). Finally, the different phonetic correlates are difficult to disentangle, since, for example, f_0 and intensity often vary synchronously (e.g. a stronger air flow increases both intensity, which creates more loudness, and fundamental frequency via a more rapid vibration of the vocal folds, which, in turn, results in higher pitch).

It appears, then, that there is no ad-hoc method for reliably distinguishing stress from post-lexical events. This can make it difficult to determine whether an accent really belongs to the word domain (*cf.* van der Hulst, 2014). Recall, however, that under certain circumstances, e.g. when stress is lexically distinctive, stress unambiguously belongs to the word. This is because, in those circumstances, the accent has a functional role in lexical processing. In this dissertation, it will be argued, that it is through the contribution of lexical stress in word processing, that we may observe its identity. Crucially, such a perspective detaches the accent from its phonetic realization, and instead defines it based on its phonological involvement in lexical processing and the structuration of the speech stream. As such, the accent is an abstract entity, independent from phonetic parameters, that serves to attract attention to the salient moments in speech so as to facilitate speech comprehension.

1.3 Rhythm and Meter

In the previous section, we saw that in any given utterance we may perceive prominences of different degrees. This is because, the different layers in the prosodic hierarchy each influence the phonological form of the utterance, resulting in prominences of different strength that reflect the hierarchical relationships between the prosodic constituents. For instance, there may be boundary tones, flagging the edges of the higher-level constituents, or there may be pragmatic accents which express the utterance's information structure. Each of these post-lexical constraints modulates the surface level realization of lexical stress. But there is an additional coordinator in the prosodic organization of speech utterances which we have yet to cover: rhythm.

The phonetically independent identity of accentuation is best evidenced by its role in the metrical organization of English rhythm. Above, it was briefly mentioned that English rhythm is stress based. Here, we will look closer as to what that means. Rhythm entails beat, meter, and timing. It presents the temporal relationship of beats or events, which is organized by an abstract, underlying metrical pattern that does not have to be regular *per se*. In other words, rhythm presents the organization of metrically strong and metrically weak beats over time. The actual phonetic parameters associated with the metrical beats can be variable. Indeed, metrical beats are often described as 'mental beats', i.e. they come from within, and do not require any phonetic manifestation. Meter, therefore, represents an abstract notion that organizes rhythm. Rhythm, in all this, is holistic, i.e. it is *not* the *sum* of a sequence of events ($strong + weak + strong + weak \neq rhythm$), but it has its own shape or form. As such, rhythm may be understood to represent an auditory Gestalt. That is, it is an assemblage of strong events, that contrasts to a background of weak events. This assemblage is perceptually regulated by a metric beat, and in English, that beat is governed by stress.

The notion of a specific level in prosodic hierarchy controlling the perception of linguistic meter stems from the isochrony hypothesis according to which each language has a periodically recurring 'unit' that determines its metrical organization (Pike, 1945; Abercombie, 1967, in Fletcher 2010). Depending on the language, this unit may be the mora (a phonological unit just under the syllable, dominant in languages such as Japanese), the syllable (in French, or for instance Spanish) or stress (the English metrical unit).

Crucially, this three-way division of languages relies on isochronic, surface-level realization of the prosodic units underlying the languages' beat. This means that in syllable-timed languages, all syllables should be of approximately equal duration, while in stress-timed languages the distance between stressed syllable should remain constant. The hypothesis therefore predicates same length syllables, irrespective of syllable complexity, in syllable-timed languages (hence the name 'machine gun' for the sound of these languages), and syllable compression or silent beats in stress-timed languages (i.e. when there are many syllables between local accents, or too few, respectively, leading to a 'morse code' sound for these languages).

The idea of strict, surface-level isochrony between language specific structures is, however, found not to hold (e.g. Roach, 1982; Dauer, 1983; Fant et al., 1991; Ramus et al., 2000; Arvaniti, 2012, see also Fletcher 2010; Turk & Shattuck-Hufnagel 2013 for recent reviews, and Cumming 2010 for a detailed, historic overview of work on linguistic rhythm). Cross-linguistic phonetic studies consistently failed to deliver evidence of differences in the temporal regularity of inter-stress intervals, showing intervals around 550 ms for both stress based and syllable based languages (Fant et al., 1991). Similarly, Delattre (1966) found equal variability in syllable duration in French and Spanish (both syllable-timed languages) as in English and German (both stress based) (see also Wenk & Wioland, 1982; Roach, 1982; Dauer, 1983; Ding et al., 2016b).[6] So, it has instead become clear that languages are more often somewhere along a continuum between the two rhythm categories (e.g. Dauer, 1983; Auer, 1993; Ramus et al., 2000; Grabe & Low, 2002; Arvaniti, 2012).

The lack of empirical support can likely be attributed to the assumption of rigid periodicity in speech, the focus on acoustic (surface-level) rather than perceptual analysis and the representation of linguistic rhythm as a one-dimensional structure. That is, there is little doubt amongst scholars on the reality of a perceptual impression of a particular kind of rhythm between languages (e.g. Turk & Shattuck-Hufnagel, 2013). It seems rhythm is not the result of an isochronic, surface-level event, but of a phonological mechanism that does not depend on phonetic correlates to be perceptually real for the listener (e.g. Lehiste, 1973). Again, rhythm is an auditory gestalt, regulated by an underlying, and abstract, metrical structure.

Rhythmically induced perceptual gestalts can be perceived all around us. Think, for instance, of the ticking of a clock, which is often perceptually grouped into groups of two beats (—tik-tak—tik-tak—). Also more irregular sounds (e.g. the dripping of leaky tab; Fletcher, 2010) are likely to be perceptually grouped if they are neither too far apart nor too close together (i.e. distances between 0.1 and 3 seconds; Fletcher, 2010). These perceptual groups will typically be of equal duration, and may present a hierarchy of relative prominence (e.g. the first pulse in a group may be perceived as more salient if it is louder or higher in pitch, while, similarly, the third pulse may be perceived as ending the group, if it is longer in duration; Fletcher, 2010).[7]

The tendency to impose rhythmic structure on irregular sequences of events, translates to the perception of speech. For example, when asked to tap along English utterances, listeners tended to tap more regularly than the actual inter-stress intervals (Donovan & Darwin, 1979, in Fletcher 2010), partially evidencing the perceptual bias towards isochrony. This tendency towards isochrony has been related to the notion of the **perceptual center** or P-center (e.g. Morton et al., 1976, in Cumming 2010; Fletcher 2010).

---

[6] Indeed, universal boundary phenomena such as pre-boundary or word-initial lengthening can drastically increase the duration, also in syllable-timed languages (e.g. Lehiste, 1973; Fletcher, 2010). Notably, the French primary accent (the final accent, FA, see section 1.5) is primarily marked by increased duration of syllabic rime, relative to surrounding syllables (Di Cristo, 1999; Vaissière, 1991; Astésano, 2001).

[7] Note that this is in accordance with the Iambic-Trochaic-Law, according to which Feet group syllables either **iambically** (weak-strong), when stress patterns are defined mostly by an alternation in duration, or **trochaically** (strong-weak), often defined by a difference in intensity.

P-centers are generally described as 'psychological moments of occurrence'. They are perceptually salient 'isochronic-like' events near the vowel onset of stressed syllables, separated by around 500 ms,[8] which appears to be a preferred rate for humans (Allen, 1975, in Astésano 1999; Fletcher 2010).

This apparent perceptual bias towards isochrony has been modeled in a recent computational effort on musical pulse perception (Large, 2008; Large & Snyder, 2009). In the computational model, neural resonance offers an explanation on how non-periodic auditory input is represented and processed in the brain to create the percept of a periodic pulse. According to the model, pulse perception emerges through non-linear coupling of two (neural) oscillators. One oscillator will closely track the physical properties of the auditory input, while the other integrates the input from the sensory system over larger time-windows (Large & Snyder, 2009). The non-linear interactions between the systems give rise to oscillatory activity not only at the frequencies present in the input, but also at more complex combinations, including the pulse frequency and its harmonics. Thus, this framework predicts that the interaction between auditory input and intrinsic neural activity leads to neural entrainment to the pulse frequency, which in turn creates the perception of periodicity.[9]

Note, that the model underlines that pulse and meter are percepts and not part of the auditory signal itself (Large, 2008).

> “ [Pulse and meter] are responses to patterns of timing [...] in the acoustic rhythm. Although responsive to stimulus properties, pulse and meter are not themselves stimulus properties. These terms refer to endogenous dynamic temporal referents that shape experiences of musical rhythms. The rhythms of music, which are temporally complex and richly articulated, are heard in relation to a relatively stable percept of pulse and meter. ”
>
> *Large (2008, p. 190)*

---

[8] Note that this is similar to the inter-stress interval durations found between both stress based English and syllable-timed French, which were shown to have inter-stress durations around 550 ms (Fant et al., 1991).

[9] We will return to the concept of neural entrainment, specifically regarding its involvement in the perception and analysis of rhythmic stimuli, in chapter 3.

The fact that rhythm perception has no clear phonetic correlates, indicates that it is based on a psychological mechanism.

But, how does all this relate to the influence of rhythm on the surface realization of lexical stress?

In speech, the relationship between rhythm and stress is bidirectional. That is, rhythmic structure is created by stress patterns (i.e. the alternation of strong and weak syllables), but, conversely, rhythm plays a role in the phonological processes which determine the distribution of stressed syllables.[10] Consider for example the distribution of surface level stressed syllables in English utterances. It was mentioned above that all English content words are marked by *at least* one stressed syllable that aids in lexical processing, but from a metrical point of view, *the number* of (surface) accented syllables depends on rhythm. And, indeed, when speech is especially slow or contains long sequences of unstressed syllables (so-called **stress lapses**), more syllables will be accented (see table 1.2 for an example). These 'new' accented syllables are considered secondary and optional in the marking of the word, contributing, instead, to rhythmic structuration.

Similar to stress lapses, in high-paced speech or when stressed syllables are too close together (resulting in what is known as a **stress clash**), prominences are reorganized such that a primary stress may be 'withdrawn' in favor of a more pronounced secondary stress located further away (see also table 1.2). This reorganization of local prominences is a well studied phenomenon better known as the **Rhythm Rule**, the **Iambic Reversal** or the (less appropriate) **Clash Resolution Shift**[11] (e.g. Liberman & Prince, 1977; Fox, 2000).

Stress lapse:

In slow speech, the word elevator operator may be pronounced as:

⇒ elevator-operator

while, in fast speech, there may be only one accented syllable:

⇒ elevator-operator's car

Stress clash:

Consider the word thirteen, wherein stress is "shifted" to the first syllable when followed by a syllable with primary stress:

⇒ thirteen men

Table 1.2: That stress is dependent on rhythm, is evidenced by the phonological rules which help evade cross-linguistically disfavored stress lapses (long sequences of unaccented syllables) and stress crashes (when two accents are too close together).

The suppression of stressed syllables that would otherwise be out of beat, and the emergence of local prominences as dictated by rhythm, again demonstrate the abstract nature of accentuation. Stress is encoded in cognitive templates underlying the representation of lexical words. In con-

---

[10] This relationship between rhythm and stress and their combined role in the organization of the speech stream has a foundation in the theory of Metrical Phonology (Liberman & Prince, 1977; Hayes, 1989). In metrical phonology, the **Principle of Rhythmic Alternation** is emphasized and stress is presented as a relative concept, belonging to the union of accented and unaccented syllables (the foot, see figure 1.1), but representing both word level and utterance level patterns in a hierarchy of prosodic constituents.

[11] Note that the term Clash Resolution *Shift* is somewhat misleading as it suggests that the primary stress is in its whole taken from its designated syllable and placed on another. This is not the case, the rearrangement of accents refers to the surface distribution of local prominences, with the underlying, abstract stress templates remaining intact.

nected speech, stress may—yet need not—be phonetically realized. Because stress underlies a language's rhythm, the location of stressed syllables can be anticipated. So the functional value of rhythm and meter—and metrical stress—is straightforward: it allows listeners to a priori direct their attention to the next occurrence of an accented syllable, again pointing to a valuable role of accentuation in comprehension processes (e.g. Fraisse, 1982; Large & Jones, 1999).

1.4 Functions of stress in models of speech processing

In the previous section we saw that prosody is best described as the coordinator of speech (e.g. Beckman, 1996). Accentuation, intonation and rhythm jointly structure speech into bite-sized chucks that can more readily be analyzed by the listener. Lexical stress plays a crucial part in speech processing. It is attached to the cognitive representation of words and, as such, serves as the gateway to the mental lexicon (e.g. Eulitz & Lahiri, 2004; Cutler, 2010). Stressed syllables may be realized with combinations of pitch, length or loudness, but these phonetic features do not define stress. Instead, stress is defined by its place in the mental lexicon (Abercrombie, 1976, in van der Hulst 2014).

That is, the surface realization of stressed syllables varies and is co-determined by structural, pragmatic and rhythmic constraints. Higher level prominence from the constituents above the word is often placed on the stressed syllable and modulates its phonetic manifestation. Additionally, accentuation interacts with rhythm such that local prominences may be suppressed when they are out of beat, or emerge when the sequence of unstressed syllables is too long.

Interestingly, this Rhythm Rule is constrained within a prosodic domain (e.g. Phonological Phrase, see table 1.1) and cannot be applied across phrase boundaries (e.g. Post, 2000; Frota, 2012), which underlines the interdependence and collaboration of the prosodic hierarchy (or intonation), rhythmic structure and underlying stress representations in the surface form of the utterance (e.g. Fox, 2000; van der Hulst, 2014). That is, higher level (pitch) prominence is anchored on the utterance's rhythmic structure (i.e. metrically strong syllables), which is itself determined by stress. This indicates that the speaker prepares the form of his message, i.e. on the surface, phrase and word level prominences are integrated according to the rhythmic plan of the utterance (e.g. Aylett & Turk, 2004; Turk & Shattuck-Hufnagel, 2013, 2014). In this plan, stressed syllables then both serve to attract attention by means of their phonetic salience and a priori harness attention by means of their metrical predictability.

In the current section, we will present three frameworks that are based on the interactions between stress and attention and help explain the functional role of accentuation in speech comprehension. Indeed, several stress properties make lexical accents excellent candidates to contribute to word processing. Stressed syllables are:

1. Perceptually stable.
2. Acoustically salient.
3. Cues to word boundaries.
4. And predictable.

Because of the combination of these properties, prosodic accentuation is tightly linked to attention; at the surface level, the physical salience of stressed syllables allows them to pop-out from their background and attract attention, while, simultaneously, the underlying accentual meter entrains attention such that it hits all the marks. As such, the partnership between accentuation and attention facilitates processing throughout speech comprehension and is of central importance in the theoretical frameworks discussed below: The Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988; Cutler, 1990), according to which listeners rely on the metrical beat in their language to mark the boundaries of the words and cue lexical access. The Attentional Bounce Hypothesis (Pitt & Samuel, 1990, ABH;), which proposes that metrical structure serves to attract attention towards the lexically stressed syllables. And, the Dynamic Attending Theory (DAT; Large & Jones, 1999), which states that attention dynamically entrains to the rhythmic (and, hence, predictable) structure in sound.

1.4.1 Metrical Segmentation Strategy

Speech is a continuous signal with no cues consistently marking the boundaries between words. Still, listeners seem to have no problem recognizing individual words and segmenting the speech stream. Much research has been devoted to understanding how such segmentation might proceed.[12]

According to the Metrical Segmentation Strategy (MSS), listeners rely on their languages' metrical structure to infer word boundaries (Cutler & Norris, 1988; Cutler, 1990). In languages such as English or Dutch, rhythmic structure is stress based, and lexically stressed syllables are often word initial. Indeed, in a study analyzing the distribution of strong (i.e. stressed) and weak syllables in English conversational speech, Cutler and Carter found that 90% of lexical (open-class) words were either mono-syllabic[13], or polysyllabic, beginning with a stress (Cutler & Carter, 1987, see also Vroomen & de Gelder 1995 for comparable results in Dutch). Because listeners are sensitive to such a statistical prevalence, they will search their mental lexicon as soon as they encounter a strong syllable, while, conversely, refraining said search on weak syllables (Cutler & Norris, 1988).

Evidence for MSS is provided by studies on juncture misperception ('slips of the ear'), wherein English and Dutch listeners more frequently erroneously inserted a word boundary when encountering a strong syllable (for instance, "analogy" → "an allergy") or deleted a word boundary before a weak syllable (for instance, "my gorge is" → "my gorgeous"), than the other way around (e.g. Cutler & Butterfield, 1992; Vroomen et al., 1996, see also e.g. Banel & Bacri 1994 for similar findings on phrasal segmentation in French). Further evidence is to be found in word-spotting studies, wherein the detection of words embedded in nonsense-words (i.e. pseudowords) was slowed down when the word straddled the boundary of a strong syllable (e.g. Cutler & Norris, 1988; Norris et al., 1995; Vroomen & de Gelder, 1995; Vroomen et al., 1996).

Note, however, that MSS proposes a universal strategy, such that the metrical units used to infer word boundaries may differ between languages. French listeners, according to MSS, are thus expected to segment on the syllable, the (alleged) French metrical unit (e.g. Pike, 1945; Abercombie, 1967; Mehler et al., 1981; Cutler et al., 1986). In chapter 2, we will see that speech

---

[12] We will return to this topic in chapter 2 wherein the problem of speech segmentation is approached from several perspectives.

[13] Because all content words in English carry at least one stress, the syllable in a monosyllabic word is, by definition, stressed.

segmentation involves a more complex strategy wherein listeners are likely to use any and all information available to them. That is, listeners presumably segment speech along the entire prosodic hierarchy (i.e. syllable, foot and phrase) using different cues on an as-needed basis. Nevertheless, MSS proposes an elegant and cross-linguistic solution as to how listeners benefit from stressed syllable in confronting the problem of speech segmentation and is able to account for much of the evidence wherein stress patterns were found to guide speech processing.

1.4.2 Attentional Bounce Hypothesis

A strategy related to MSS is proposed in the Attentional Bounce Hypothesis (ABH; Pitt & Samuel, 1990). In ABH, the listener also relies on the metrical structure in speech for word processing, but now does so anticipatory. That is, ABH posits that the temporal regularity of stressed syllables provides a structure that allows anticipation of future occurrences of perceptually clear segmental speech[14] (Pitt & Samuel, 1990). The listener can then selectively tune attention to "bounce" from one stressed syllable to the next. In ABH, metrically strong syllables then serve both to attract attention by means of their stability and acoustic salience as well as a priori distribute attentional resources through their predictability. The hypothesis thus proposes that predictably located beats in a metrical structure of an utterance guide the attention of the listeners to the stressed syllables.

Evidence for ABH is provided by studies on phoneme monitoring, which is facilitated on syllables that are expected to be stressed based on a previous rhythmically regular context (i.e. sentence context or word list; e.g. Pitt & Samuel, 1990, see also Cutler 1976; Quené & Port 2005; Breen et al. 2014). Listeners were instructed to identify target phonemes in minimal stress pairs. For instance, they were asked to detect the phonemes [p] or [m] in the minimal pair 'permit'.[15] Importantly, the acoustic signal of the target word was manipulated such that both syllables were equally salient. Behavioral results (error rates and reaction time) showed a general trend suggesting that participants performed better when phonemes were located

---

[14] Recall that stressed syllable are perceptually stable and survive under noise, making them more reliable cues than, for instance, segmental information.

[15] Notice that 'permit' refers to a noun when stressed on the first syllable ('permit'), and to a verb when stressed on the second ('permit').

in syllables that were expected to be stressed. This suggests hat attention had been preferentially assigned to the (anticipated) stressed syllables and facilitated speech processing.

It is important that syllable stress was normalized, since, while it had already been shown that phonemes are detected faster or more accurately when they occur in stressed syllables (e.g. Cutler & Foss, 1977; Bond & Garnes, 1980), it had been unclear whether the perceptual advantage was due to the bottom-up acoustic salience of local accents, or due to top-down temporal anticipation. In presenting the minimal pairs with equal stress (i.e. no stress) on both syllables, the results in Pitt & Samuel (1990) hint towards the latter interpretation, wherein the metrical structure in speech guides attentional resources so as to hit all the accentual landmarks. Since then, various studies have backed-up this interpretation, and demonstrated that metrical regularity induces temporal expectations for the location of local prominences, which are subsequently preferentially processed (e.g. Schmidt-Kassow & Kotz, 2008; Schmidt-Kassow et al., 2009; Rothermich et al., 2010, 2012; Cason & Schön, 2012; Roncaglia-Denissen et al., 2013; Falk & Dalla Bella, 2016; Harding, 2016).

1.4.3 Dynamic Attending Theory

Closely related to ABH, the Dynamic Attending Theory (DAT; Large & Jones, 1999) assumes that the degree of temporal predictability of auditory information dynamically entrains attention on multiple temporal planes. DAT proposes a mechanism comparable to the pulse resonance model by the same authors, which was presented in section 1.3 (Large, 2008; Large & Snyder, 2009). In DAT, the perception of meter is held to result from the interaction between a sound signal and intrinsic neural dynamics. The model further posits that, if there are regularities in the signal, these can be used to guide attention to particular points in time. In complex rhythms, with a hierarchy of temporal regularities, multiple oscillators synchronize and nest such that the different frequencies lead to a united perception of metrical layers (Large & Snyder, 2009). That is, attention both tracks the hierarchy of predictable events as well as groups them into nested domains.

The theory finds support in a series of studies, in which a rhythmic auditory cue was shown to facilitate perceptual processing of a subsequent target, with highest accuracy for targets that were phase-aligned with the rhythmic cue (Jones, 1976; Large & Jones, 1999; Jones et al., 2002). Additional evidence is provided by studies showing the degree of rhythmic regularity to modulate the strength of temporal anticipation (e.g. Schroeder & Lakatos, 2009; Jones, 2010, see also Henry & Herrmann 2014 for a literature overview). Moreover, the theory appears biologically plausible, with recent developments in the field of neuroscience demonstrating ongoing neural oscillations to be entrained by temporally regular stimuli and align neural excitability to external rhythmic structures (e.g. Schroeder & Lakatos, 2009; Rohenkohl & Nobre, 2011; Giraud & Poeppel, 2012; Henry & Herrmann, 2012; Gross et al., 2013; Henry et al., 2014; Ten Oever et al., 2017), which we will return to in chapter 3.

Dynamic Attending of Speech

DAT does not directly address speech processing, but there are several parallels. In Metrical Phonology, for instance, the speech utterance is organized by multiple levels of prosodic constituency. Each domain has its own influence on the phonological form of the utterance. This means that in any language, rhythmic structure results from the combined efforts of syllable, stress, and phrase timing (e.g. Cummins & Port, 1998; Arvaniti, 2009; Tilsen & Johnson, 2008; Nolan & Asu, 2009, *cf.* Martin 1972). Indeed, this is similar to the suggestion that the metrical structure of languages is unlikely to be categorized into either stress or syllable-timed, and is presumably better placed somewhere along a continuum (e.g. Dauer, 1983; Cummins & Port, 1998; Arvaniti, 2009, 2012). That is, in any language, rhythm is based along the entire prosodic hierarchy, and thus based both on the syllable as well as the stress (*cf.* Astésano, 2001).

Cummins & Port (1998), for instance, proposed an adaptive oscillator model, which is similar to DAT but integrates oscillators that more specifically track the different layers in the prosodic hierarchy. That is, Cummins and Port propose speech rhythm to involve a set of hierarchically ordered oscillators that entrain to the different metrical levels so as to direct and attract attention (see also Port, 2003).[16] Evidence for the adaptive oscillator model, is provided by studies using the Rhythmic Cycling paradigm that is designed

---

[16] The model specifically addresses timing in speech *production*. As such, another important function of the hierarchically ordered oscillators is to bias the motor system such that prominent motor events are coupled to the attentional pulse (e.g. Port, 2003).

to reveal the rhythmic organization in speech production, in particular with respect to inter-stress intervals and their interdependence to the cycle of the phrase-oscillator (e.g. Cummins & Port, 1998; Port, 2003, see also Tajima & Port 2003 for a study on rhythm in Japanese speech production).

In the paradigm, English speakers are asked to repeat a phrase (e.g. "big for a duck") while aligning to a two-tone metronome. That is, the first and last stressed syllables of the phrase (i.e. "big" and "duck", respectively) should be aligned to the two tones. While successful initially, the metronome that guides the speakers fades out after a few repetitions, at which point the timing of the produced syllables starts to stray from the original stress-beat, but, interestingly, remains at simple harmonic fractions of the phrase period. That is, when speakers lost the metronome that initially guided their production, they did not place the stressed syllable at random locations but were biased towards the phrase harmonic.

This consistent placement of stress within the phrase suggests that speech is temporally organized according to hierarchical principles wherein the lower levels are nested under the higher levels with integer relations. That is, the window of the phrase constrains the possible locations of the stress beats, presumably because the longer time-window (i.e. lower level harmonic frequency) is more stable (Cummins & Port, 1998; Port, 2003, see also Haken et al. 1985). A fixed number of small prosodic units (e.g. 2 or 3 syllables, or 2 or 3 feet) then nests under one cycle of the oscillator representing the higher level unit (e.g. the foot or AP, respectively). Notice how the model mirrors the nested layers in the traditional prosodic hierarchy (see figure 1.1) and presents an early account on a dynamic system wherein the origin of speech rhythm is represented as multi-timescale, parallel and coupled hierarchical entrainment that serves to direct attention to the salient points in time.

IN CONCLUSION, stress interplays with attention in its functions in speech comprehension. That is, stress harnesses attention from bottom-up through its acoustic prominence, and guides attention, top-down, through its temporal regularity. According to MSS, stress plays an invaluable role in speech processing, because it marks the boundaries of the words in an otherwise continuous speech signal and indicates to the listener when to initiate lexical access (e.g. Cutler & Carter, 1987; Vroomen & de Gelder, 1995; Cutler & Norris, 1988; Cutler, 1990; Cutler & Butterfield, 1992; Vroomen et al., 1996).

ABH more specifically focuses on the predictability of stressed syllables and proposes attentional resources to fluctuate while making sure attention peaks on the accents which can then optimally be analyzed (Pitt & Samuel, 1990). DAT proposes a similar mechanism to optimize processing by relying on the rhythmic structure in sound, but recognizes the complexity in rhythm and allows for tracking along multiple temporal dimensions (e.g. Large & Jones, 1999; Jones et al., 2002; Jones, 2010). Indeed, the time-scales in speech rhythm appear to be highly interrelated with the larger windows dominating the temporal locations of the smaller units (Cummins & Port, 1998; Port, 2003). Note, that this hierarchy of time-scales ensures that even if stress intervals are not strictly isochronic, they are still temporally predictable through the longer time-scale that determines their presentation in time (see also Byrd & Saltzman, 2003; Aylett & Turk, 2004; Turk, 2010).

Recall, that this is similar to the conclusion drawn in the previous section. Rhythm must be considered along the entire prosodic hierarchy to appreciate that the different layers impose their own constraints to the timing of speech events. Indeed, the idea that higher level prosodic structures control the rhythmic structure in speech and allow for local prominences to be temporally predictable, is echoed in another more recent theory on utterance planning (the Smooth Signal Redundancy Hypothesis; Aylett & Turk, 2004; Turk, 2010), according to which speakers actively plan the phonological form of an utterance in order to make less anticipated but important elements in speech more acoustically salient (see also Turk & Shattuck-Hufnagel, 2013, 2014).

So, metrical stress provides attention with a regular beat to which it can synchronize its limited cognitive resources. Surface-level accentuation, on the other hand, attracts attention and may serve to reset the attentional oscillation when speech is not perfectly periodic and the synchronization starts to stray. The term ‘accent’ thus refers to an abstract property of the word, crucial in its function during speech comprehension. Now that the accent has been defined, let us turn to the language investigated in the present dissertation: French—a language, said, without accent.

1.5 Stress in French prosody

French accentuation holds a low phonological and post-lexical status. French rhythm is traditionally held syllable based with syllables of approximately equal metrical weight. In connected speech, some syllables may receive additional prominence. These accents are, however, not considered to directly apply to the word domain, but, instead, thought to belong to the phrase. This means that the accents have post-lexical functions, and are not considered to contribute to word processing. For example, French accents may signal phrase boundaries or present the utterance's information structure, but they never distinguish the semantic content of a word. French accentuation is thus not lexically distinctive, and, as we will see below, tightly intertwined with post-lexical, intonational prominence.

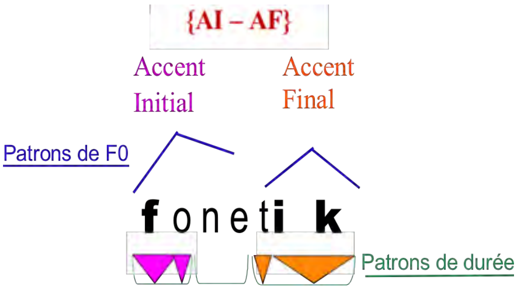
In 1980, Rossi published an influential article in which he discusses these two properties of accentuation and questions the existence of the accent in French (Rossi, 1980). The article had a substantial impact on the scientific community, since, soon after, the notion of French as *a language without accent* became the generally accepted view on French prosody, and accentuation was attributed a rather trivial role in speech processing.

Indeed, as some authors have argued, if French language does not know lexical stress, it is reasonable to assume that its speakers are confronted with stressed syllables too infrequently to be able to hear the accents (e.g. Dupoux et al., 1997). That is, the rare interactions with local prominences in 'a language without accent' are presumed insufficient for speakers to develop a sensitivity to accentual information, essentially leaving them 'deaf to stress'.[17] Because listeners can still readily decode speech, despite their supposed 'phonological deafness', it—according to these scholars—stood to reason that accentuation is unlikely to play an important function in French comprehension processes. Consequently, and understandably, French accentuation has attracted rather little interest in the linguistic field.

Below, we will examine in more detail why French accentuation is thought to belong to the 'group of words' and precisely which functional roles accentuation may have in this language. Next, we will present two Metrical Models, which distinguish surface level realizations from phonological representations and allow us to posit that French accentuation may belong to the word and play a much more valuable role in speech comprehension than is currently acknowledged.

---

[17] Note that while the term 'stress deafness', when taken literally, implies a phonological deafness for French listeners, and is in fact often interpreted as such, Dupoux et al. (1997) intended for a more nuanced interpretation, wherein speakers of languages with fixed, non-distinctive stress do not encode stress templates into their mental lexicon and are consequently less sensitive to variable, lexically distinctive stress in foreign languages.



	Final Accent (FA)	**Initial Accent (IA)**
Phonological status	primary and compulsory	secondary and optional
Location	last syllable word(group)	first syllable word(group)
Demarcative function	right boundary	left boundary
Phonetic characteristic:		
—*Primary*	duration (rime)	f_0 rise
—*Secondary*	f_0 movement	onset duration

Table 1.3: The schematic representation (on the left, extracted from Astésano 2016) shows a the phonetic characterization of IA and FA (here on the word 'phonétique', [fonetik] (English: phonetics)), in terms of tonal configurations and durational patterns. IA is characterized by an asymmetric local f_0 configuration, and by a short syllabic duration with significantly longer onset; FA is characterized by a symmetric f_0 configuration, and by a long syllabic duration with significantly longer rime. The table (on the right) provides a general overview of the characteristics of IA and FA. FA is the primary accent in French and obligatory marks the right boundary of AP. IA is the secondary accent, optionally marking the left group boundary.

1.5.1 Group level accentuation

French accentuation applies to the 'group of words'. Word groups consist of at least one content word and the associated clitic words (Jun & Fougeron, 2000). Depending on the author, the word group is referred to as 'clitics group' (groupe clitique; Garde, 1968), 'accentuel group' (Mertens, 1993), 'Accentual Phrase' (AP; Jun & Fougeron, 2000, see table 1.1) or 'rhythmic group' (groupe rythmique; Di Cristo, 1999). In the current work, the word group will mostly be referred to as the accentual phrase (AP; Jun & Fougeron, 2000), i.e. the domain just above the word (see figure 1.1).

Two (surface-level) group accents are generally recognized in French, the final accent (FA) and the initial accent (IA) (see table 1.3). FA is the primary stress, obligatory marking the right boundary of AP with a lengthened syllable rime, sometimes supported by an additional fluctuation in f_0. This accent is the compulsory accent in French and falls on the last syllable of the last word of AP, i.e. FA typically co-occurs with the right prosodic constituent boundary. The second accent, IA, is the secondary stress, marking the left boundary of AP. This accent is primarily cued by a rise in f_0 and a secondary lengthening of the syllabic onset (Astésano, 2001). IA is mostly associated with its rhythmic function, i.e. it intervenes when a long stretch of syllables is pronounced without FA (a so-called stress lapse, see section 1.3). So, while the accent is associated with the word level by some authors (e.g. Di Cristo, 1999; Vaissiere, 1997; Welby, 2003), IA is still regarded as secondary and optional, which defeats its function in the structuration of the speech stream (but see Astésano et al., 2007, and following).

As is discussed below, the accents may have functions other than boundary marking as well (e.g. they group words together, rhythmically structure speech or mark the utterance's information structure), but note that these functions remain post-lexical in nature.

As we have seen in section 1.2, accentuation is not insulated from intonation (nor is intonation from accentuation). This fact is especially pronounced in French. The intonation contour is made up of slow, global f_0 movements, while rapid, local variations in f_0 mark specific syllables in accentuation. Because accentuation and intonation rely on the same acoustic-phonetic parameter in French, local prominences near phrase boundaries will blend with intonation so that their phonetic parameters are spread and diluted over adjacent syllables (e.g. Rossi, 1980; Fónagy, 1980).Particularly regarding the French primary accent, FA, the claim is made that the accent perceptually disappears at the level of the intonational phrase (IP).

Recall, that this phonetic confound is not reserved to the French prosodic system. Teasing apart lexical and post-lexical stress based on their phonetic characteristics is problematic in many languages. It was therefore argued that the study of stress calls for a functional approach (see section 1.1). But, in French, accentuation and intonation overlap functionally as well. That is, accentuation interacts with intonation in the common purpose to delimit phrase boundaries.

Phrasal de-accentuation

In French, the primary stress (fa) on 'jolie' in 'Jolie Fille' may be de-accented to avoid a stress clash between the successive syllables carrying primary stress *and* to favor boundary marking of the accentual phrase, as in:

⇒ jolie fille

But an initial accent (ia) may surface to re-equilibrate the phrase:

⇒ jolie fille.

Figure 1.2: Example of an accentual phrase (AP) wherein the final accent on 'jolie' is reduced in favor of a clear marking of the phrasal boundary. Note, however that the de-accentuation also serves to evade a stress clash, which is universally dispreferred, and may lead to the initial syllable of 'jolie' being accented instead: the initial accent.

When a word is embedded into a phrase, accents within the phrase may be phonetically reduced, or de-accented (e.g. Di Cristo, 1999; Astésano, 2016), to favor a more prominent marking of the phrase boundary (hence the label *boundary language* for French, Vaissière, 1991). Take, for instance, the example in figure 1.2 wherein the French primary accent (FA) on the phrase internal word is de-accented such that the group boundary is more pronounced (Delattre, 1966; Rossi, 1980). This means that group marking takes precedence over word marking, resulting in an ambiguous functional distinction between the accent and intonation. In fact, as some scholars have argued, the interaction between the phonological entities may render the accent 'redundant' (Garde, 1968).

It is important to realize, however, that 'de-accentuation' does not mean that the accent is deleted and disappears completely. Instead, the accent is reduced to various degrees depending on rhythmic, contextual and pragmatic circumstances. This means both that 1) a trace of the local prominence survives and that 2) de-accentuation does not exclusively serve a clear marking of phrasal boundaries. In the example in figure 1.2, for instance, the de-accentuation also helped dodge a stress clash when the primary French accent (FA) located on the last syllable of 'jolie' was followed by the monosyllabic 'fille' (also carrying primary stress). As we saw in section 1.3, the occurrence of two consecutive stressed syllables is universally disfavored and may be avoided by restructuring the surface realization of the underlying prosodic representations (Liberman & Prince, 1977; Nespor & Vogel, 1983).[18] This is similar to the 'Rhythm Rule' presented in table 1.2, wherein the primary stress on the last syllable of 'thirteen' is (phonetically) reduced to evade the stress clash in 'thirteen men'.

In fact, where in English the suppression of a primary accent gives rise to a secondary accent on the first syllable, in French, de-accenting FA to evade stress clashes may lead to the first syllable of the phrase being accented instead (IA, see figure 1.2). This is one of the reasons for which the initial accent is interpreted as the secondary accent in French. The initial accent serves a number of functions, for example, the initial accent marks the left boundary of AP and helps group the words into a cohesive union (Di Cristo, 1999; Astésano, 2016). That is, the union of IA and FA, called *an accentual arch* (Fónagy, 1980), presents a bipolar stress template which underlies AP and groups the words it contains (see also Rolland & Lœvenbruck, 2002).

The initial accent is, however, not exclusively a result of the stress clash resolution. Similar to the role of secondary accents in English stress lapses, IA also serves a rhythmic balancing function to break long stretches of unaccented syllables, again contributing to its status as a secondary accent. So IA serves to both flag the beginning of a phrase as well as rhythmically balance the words within it. Finally, the accent is often confused with the emphatic accent. That is, the initial stress may also be expressive, pragmatically contrasting sentence meaning with an accentual emphasis.

1.5.2 Functional distinction between accents

Di Cristo distinguishes between the different functions of both the initial and final accent, and groups them into 'emphatic' and 'non-emphatic' accents (see figure 1.3; Di Cristo, 2000).

---

[18] Note that two stressed syllables can actually co-occur when they are separated by a pause, are manifested with different phonetic parameters, or hold different functions in the structuration of the speech stream (Astésano, 2016). Because IA is realized with different acoustic correlates than FA (see table 1.3), and likely holds different functions (which we will return to in chapter 6), the two accents can happily coincide. Combination of IA – FA or FA – IA pose no problem to speech encoding and decoding.

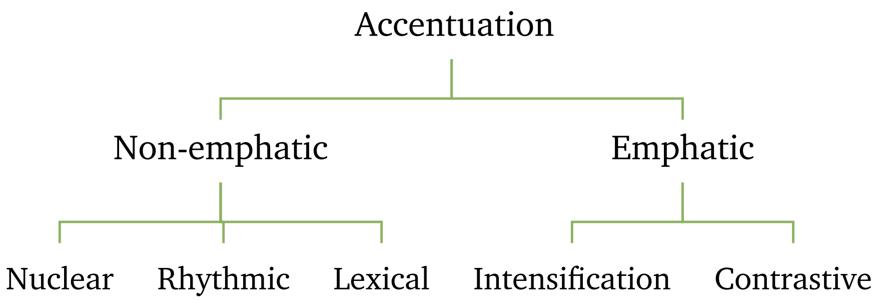


Figure 1.3: Di Cristo (2000) distinguishes accents based on their function in the organization of the stream. That is, both IA and FA are encoded in templates underlying the representation of words, but in their surface realization they may fulfill a combination of non-emphatic and emphatic function.

For instance, 'non-emphatic' accents can serve rhythmic, nuclear, and lexical functions. The **rhythmic accent** serves, as was mentioned above, a rhythmic balancing function (Di Cristo, 2000). On top of the role in rhythmic structuration, the accents may carry additional nuclear and lexical functions. That is, epending on speech rate or the utterance's information structure, any utterance can present multiple 'accentual groups' within one coherent 'intonation group'. Only the intonation group carries the **nuclear accent**, which is anchored to the intonation contour. It is this accent that is truly intertwined with intonation. The accentual groups, on the other hand, carry accents that come close to word level stress, i.e. for Di Cristo the **lexical accent** (see figure 1.3).

Di Cristo distinguishes these three types of accents from 'emphatic' stress. The first 'emphatic' accent, the **intensification accent**, expresses the speaker's emotional attitude and thus serves a para-linguistic function. The other 'emphatic' accent, the **contrastive accent**, conveys the utterance's information structure, i.e. it indicates the topic or focus of the utterance, and typically lands on IA. Di Cristo refers to this accent as a 'hyper-realization' of IA. Astésano (2001) suggests emphatic IA can be phonetically distinguished from structural/rhythmic IA by its longer rime and substantial movement in f_0 (see also Astésano, 2016). However, she is quick to add that the two types of accents are not binary in their surface realization. That is, structural/rhythmic IA can carry different degrees of additional emphatic stress; it is up to the speaker to what extent s/he wishes to (pragmatically) emphasize a certain word or syllable.

The functional division at the surface level raises an important question:

Do the roles of FA and IA in post-lexical speech processing negate the possibility for additional lexical functions, and is such a role in lexical processing dependent on surface realization?

When one takes into account the difference between surface-level prominence and underlying accentual representation (or "the difference between actual and potential accents"; Fox, 2000), it becomes possible to envision IA and FA maintaining a metrical weight at the level of the word, even when they are either not fully realized or 'hyper-realized' in the speech signal. That is, FA and IA may readily contribute to word processing by both their partial or hyper-realizations at the surface level—attracting attention from bottom-up—and their underlying metrical weight—harnessing attention through prediction.

1.5.3 Metrical weight for the French final and initial accent

Many excellent models have been proposed that describe the French prosodic system, but in the current work we will focus on Di Cristo (1999, 2000)'s Metrical Model which addresses French accentuation specifically regarding the distinction between surface realization and underlying representation (the reader is however referred to Astésano & Bertrand 2016; Astésano 2017; Garnier 2018 for elaborate overviews of the different models of French prosody). In his Metrical Model, Di Cristo underlines the distinction between acoustic reality and phonological perception and distinguishes phonetic realization from metrical weight (Di Cristo, 1999). That is, syllables can be metrically strong without necessarily being phonetically manifested (as we established in section 1.3).

The model defends Fónagy's accentual arch which is founded on **accentual bipolarisation** and **final dominance**, but imagines a different domain; where for Fónagy the accentual bipolarisation applies to AP, Di Cristo proposes the accentual pair is encoded at the lexical level in the form of latent, cognitive stress patterns. That is, latent stress patterns mark both left (IA) and right (FA) lexical boundaries with accents that are metrically strong even if their surface realization may depend on structural, rhythmic and pragmatic constraints. However, consistent with the second principle of final dominance, FA remains the primary marker of the right boundary of the word, while IA is presumed to take a more subordinate role as secondary and optional marker of the left boundary. Hence, in this Metrical Model (Di Cristo, 1999), IA is still considered a predominantly rhythmic device deemed unnecessary on constituents small enough for FA to provide the beat. Consequently, its occurrence is held less consistent than FA (Di Cristo & Hirst, 1993).

In a recent paper, Astésano & Bertrand (2016) question the optional and secondary nature of IA and suggest the accent to have a metrical weight that is similar to FA (see also e.g. Astésano, 2001; Astésano et al., 2012; Astésano, 2017). The authors further argue that part of the reason accentuation is undervalued in French, and that the nature of IA and FA are not well understood, is that most prosodic models have been intonation-based, leaving out the metrical and durational characteristics of French accentuation. As a result, local prominences marked not in pitch but in duration—or, indeed, the accents' more phonetically independent *metrical weight*—are often left unidentified. Astésano and Bertrand call for a clearer distinction between intonation and accentuation, which, indeed, do not serve the same purpose. Intonation presents the relationship between the different domains in the prosodic hierarchy and the modality of the utterance, while accentuation—as we have already seen—serves to:

1. Stabilize the syllable perceptually.
2. Signal word boundaries.
3. Bottom-up attract attention.
4. Top-down guide attention.

Indeed, they find support for both propositions in perception studies wherein syntactically ambiguous sentences were disambiguated based on prosodic constituent marking. Results showed that both FA and IA are perceived independent from prosodic boundaries (i.e. FA did not perceptually disappear under the intonation contour), indicating French accentuation not solely flags constituent boundaries but is metrically strong independent of them (e.g. Astésano et al., 2012; Garnier et al., 2016; Garnier, 2018). Furthermore, IA was found to be a more reliable cue to lexical boundaries and to regularly be perceived as more prominent than FA at both the phrasal and lexical level (Astésano et al., 2007, 2012; Garnier et al., 2016; Garnier, 2018). These results, therefore, challenge the secondary status of IA. That is, they show that IA is more than a rhythmic counterweight and more than a heavy emphatic stress, but, instead, actively involved in the organization of the speech stream.

A series of perception and neuroimaging studies further dispute the secondary status and show a strong anticipation for words to be marked with IA in their underlying cognitive representation (e.g. Jankowski et al., 1999; Astésano et al., 2007, 2012; Garnier et al., 2016; Garnier, 2018; Astésano et al., 2013; Aguilera et al., 2014, see also Roux et al. 2016 for a study on spontaneous speech). For instance, in a perception study wherein the phonetic parameters of IA were suppressed, IA was still readily perceived (Jankowski et al., 1999). This indicates that IA is phonologically expected by the listener. Recall from section 1.3, that a metrical beat does not depend on phonetic correlates to be perceptually real for the listener (*cf.* Lehiste, 1973). Perceiving the initial stress, even when its phonetic correlates were suppressed, underlines the metrical status of IA. Similarly, when the f_0 rise of IA peaks further along in the word, the prominence is still perceived on the initial syllable (Astésano & Bertrand, 2016; Garnier et al., 2016; Astésano, 2017; Garnier, 2018), contradicting the view wherein the place of IA within the word is variable, and questioning its status as "loose boundary marker" in French (e.g. Jun & Fougeron, 2000; Welby, 2003; Welby & Lœvenbruck, 2006). The results are therefore in line with Di Cristo (1999)'s notion of metrical stress templates underlying the word, but also show IA may be at least as heavy as FA.

Furthermore, recent neuroimaging studies using the event-related potentials technique (ERP; presented in section 3.2) have provided evidence against the notion of stress deafness in French. In a study directly addressing the perception of FA on monosyllabic words, participants showed little difficulty recognizing whether or not the word was marked with the primary stress (Michelas et al., 2016, see also Michelas et al. 2018 for a perceptual study of FA). Moreover, Aguilera et al. (2014) showed that IA (i.e. the secondary accent traditionally held to make less frequent occurrences) is not only perceived, but anticipated by listeners as belonging to the abstract representation of the word (see also Astésano et al., 2013).[19] The authors manipulated the phonetic realization of IA on trisyllabic words in an oddball paradigm. When the oddball had been presented without IA, a clear MisMatch Negativity component (MMN) emerged (Näätänen et al., 2007, see also section 3.2.1 for a presentation of the MMN). This MMN was however significantly smaller, when the oddball was presented with IA. Because an oddball paradigm typically elicits an MMN when a low-probability stimulus (the oddball) occurs within a train of high-probability stimuli, finding a reduced MMN when presenting the oddball with IA indicates a long-term representation of the accent and underlines the preference and expectation

---

[19] This study plays a central role in the current work and will be returned to regularly, most elaborately in section 6.1.

for stress templates with IA.

Finally, in another ERP study, this time investigating the relationship between metrical structure and late speech processing in French, it was found that lengthening the medial syllable on trisyllabic words obstructed semantic processing (Astésano et al., 2004; Magne et al., 2007).[20] In the study, participants listened to sentences in which semantic and/or metrical congruity was manipulated. Semantic congruity was manipulated by presenting sentences in which the last word was incoherent with the semantic context of the sentence, while metrical congruity was manipulated by lengthening the medial syllable of the last word, an illegal stress pattern in French. The metrical violation resulted in an increased N400 (a component held to reflect difficulties in lexico-semantic processing, described in section 3.2.3), even when the sentences were semantically congruent. This not only indicates that metrical patterns interact with word-level processing, but also that an acoustically prominent syllable does not necessarily facilitate speech comprehension.

Indeed, it was previously underlined that prominent syllables attract attention to the word by means of their acoustic saliency and as such facilitate lexical processing. However, crucially, the position of the word stress must be meaningful to the listener, i.e. the listener must expect the stressed syllable, presumably based on its cognitive representation underlying the word. In a language alleged without lexical stress, and wherein stress is not lexically distinctive (placing stress on whichever syllable never changes the meaning of the word), the attentional grasp could have facilitated semantic retrieval, regardless of its position on the word. But it did not, and in fact hindered word processing, further demonstrating the reality of stress templates in French which are encoded at the lexical level and clearly do not include stress on the medial syllable. The results thus reinforce the notion of bipolar stress patterns underlying the cognitive representation of words, as is suggested in Di Cristo's metrical model.

IN CONCLUSION, French accentuation traditionally holds a post-lexical status wherein the final primary accent (FA) and secondary initial accent (IA) together deliminate the group of words (i.e. AP; Jun & Fougeron, 2000). Depending on the intent of the speaker, these initial and final accents fulfill different emphatic or non-emphatic functions. However, it has been shown that it is not always easy to pin down the function of the stressed syllables, especially since they may serve multiple functions simultaneously.

---

[20] This study also holds great inspirational value in the current work and will be returned to regularly, notably in section 6.3.

Moreover, the speaker controls the surface realization of the accents, which may be more, or less pronounced depending on, for instance, speech rate or pragmatic considerations.

The same constraints hold for languages wherein lexical stress is generally assumed (e.g. English or Dutch). However, French differs from these languages in several respects. First, in French, stress is not lexically distinctive. This means that while the location of stress can differentiate between the semantic content of words in languages such as English or Dutch, it never does so in French. When an accent decides on the semantic content between minimal stress pairs, it is straightforwardly a property of the word. Indeed, as was argued before, whereas it can be difficult to determine an accent’s domain based on its surface realization, its functional role in distinguishing word meaning can be unambiguously attributed to the lexical domain. Because, in languages such as English or Dutch, stress is lexically distinctive, its lexical status is never questioned despite post-lexical factors modulating its surface realization. Conversely, because, in French, stress is not lexically distinctive, its phonological status is less readily assumed, which ultimately has led to the notion of ‘language without accent’ for French.

However, we argued that lexical stress does not *depend* on its function in distinguishing between minimal stress pairs, i.e. stress can also belong to the word domain when it is not lexically distinctive. To clarify this, consider the lexical value of a phoneme. The phoneme [eɪ] does not to a greater extent apply to the word b̀rain’ [breɪn] than it does to ‘train’ [treɪn] simply because, in the former, the phoneme is lexically distinctive (e.g. bran), while it is not in latter (tron, tran, trin, trun, . . .). Moreover, while the role in distinguishing semantic meaning is certainly a useful, and efficient, feature, minimal pairs are uncommon, even in languages with word level stress. Lexical stress will sooner serve in other lexical processes, such as word recognition, semantic retrieval or speech segmentation.

This brings us to the second difference between French and languages such as English or Dutch. Where in the latter languages, stress is generally assumed and its role in word processing has attracted much scientific interest, in French, accentuation is undervalued. Consequently, the contributions of IA and FA to lexical processing are relatively unexplored. Most of the description of French prosody is focused on the tonal or intonational characteristics of the language. This means that many studies only present the (nuclear) final accent that is intertwined with the intonation contour, while they miss the accents marked in metrical weight. The current work wishes to address this gap in the academic field. As Astésano and colleagues argue, (metrical) accentuation should be given a more prominent place in the descriptions of French prosody and not be regarded solely as subordinate to intonation. Work to that effect indeed suggests French listeners to have metrical expectations for words to be marked by IA and FA in their underlying stress template (e.g. Jankowski et al., 1999; Astésano et al., 2013; Aguilera et al., 2014; Michelas et al., 2016, 2018). In the current work, we seek to build on that work and determine to what extend presenting words without their expected bipolar stress template disrupts word level processing during speech comprehension.

1.6 Chapter summary

In this chapter, we have attempted to define the accent as an abstract lexical entity that is part of a prosodic organization, which is, itself, the coordinating principle behind the hierarchical framework for speech (Beckman, 1996). That is, the accent is one of the three phonological phenomena (accentuation, intonation and rhythm) which jointly structure speech into bite-sized chucks in order to facilitate speech processing. We have discussed how accentuation, intonation and rhythm each imposes its own sets of phonological rules to the speech utterance such that the surface form of any utterance, is the result of the collaboration between the three prosodic phenomena. Consequently, the accent itself rarely surfaces in its canonical shape but is instead regularly modulated by higher level and post-lexical re-structuration. In other words, the phonetic form of the accent is highly dependent on structural, pragmatic and rhythmic considerations. Therefore, we have argued that the accent refers to an abstract concept, attached to the representation of the lexical word which is independent from its phonetic manifestation.

We have shown that while the post-lexical confound on the phonetic realization of stress is cross-linguistic, it has proven especially troublesome for the study of French accentuation. In languages such as English or Dutch, the lexical status of stress is never questioned. In these languages, stress has a place in the dictionary. This means that even when the accent receives additional intonational emphasis or is reduced due to rhythmic constraints, no one doubts its phonological status nor its lexical domain. Moreover, in these languages, stress is lexically distinctive, i.e. it can differentiate between the semantic content of words. Because in those cases, words are recognized based on their underlying stress patterns, the accent's role in lexical access demonstrates its lexical identity. In French, accentuation does not have these advantages, i.e. in French, accentuation is not an entry in the dictionary and not lexically distinctive. However, we argued that, even if stress can not distinguish semantic content in French, it may still be metrically heavy and play an important role in lexical processing.

Indeed, presently, the role of French accentuation in lexical processing is ill-understood. Most descriptions of French prosody are focused on the tonal organization of the language, such that accents marked in duration or, more generally, *metrical weight* are rarely recognized. This means that IA and FA, which we have argued to underlie the representation of the word, are often overlooked. Consequently, they have attracted little interest in the linguistic field and their contributions in lexical processing have remained relatively unexplored. However, as we have shown, the studies that *have* addressed the representation of accentuation in French, paint a picture wherein French listeners expect words to be marked by both IA and FA (e.g. Jankowski et al., 1999; Astésano et al., 2012, 2013; Aguilera et al., 2014; Garnier et al., 2016; Garnier, 2018). The studies therefore point to a functional role for the accents in word-level processing, and suggest French accentuation plays a more substantial part in speech comprehension than is currently acknowledged.

Indeed, stress is generally known to be crucially involved in the process of speech comprehension. It actively contributes to processes such as speech segmentation, lexical access and even post-lexical processes such as semantic retrieval and integration. Stress interplays with attention throughout comprehension, i.e. it harnesses attention from bottom-up through its acoustic prominence, and guides attention, top-down, through its temporal regularity. According to MSS, stress plays an invaluable role in speech processing, because it marks the boundaries of the words in an otherwise continuous speech stream and indicates to the listener when to initiate lexical access (e.g. Cutler & Carter, 1987; Vroomen & de Gelder, 1995; Cutler & Norris, 1988; Cutler, 1990; Cutler & Butterfield, 1992; Vroomen et al., 1996). ABH more specifically focuses on the predictability of stressed syllables and proposes attentional resources to fluctuate while making sure attention peaks on the accents which can then be optimally encoded (Pitt & Samuel, 1990). DAT proposes a similar mechanism to optimize processing that relies on the rhythmic structure in sound, but recognizes the complexity in rhythm and allows for tracking along multiple temporal dimensions (e.g. Large & Jones, 1999; Jones et al., 2002; Jones, 2010).

In the next chapter we will look closer as why the theoretical framework described above place such an emphasis on the extraction and analysis of stress patterns. That is, what does speech processing precisely entail, which processes are involved in speech comprehension? But for now, keep in mind that French, traditionally a language without accent, was difficult to position in these theoretical frameworks. However, if we accept IA and FA to carry metrical weight, the accents are more readily integrated.

THAT IS, IA and FA then attract attention both by their acoustic salience (when they are realized, whether partially, fully or 'hyper') and by their metrical predictability. Such a perspective opens the door to a functional role for both accents in word-level processing during speech comprehension.

2 Speech processing

In order to appreciate the functional roles of French accentuation in the process of speech comprehension, it will be necessary to better understand the challenges a speech system faces when confronted with an acoustic speech signal. Speech perception is one of the hardest processes to model computationally and unfolds in three stages: 1) an auditory stage during which sound is spectrally decomposed, 2) a pre-lexical stage during which segmental and supra-segmental information (i.e. phonemes and stress templates, respectively) is matched to their phonological representations such that word form hypotheses can be derived and activated, and, finally, 3) a lexical stage wherein the activated lexical representations are evaluated and compete until one lexical representation can be selected for word recognition. Complicating this process, is the fact that, first, neither segmental nor supra-segmental information typically presents itself in its canonical form (i.e. the **variability problem**), and that, second, speech is a continuous signal with no cues consistently marking the boundaries between words or even between phonemes (i.e. the **segmentation problem**).

In the current chapter, the three stages involved in speech perception will be described in more depth, and we will discuss some theories on how the speech system confronts segmental and supra-segmental variability, and how it segments the signal despite the lack of consistent cues to boundaries between linguistic units. The two challenges to speech perception as well as their assumed solutions will be discussed through the presentation of well-known computational models on speech perception (i.e. the Cohort model, Marslen-Wilson & Welsh 1978; Wilson 1990, TRACE, McClelland & Elman 1986, and Shortlist, Norris 1994; Norris & McQueen 2008). The aim of the chapter is to show that—and why—speech comprehension requires more than a passive, bottom-up analysis of the sound signal, and instead relies on a top-down analysis of (amongst other information) metrical structure.

Indeed, we will see that, while speech perception was first modeled as a purely bottom-up and feed-forward process, initially reporting increasingly better performances (a popular quote back then was: "Anytime a linguist leaves the group the recognition rate goes up", in Davis & Scharenborg, 2016), these early improvements are now considered to be, for a large part, due to the steep development and increase in computer power in that period. That is, the way speech processing was modeled, was unrealistically inefficient and mostly relied on high CPU (i.e. computer power). When development in computer power slowed down, the computational models still under-performed compared to humans.

More recent models take into consideration that speech perception is likely aided by top-down prediction and cues hidden in the speech signal which help locate, for instance, word boundaries. That is, in newer models, pre-lexical cues are taken into consideration, allowing for more efficient models to speech processing. In section 1.4, we presented one of the possible cues which help confront the problem of speech segmentation and facilitate lexical access: metrical stress. The theory of metrical segmentation will be related to French, a language wherein, as we have discussed, the syllable is held to be its metrical unit. French listeners are therefore generally held to segment on the syllable, however, we will show that speech segmentation involves a more complex strategy. That is, we will show that the listener, also French, is more likely to use all cues available on an as-needed basis, and segment speech on multiple time-windows, i.e. along the entire prosodic hierarchy (i.e. syllable, foot and phrase), in parallel. Together with the notion of metrical weight for the French initial and final accents (Di Cristo, 1999; Astésano, 2001, 2017), this gives a new perspective to the role of accentuation in French in speech processing.

2.1 Three stages in speech perception

Sound arrives in the inner ears in the form of physical vibrations which are subsequently converted into an electrical signal that the brain can decode. Decoding initially involves passing the sound signal to the auditory cortex through the cochlea. The cochlea is a type of "frequency analyzer" that translates a range of acoustic frequencies (between 20 kHz and approximately 20 Hz) onto a so-called tonotopic map. The place where a frequency is encoded is mainly dependent on physical characteristics of the basilar membrane (e.g.

the stiffness and width of the membrane) in the cochlea. These physical characteristics vary gradually, such that each part in the cochlea is sensitive to a slightly different resonant frequency. The base of the cochlea is sensitive to high frequencies, and gradually becomes more sensitive to the lower frequencies towards the tip of the cochlea. So, the cochlea analyzes sound by way of what can be seen as a spatially distributed Fourier Transform.

This spatial selectivity also informs us about where the decoding of linguistic information takes place. In the lower, sub-cortical regions, the transient sounds with a high temporal resolution are processed, while the slower sounds are processed by and surrounding the auditory cortex. In other words, the auditory system is increasingly sensitive to more complex linguistic information; it extracts fine-grained phonetics (such as f_0) in the sub-cortical regions to then gradually continue on processing segmental (e.g. phonemes) and supra-segmental (e.g. syllables, stress) information near and beyond the primary auditory cortex (Hickok & Poeppel, 2015; Poeppel et al., 2008; Hickok & Small, 2015). Note that, as a result, already at this early stage, speech is processed hierarchically.

Speech perception is however more complex than the spectrotemporal analysis of the auditory signal. The listener must additionally extract meaningful units from the signal; s/he must recognize words. Word recognition involves at least two more processing stages; a pre-lexical stage, during which the phonological information required for lexical retrieval is extracted, and a lexical stage, involving the competition of multiple lexical candidates and ultimately the selection of the appropriate phonological word form.[1]

During the pre-lexical stage, segmental (for instance phonemes) and supra-segmental (e.g. syllables or stress patterns) information is extracted from the sound signal. Just as with lexical stress, the functional value of segmental information for lexical access is especially apparent in minimal pairs. In chapter 1, it was shown that stress templates can distinguish between segmentally similar word pairs (e.g. suspect refers to a noun, while suspect refers to a verb). Because these minimal pairs differ only in the location of the stress, word recognition clearly depends on the analysis of their underlying metrical stress patterns. The same holds for phonemic information. For instance, the word [breɪn] ('brain') differs from the word [treɪn] ('train') only in its initial phoneme. Therefore, in order to access the correct lexical representation, the word's segmental content must be analyzed beforehand

---

[1] Note that a word's phonological form refers to the word's phonological representation in the mental lexicon, and not to its semantic content. That is, word representations are generally assumed to be stocked in the mental lexicon and contain diverse information about the word, e.g. syntactic category, orthographic form, semantic meaning, and, its canonical acoustic realization which is also referred to as the word's phonological form.

(i.e. pre-lexically). So, word recognition requires a pre-lexical analysis, both of segmental and of supra-segmental information.

The necessity to extract and categorize phonemic and prosodic information is not restricted to the recognition of minimal pairs. Even for words that are not part of a minimal pair, phonological categorization facilitates word processing by *constraining* the *number* of lexical hypotheses that are considered in the lexical stage. In this lexical stage, the mental lexicon is searched for word form representations matching the pre-lexically analyzed phonological input. These word forms are activated and will compete for lexical access, a process that continues up until one word best matches the input and can be selected. This stage, then, involves two main processes: the evaluation of lexical candidates and the competition between them. The more discriminating the information fed to this stage, the faster and easier the process of word recognition. For instance, as we will see later, words with early uniqueness points (i.e. the point in the word where only one lexical candidate matches the input) are recognized faster than words with later uniqueness points (e.g. Wilson, 1990; Radeau & Morais, 1990), indicating the process of lexical access to be completed as soon as the speech system receives enough discriminating information to make a motivated decision.

In sum, speech perception unfolds in three stages. First, during the auditory analysis, the speech signal is spectrally decomposed and distinguished from non-speech sounds. Then, phonological information is assembled during the pre-lexical stage and passed on to the lexical stage in which lexical hypotheses are evaluated up until one word can be selected for lexical access. As such summarized, the process may seem relatively straightforward. However, as mentioned before, word recognition is one of the hardest processes to model computationally because of two challenges in speech: the **variability problem** and the **segmentation problem**.

2.2 Computational problems

The variability problem refers to the tendency for the acoustic-phonetic realization of phonological entities to deviate from their canonical representation. That is, both segmental and supra-segmental speech sounds are highly contextually variable. Indeed, in the previous chapter we saw that the surface realization of stressed syllables depends on a range of contextual (e.g. structural, rhythmic and pragmatic) factors. The same holds for the realization of segmental information, such as phonemes. Such variability often results in phonemes and stressed syllables that are, objectively, rather ambiguous. Still, despite such ambiguity in the speech signal, listeners typically appear unfazed, and rapidly waltz through the process of word recognition. Below, we will discuss two frameworks that have addressed the variability problem and help explain how listeners understand speech with such ease; the **abstractionist** (Eulitz & Lahiri, 2004, see also Cutler 2010 for a review) and the **exemplar** (Johnson, 1997, see also Pierrehumbert 2001 for an application specifically to linguistic information) frameworks.

The second major challenge, the segmentation problem, concerns the absence of consistent cues to boundaries between word or even between phonemes. We already encountered this problem when we presented Cutler and Norris's Metrical Segmentation Strategy (MSS) in section 1.4, which holds that listeners use metrical information to segment speech. Here, we will discuss a number of other devices available to the listeners, and their relation to the better known computational models that have addressed the problem of speech segmentation in word recognition: the Cohort model (Marslen-Wilson & Welsh, 1978; Wilson, 1990), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994; Norris & McQueen, 2008).

2.2.1 Variability problem: abstractionist or exemplar

Any speech segment can be pronounced in an infinite number of ways depending on factors such as co-articulation, speech rate, speaker identity and noise. For example, the phonetic outcome of phonemes constantly changes as a result of **phonological assimilation** (e.g. Nguyen et al., 2009; Nguyen, 2012). In assimilation, the place of articulation of adjacent segments shapes the acoustico-phonetic realization of the phoneme. For instance, a given phoneme will necessarily be differently pronounced when nearby a voiceless consonant, such as [p], or nasal consonant, such as [ŋ]. Micro-prosodic variation presents another example of context dependent variability. For instance, the fundamental frequency (f_0) of a given vowel will be higher when it is preceded by a voiceless consonant, than when it is preceded by a voiced consonant (Di Cristo, 1976). These examples show that the pronunciation of phonemes constantly deviates from their canonical representation due to segmental context.

Also linguistic structures larger than phonemes are affected. Obviously, as we have already established, the acoustic manifestation of a lexical stress is highly variable, but even the shape of speech sounds informing on larger structures or para- or extra-linguistic features such as speaker identity depend on many contextual factors. In fact, even if the *same* speaker were to repeat the *same* sentence guided by a metronome beat twice, the acoustic realizations are likely to differ. How then does the speech system map these ever-changing speech sounds to their corresponding representation?

In the **abstractionist theory** it is assumed that phonological units are encoded in the form of abstract, cognitive templates (e.g. Eulitz & Lahiri, 2004; Cutler, 2010). These templates are registered in long-term memory and phonetically underspecified (i.e. fine-grained phonetic detail is omitted). As such, they serve as the phonological archetypes to which the phonetically diverse speech sounds can be generalized. This view therefore clearly distinguishes between the surface phonetic form of speech events and their underlying phonological representation. Phonological templates are abstracted from actual speech input and, in speech production, they may be modulated or modified according to context, while, in speech perception, they serve as prototypes for incoming speech to be compared to for similarity.

Indeed, studies on word learning have provided evidence for this type of phonological generalization (e.g. Shatzman & McQueen, 2006; Sulpizio & McQueen, 2012, see also Cutler 2010 for a recent review). In the studies, listeners were taught pairs of words that sounded like words but did not exist in their language (i.e. pseudowords). Importantly, the pseudowords were presented without the suprasegmental information familiar to the listener, i.e. without the stress template typically underlying their lexical words. Results in word recognition showed listeners to generalize/map the prosodic characteristics of their language to the newly learned words.

Additional evidence for underspecification is provided by the method of event-related potentials (ERP; see section 3.2 for a description of this method). In these studies, underspecification is related to the principle of optimal or predictive coding in neuroscience (e.g. Friston, 2005, see also Scharinger et al. 2012, 2016). That is, in predictive coding, perception is less concerned with the fine-grained analysis of sensory information, but instead crucially depends on the ability to generate expectations about upcoming sensory input and compare or generalize the incoming, bottom-up information to those predictions. Bottom-up evidence that matches with the prediction is categorized with little effort, while only the mismatching information requires additional processing, a cognitive effort that is reflected in a modulation of the ERP (e.g. Scharinger et al., 2012, 2016).

In stark contrast to abstract representations, the **exemplar theory** assumes each encountered speech segment to be encoded in detailed, short-term memory traces called *exemplars* (e.g. Pierrehumbert, 2001, see Nguyen et al. 2009 for an overview of different types of exemplar models). These memory traces contain a vast amount of information, ranging from fine-grained acoustic structure, to speaker identity, to the specific situation wherein the utterance occurred. However, unlike the representations assumed in the abstractionist framework, exemplars will not be maintained indefinitely. If exemplars are not strengthened by other perceptually (nearly) identical speech sounds within a certain time limit, they expire and are 'forgotten'.

Exemplars are categorized into so-called *exemplar clouds*. These exemplar clouds carry a label representing the ensemble of the strongest exemplars they contain. With these labels, exemplar clouds may resemble, what can be considered, representations of canonical phonological units. As such, the exemplar clouds allow for the infinitely diverse speech sounds to be categorized into discrete phonological entities, again solving the variability problem. Note however that if the cloud is represented by the ensemble of exemplars which constantly change, then the phoneme in this theory is more continuous than discrete and certainly more continuous than the discrete representation in the abstractionist theory (Välimaa-Blum, 2009).

Support for the exemplar theory, again, is provided by studies on word recognition. In these studies, results indicate that high-frequent pronunciation variants are recognized faster than variants that are encountered less frequently (e.g. Connine, 2004; Connine et al., 2008; Pitt et al., 2011). Furthermore, it appears that some pronunciation variants of words are stored in the mental lexicon. For example, when Dutch listeners have to recognize that the shorter pronunciation of [natylək] → [tyk] (meaning 'of course'), means the same thing as the longer word, they appear to store the new variant rather than reconstruct the longer word through pre-lexical processes (e.g. Ernestus et al., 2002; Ernestus, 2014).

The truth on the representation of linguistic information likely lies somewhere in the middle, such that adaptability as well as representations specific to speakers can be accounted for (e.g. Nguyen et al., 2009; Nguyen, 2012; Eisner & McQueen, 2018). In the current thesis it is therefore not the intent to disambiguate between the two frameworks, although we will report results that are more in line with the abstractionist theory.

2.2.2 The segmentation problem

As was explained previously, the segmentation problem refers to the lack of spaces or cues that reliably and unambiguously mark the boundaries between words or even between phonemes in continuous speech. An additional difficulty that, as we will see below, is closely related to the segmentation problem, is the **embedding problem**. Due to there being only a limited number of phonemes available to any speech system, words often sound alike and can have other words partially or wholly embedded within them (e.g. ‘cap’ is a word on its own, but can also be the initial syllable of ‘captain’ or ‘capital’). This means that the speech stream usually matches with multiple lexical candidates. Still, the listener is perfectly able to segment continuous speech into separate words.

We will discuss the problem of speech segmentation through the presentation of three computational models that have, as their main goal, sought to address it: Cohort (Marslen-Wilson & Welsh, 1978; Wilson, 1990), TRACE (McClelland & Elman, 1986), and Shortlist (Norris, 1994; Norris & McQueen, 2008). Computational models on word recognition typically assume a simultaneous evaluation and competition of multiple lexical candidates that have been activated based on segmental and supra-segmental phonological information. They are often connectionist models with nodes symbolizing one or more levels of pre-lexical representations and nodes symbolizing the lexical representations. Upon auditory input, matching pre-lexical nodes are excited and, in turn, activate the lexical nodes they are connected with. Ultimately, the strongest activated lexical node ‘wins’ and is selected for lexical access.

Whether the strength with which lexical nodes are activated is gradient, depending on goodness-of-fit (i.e. whether processing is serial or cascaded, discussed below), and whether explicit boundary cues are considered, differs per model. Below, it will be explained how each of the models fares in face of constant variability and lack of consistent boundary cues in the speech signal.

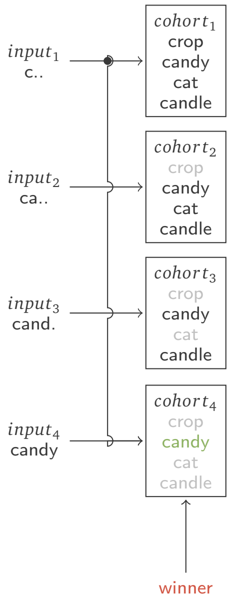


Figure 2.1: Cohort model presents speech processing as a serial process that continues up until the uniqueness point. Note that in Cohort there is no pre-lexical analysis of phonological information.

Cohort One of the first computational models to address word recognition is the Cohort model (figure 2.1; Marslen-Wilson & Welsh, 1978; Wilson, 1990). In the model, a word's initial phonetic sequence activates a set of segmentally similar lexical representations, called **the cohort**. As the speech signal continues, activated representations that seize to match the acoustic stream are disregarded from the cohort, while matching candidates are additionally activated. The Cohort model presents speech processing as a serial, bottom-up and feed-forward process that continues up until only one lexical item matches the incoming sound and can be selected.

The Cohort model proposes an elegant solution to the segmentation problem in the sense that, if words can be identified early in the process of word recognition and before their acoustic offset (i.e. at their **the uniqueness point**, the point at which there is definitive support for one particular lexical hypothesis and all other candidates have been disregarded), inferring the onset of a subsequent word is straightforward. That is, there is no need for word boundaries to be marked, because their onset can be determined through the recognition of the previous word. So, in the model, speech is segmented through lexical access.

Indeed, segmenting on lexical access has proved a commonly used strategy amongst listeners (e.g. Mattys et al., 2005). When listeners are asked to recognize unfamiliar words in a speech stream, they often rely on their lexical knowledge in a strategy called 'segmentation by lexical subtraction' (e.g. Mattys et al., 2005; Cunillera et al., 2010, 2016; Palmer et al., 2018). Further evidence for Cohort comes from the finding that words with an early uniqueness point are recognized faster than words with a later uniqueness point (e.g., Wilson, 1990; Radeau & Morais, 1990). Again, this indicates lexical access to be completed as soon as the cohort is reduced to one word, consistent with the Cohort model.

However, uniqueness points also present one of the model's biggest weaknesses. That is, the model too heavily relies on the assumption that the point at which the sound signal matches with one unique lexical candidate generally occurs before the word's offset, which is not always the case. Embedded words are especially problematic for Cohort, since, with these words, it is not possible to rule out longer competitors before the offset of a word.

Another assumption of Cohort that has since been heavily criticized, is the assumption that speech processing is serial. Results from a large number of priming and cross-splicing studies (e.g. Marslen-Wilson et al., 1996; Marslen-Wilson & Warren, 1994; McQueen et al., 1999; Dahan et al., 2001; Toscano et al., 2010; Gwilliams et al., 2018, see also McQueen 2007 for an excellent overview) point towards a more cascaded processing flow. In these studies, the fine-grained, sub-segmental acoustics of the initial consonant of a given word was rendered ambiguous. The idea is that, if speech processing is serial, then the ambiguous phoneme should be categorized during the pre-lexical stage of speech processing and lexical hypotheses will be equally activated. Conversely, if processing is cascaded, then the phonemic ambiguities may be relayed to the lexical stage where their ambiguity will modulate the activation strength of lexical candidates. As an example, the Voice Onset Time (VOT) of the voiceless stop-consonant [k] may be shortened such that the phoneme sounds increasingly more like the voiced consonant [ɡ]. Studies manipulating these voice onset times indicate activation levels of lexical hypothesis to match with the ambiguity of the phoneme, demonstrating postponed pre-lexical categorization. Processing is then not serial, but cascaded, and inconsistent with the Cohort model.[2]

Finally, the Cohort model is completely dependent on a correct analysis of the initial phonetic information and irreversibly fails when a word onset is not recognized. While onsets have indeed been found to play a privileged role in successful word recognition compared to the word's offset (e.g. Content et al., 2001a; Dumay et al., 2002; Sanders et al., 2002; Astheimer & Sanders, 2009; Breen et al., 2014), such failure does not do justice to human ability in restoring missed onsets, which occurs frequently especially in situations of noise. Moreover, lexical representations that differ in their initial segmental sequence, but share a global phonological similarity, are found to also be activated (Luce & Pisoni, 1998). That is, words do not necessarily start the same as the sound input to be considered as lexical candidates, there is also competition among words sharing similar segmental sequences at different locations.

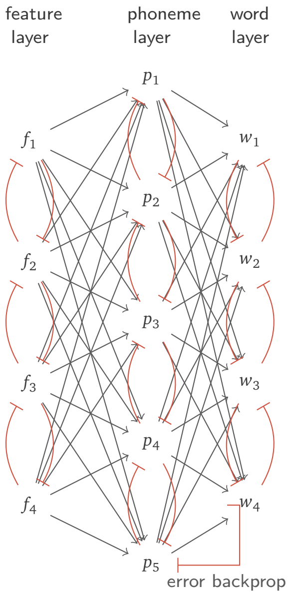


Figure 2.2: TRACE models speech processing as a three-layered architecture wherein features, phonemes and words are analyzed and compared in parallel. TRACE further implements both back-propagation and mutual inhibition.

TRACE In the Cohort model, lexical representation are activated based on their initial match with the auditory input and disregarded from the cohort as soon as the input diverges. In contrast, TRACE proposes an evidence-based account to lexical selection, allowing for words with global similarity to be activated as well (McClelland & Rumelhart, 1981; McClelland & Elman, 1986).

---

[2] Note that there is also evidence for a cascade of metrical or suprasegmental information up to the lexical stage, such that this (pre-lexical) information modulates the lexical competition process (e.g. Davis et al., 2002; Salverda et al., 2003; Shatzman & McQueen, 2006), suggesting suprasegmental information to be continuously passed forward in word processing.

That is, TRACE portrays lexical access as a highly interactive and competitive process, during which sequential time windows are analyzed separately and on three different planes (features, phonemes, and words; see figure 2.2).

Furthermore, while in TRACE lexical units are activated in parallel in a manner comparable to the Cohort model, lexical units are also interconnected with inhibitory links which rule out mutually exclusive lexical candidates in a mechanism called lateral inhibition. That is, the activation flow is bidirectional and spreads both from the lower levels to the higher levels (i.e. cascaded influence of pre-lexical sub-segmental information on lexical processing) as well as can backtrack from high to low (i.e. top-down lexical bias on phoneme categorization)[3].

However, in the connectionist model, the analyzed temporal information is represented spatially (i.e. different nodes for each time segment) which poses a problem when the same linguistic unit, for example a syllable, is repeated (Davis, 2003). Moreover, words are selected purely through activation between levels, and inhibition within levels. That is, lexical analysis is completed on the entire network of interconnected feature, phoneme and lexical representations, which get reanalyzed at each new time step. Finally, segmentation is still based on lexical access, with word offsets marking the lexical boundaries. That is, similar as in Cohort, in TRACE, lexical hypotheses are constrained only on their segmental structure, despite the abundant evidence for supra-segmental information to play a valuable role in word recognition (e.g. Cutler & Clifton, 1984; Cutler & Norris, 1988; Mattys & Samuel, 1997). These properties make the TRACE model biologically unrealistic and unreasonably inefficient.

Shortlist In the Shortlist model, the speech stream is segmented by relying on statistical cues to word boundaries rather than looking for words (Norris et al., 1995; Norris & McQueen, 2008). That is, the model relies on phonotactic and supra-segmental information by decoding syllable probabilities and metrical structure in parallel. Specifically, the model incorporates

---

[3] Note that demonstrating phonological processing to actually involve lexical feedback is far from straightforward. That is, while it has been shown that pre-lexical, phonological processing can be affected by lexicality and by word-frequency (phoneme monitoring (e.g. Cutler & Carter, 1987), phonemic ambiguity resolution (e.g. McClelland & Elman, 1986, known as the Ganong effect), and phonemic illusion effect (e.g. Samuel, 1981; Samuel & Ressler, 1986; Samuel, 1991; DeWitt & Samuel, 1990; Samuel, 1996), the results are equally compatible with postponed phonological processing.

the combined constraints of the Possible Word Constraint (PWC; Norris et al., 1997) and the Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988) to bound speech segmentation.

In the Possible Word Constraint (PWC; Norris et al., 1997), boundaries around one or a sequence of consonants that can never be real words are disallowed. This means that, for example in English, 'fapple' may not be segmented in 'f' + 'apple', because 'f' cannot be a real word in that language. Other phonotactic statistics concerning which consonant clusters can occur within versus between syllables, or which sequences are more likely to be at the boundary of a word, also serve to cue word boundaries. For instance, it is assumed that illegal or infrequently occurring segmental sequences are likely to contain a lexical boundary, so that these sequences encourage segmentation in Shortlist. In English, for example, consonant sequences such as [br] are allowed as word onsets (e.g. [breɪn], 'brain'), but are never at a word's offset. Conversely, sequences such as [kt] may be found at a word's offset (e.g. [sʌspɛkt], 'suspect'), but not at its onset. Also, a sequence such as [ntʃbr] is assumed to contain a word boundary (e.g. [lʌntʃbreɪk], 'lunch break').

The Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988; Cutler, 1990), previously presented in section 1.4, holds that listeners rely on their language's metric structure to infer word boundaries. Recall that in languages such as English or Dutch, rhythmic structure is stress based, and lexical stresses are often word initial (Cutler & Carter, 1987; Vroomen & de Gelder, 1995). This statistical prevalence makes stressed syllables reliable cues to the locations of word onsets in continuous speech. Moreover, stressed syllables are perceptually stable (surviving under noise) and acoustically salient, hence automatically attracting attention. Therefore, in Shortlist, all (English) lexical candidates are given a boost when they begin with a stressed syllable.

In sum, similar to TRACE, in Shortlist word candidates are activated based on their segmental overlap with the input signal. But then, a shortlist is created with only the segmentally and prosodically most likely lexical candidates, i.e. additional pre-lexical classification further constrains the search space. Only the shortlisted word hypotheses will compete in the following interactive activation process leading to recognition, making Shortlist a much more efficient model.

2.3 The role of the syllable in French speech processing

In the previous section, we saw that the processes of lexical access and speech segmentation are likely more efficient if the speech system can pre-lexically constrain the number of word hypotheses and identify word onsets (Briscoe, 1989), for instance based on metrical structure. However, whereas in English or Dutch, stress is the metrical unit underlying the languages' rhythm, the French metrical unit is considered to be the syllable (Pike, 1945; Abercrombie, 1976; Cutler et al., 1986).[4] In French, all syllables remain full, as opposed to, for instance, English, wherein unstressed syllable-vowels may be phonetically reduced. Furthermore, syllable structures tend to be open (vowel coda) in French (Adda-Decker et al., 2005, in Shoemaker 2009), allowing for syllable boundaries to be easily recognized. This is in contrast to the English syllable, wherein boundaries tend to be more ambiguous, also referred to as **ambisyllabicity**. For instance, the [l] in 'palais' (palace) is assigned to the final syllable by French listeners (Content et al., 2001a), while English listeners tend to be undecided about whether the [l] of 'balance' belongs to the onset of the second syllable or the coda of the first (Kahn, 1980, in Content et al. 2001a).

Support for a privileged position for the syllable in French comes from a series of studies wherein French listeners were found to detect phoneme sequences faster in words in which the sequence constituted a whole syllable, than in words in which the sequence contained a syllable boundary (Mehler et al., 1981; Cutler et al., 1986), while English listeners were insensitive to the difference (Cutler et al., 1986). For instance, Mehler et al. (1981) presented French participants with a visual cue of a cluster of phonemes (e.g. 'pa' or 'pal') followed by an auditory presentation of a word that started with the same sequence (e.g. 'palace' or 'palmier'). Participants were asked to indicate as quickly as possible whether the visual cue was part of the target word or not. Results showed that French participants were faster to respond when the sequence of phonemes did not straddle a syllable boundary (Mehler et al., 1981).[5] That is, 'pa' was detected faster in 'pa-lace' than in

---

[4] Recall that there is no real evidence for the rhythm classes division and, moreover, as was explained in section 1.5, IA and FA may hold metrical weight as they underlies the representation of the word, and underlie the representation of the word.

[5] Note that, Content et al. (2001b), although partially replicating these results, found a syllable effect in French only where the critical consonant was a liquid (such as the [l]) and that even for such stimuli the effect relied on several experimental factors.

'pal-mier', while, conversely, 'pal' was detected faster in 'pal-mier' than in 'pa-lace'. Mehler and colleagues interpreted this effect to mean that French listeners were **syllabifying**, meaning that the listeners were segmenting the words into whole syllables to prepare for lexical access. It is then through the syllable that they accessed their mental lexicon.

Directly comparing French speech segmentation to segmentation by English listeners, Cutler et al. (1986) tested speakers of both languages with an English and French version of the task used in Mehler et al. (1981). They obtained comparable results for the French participants (even in the English version), while there was no evidence of syllabification for the English participants (i.e. similar response times for 'ba' or 'bal' in both 'ba-lance' or 'bal-cony'). Cutler and colleagues were, however, skeptical that listeners should have segmentation strategies unique to their language. They argued that there should be a universal rule behind the syllabic segmentation for the French listeners and the stress segmentation for the English. And so they proposed the universal segmentation strategy based on meter: the Metrical Segmentation Strategy (MSS), wherein English (traditionally classified stress based) segment on metrical stress, and French (traditionally classified syllable-timed) segment on the 'metrical syllable'.

However, an alternative theory is provided by Content et al. (2001a) in the Syllable Onset Segmentation Heuristic (SOSH). According to SOSH, listeners tackle word segmentation based on trial-and-error. That is, listeners concede that there is no linguistic cue consistently marking the word boundaries, but assume that strong syllable onsets are most likely to coincide with the onsets of words. Importantly, in the model, it is emphasized that is the *onsets* of syllables that cue lexical access, and not the syllable itself (as is presumed in MSS). As Content and colleagues argued, while the syllable onset may flag the beginning of a word, its offset is much less informative and unlikely to be used in speech segmentation or lexical access.[6]

The heuristic accounts for differences between French and English segmentation strategies. In English, unstressed syllable vowels tend to be phonetically reduced, while French vowels remain full. This means that where the MSS proposes the English listener segments on metrical stress while the French listener uses the 'metrical syllable', SOSH proposes strong syllable onsets to signal the word boundaries for both listeners.[7]

---

[6] Recall that the Cohort model assumes a similar privileged role for onsets (Marslen-Wilson & Welsh, 1978; Wilson, 1990).

[7] Note how this theory puts IA in new light: if IA is strong both acoustically and metrically *and* signals word onsets, segmenting on IA provides listeners with a much more efficient heuristic than the 'general' syllable onset.

SOSH finds support in a word-spotting study by Dumay et al. (2002) wherein misaligned syllable *onsets* hindered processing more than misaligned syllable *offsets*. Furthermore, similar hindrances to word recognition were observed in stress based languages such as Dutch, indicating SOSH could be applied cross-linguistically (Vroomen et al., 1996). However, SOSH has also been criticized. First, inferring word boundaries at each syllable onset arguably results in many unsuccessful attempts to lexical access, making the strategy rather inefficient (although see footnote 7). But, also, syllable boundaries are not that easy to detect, especially not in French. That is, in French, phonological processes frequently blur word boundaries, such as in the case of the highly common French liaison (e.g. Wauquier-Gravelines, 1999; Dumay et al., 2002; Shoemaker, 2009), arguably making the syllable a less reliable cue in processes such as lexical access or speech segmentation.

It should be pointed out though, that neither MSS nor SOSH claim strong syllable onsets or strong metrical syllables to be the *only* cue to lexical access. In their 1986 article, Cutler and colleagues explicitly add that speech may provide multiple signals for the listener to use when confronting the segmentation problem, and that under certain circumstances, segmentation is likely guided by more than one by of them (Cutler et al., 1986). Similarly, for SOSH it is made clear that initiating lexical access on strong syllable onsets is a *heuristic* strategy, which can be adjoined or even replaced when other cues are also available (Dumay et al., 2002).

2.4 Hierarchical weights between segmentation cues

The studies presented so far, have, in their investigation of speech segmentation and word recognition, mostly zoomed in on one particular lexical or pre-lexical cue. However, listeners typically have a range of cues available to them. That is, besides lexical knowledge, word stress and phonotactic probabilities, word boundaries can also be marked by supra-segmental lengthening, de-coarticulation,[8] pauses,[9] and many more, which the listener appears to take advantage of on an as-needed basis.[10]

Mattys et al. (2005) recognized that listeners have a selection of strategies to choose from, which may not always point to the same segmentation solution. They developed a hierarchical framework for speech segmentation, wherein certain cues take precedence over others, mostly dependent on listening circumstance (see figure 2.3).

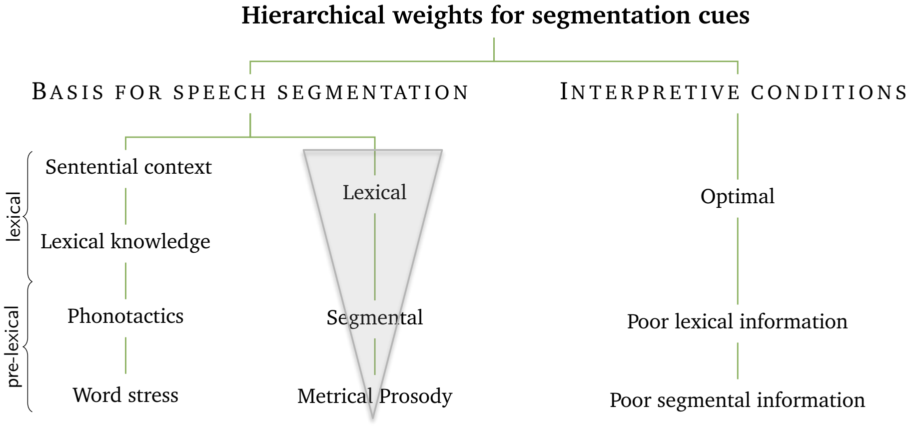


Figure 2.3: Mattys et al. (2005)'s hierarchical approach to speech segmentation. The width of the inverted triangle indicates the relative weight of the segmentation cue. Segmentation is dynamic such that listeners continuously adapt their segmentation strategy according to listening circumstances and availability of information. (Adapted from Mattys et al. 2005)

That is, they showed that in ideal listening circumstances, with clear and distinct speech, segmentation based on lexical knowledge sufficed (see also White et al., 2012). However, when speech was more noisy, listeners additionally employed pre-lexical strategies for segmentation. In conditions wherein the speech signal was especially corrupted, supra-segmental or prosodic cues proved most effective. In other words, where word recognition and speech segmentation can be achieved through lexical knowledge, pre-lexical cues

---

[8] Reduced co-articulation also cues word boundaries. Within words, phonemes are often co-articulated, i.e. they phonetically overlap. However, phonemes near word boundaries tend to be strengthened, reducing the phonetic overlap (e.g. Fougeron & Keating, 1997, see also Mattys 2004).

[9] Although silences are generally considered unreliable cues to word boundaries, notably due to co-articulation between words and plosive consonants, pauses have been found to facilitate speech segmentation in infant-directed or hyper-articulated speech. Especially when learning a new language (whether the first language for infants or a second language for adults), pauses can help listeners parse the speech stream. Note, however, that in these contexts, pauses will predominantly delimit small prosodic phrases, rather then each individual word.

[10] Note that IA and FA both mark word boundaries (left and right, respectively), which, although seemingly redundant, may have a functional value such as signaling different constituent domains or signaling lexical access versus speech segmentation, as we will be discussed further in chapter 6.

are ignored by listeners, however, in more natural speech contexts, listeners must employ strategies wherein they use pre-lexical cues to infer word boundaries and identify individual words. Recall that stressed syllables are, indeed, often more clearly articulated and prosodic information survives under conditions of noise, which is less likely for segmental information. Additionally, prosodic prominence typically renders stressed syllables more phonetically salient, such that they automatically attract attention (*cf.* Vroomen & de Gelder, 1995; Quené & Koster, 1998). It is then understandable that listeners should turn to the prosodic information to understand speech in more natural and noisy situations.

However, critically missing thus far is the use of information outside speech perception, i.e. contextual information also prepares the listener for lexical access. For instance, while oronyms such as 'I scream—ice cream' or 'that's tough—that stuff' may be distinguished by a pre-lexical analysis of fine-grained supra-segmental durational information, such an analysis is not always necessary and may be resolved a priori based on context.

There is some disagreement as to whether or not contextual knowledge mainly influences initial lexical activation levels (as would be assumed by, for instance, the theory of predictive coding) or the outcome of lexical competition (see e.g. Eisner & McQueen, 2018, for a discussion). Furthermore, some studies suggest that while contextual information is indeed used in word recognition, acoustic-segmental information takes precedence (e.g. van den Brink et al., 2001; Dahan & Tanenhaus, 2004; McQueen et al., 2009), but others show that listeners ignore pre-lexical cues (such as phonotactic regularities) when words can be recognized based on post-lexical information (e.g. White et al., 2012; Kim et al., 2012, see also Mattys et al. 2005, presented above). Indeed, in some cases, the listener's personal experiences may bias him towards specific segmentation solutions such as in the famously misinterpreted oronym "D'you":

> " You know, I was having lunch with some guys from NBC, so I said: 'Did you eat yet or what?'...
>
> and Tom Christie said: 'No, jew?'
>
> Not 'Did you?'! Jew eat? Jew! You get it?! Jew eat! "
>
> *Woody Allen, in Annie Hall, 1977*

Contextual constraints guide segmentation and word identification in a top-down, forward looking fashion by allowing the listeners to predict. These predictions can be motivated on the utterance's previous semantic content, syntactic structure, phonological form, and—of course—its rhythm.

2.5 Parallel temporal segmentation

We have discussed how supra-segmental information, and, in particular, stress, provides a solid cue to word boundaries. This is because stressed syllables are perceptually stable, acoustically salient, often located at word onsets, but also, because they are temporally predictable (e.g. Fant et al., 1991). Stress patterns make up a metrical framework that guides attention during speech segmentation (e.g. Martin, 1972; Cutler & Foss, 1977; Pitt & Samuel, 1990). That is, prominences guide the listener through the speech stream by providing a beat that the listener can use to anticipate the timing of the next stressed syllable. The listener can then direct attention in an efficient, forward-looking manner and maximize encoding resources at stressed syllable, while planning for decoding to proceed during time windows with less salient information (e.g. Pitt & Samuel, 1990). That is, if accentuation is metrically strong and also cues lexical boundaries—such as the French initial and final accents—such a predictive mechanism clearly should lead to faster word recognition and facilitated speech segmentation.

Indeed, we are reminded of the Attentional Bounce Theory (ABH; Pitt & Samuel, 1990), presented in section 1.4. In ABH, the listener anticipatory relies on metrical structure for speech processing. The listener can then selectively tune attention to "bounce" from one stressed syllable to the next and derive predictions on where to segment speech and when to initiate lexical access. Furthermore, as was argued in section 1.1, listeners may also rely on predictive information from other domains in the prosodic hierarchy. Indeed, rhythm must be considered along the entire prosodic hierarchy to appreciate that the different layers impose their own constraints to the timing of speech events (Cummins & Port, 1998; Port, 2003; Aylett & Turk, 2004; Turk, 2010; Turk & Shattuck-Hufnagel, 2013, 2014).

For example, listeners may furthermore rely on the intonation contour to predict the temporal location of valuable information. Studies on the use of prosody in English speech comprehension have shown that a preceding intonation contour facilitated phoneme detection if this preceding contour predicted the phoneme to be located in a stressed syllable (Cutler, 1976, see also section 1.4) and intonational boundary cues have been found to take precedence over phonotactic cues in artificial language segmentation. For instance, when Shukla et al. (2007) directly compared the use of phonotactic probability transitions within versus across IP-boundaries by Italian listeners segmenting an artificial language, they found that participants recognized words only when they occurred within the phrases, but not when they straddled the boundaries. The authors argued that prosodic boundary cues (i.e. the declining pitch contour) were used as "filters" to suppress possible statistically well-formed words that occur across Intonational Phrase boundaries. In fact, because pitch movements at intonational boundaries (e.g. the left initial rise or right falling pitch) are observed across languages, these cues could well be universal.

Furthermore, also cross-linguistically, both right and left edges of words and higher-level prosodic domains tend to be lengthened in duration (e.g. Fougeron & Keating, 1997; Turk & Shattuck-Hufnagel, 2013; White et al., 2015, see also chapter 1) and phrase boundaries have indeed been found to also cue speech segmentation in French (e.g. Banel & Bacri, 1994; Bagou et al., 2002; Rolland & Lœvenbruck, 2002; Welby, 2007; Spinelli et al., 2010). Note that this additionally means that, while in this chapter speech segmentation mostly referred to the chunking of utterances into individual lexical words, it is reasonable to assume segmentation unfolds over multiple time-windows in parallel (e.g. Poeppel et al., 2008; Giraud & Poeppel, 2012; Ghitza et al., 2012):

LISTENERS may be segmenting speech into phonemes *and* syllables *and* feet *and* phrases.

At this point, we are reminded of the Dynamic Attending Theory (DAT; Large & Jones, 1999, presented in section 1.4), which also assumes that the degree of temporal predictability of auditory information dynamically entrains attention to particular points in time, but further predicts that in complex rhythms, such as speech, several nested oscillators may synchronize to the temporal regularity in time-windows of different durations (Jones, 1976; Large & Jones, 1999; Jones et al., 2002; Large & Snyder, 2009). Attention, according to DAT, then both tracks the hierarchy of predictable events as well as groups them into nested domains.

We noted, in section 1.4, that while DAT does not directly address speech processing, it does provide insight on how listeners deal with complex rhythmic structures such as the hierarchically organized layers we find in speech. Combined with Cummins and Port's adaptive oscillator model on metrical structure in speech production (Cummins & Port, 1998, also see section 1.4), we can derive a model wherein speech processing is portrayed as a dynamic system that uses temporal expectations to entrain attention to multiple parallel time-scales so as to facilitate comprehension.

Indeed, recent developments in the field of neuroscience have suggested that neural oscillations may play a significant role in the segmentation and processing of the speech stream (e.g. Giraud & Poeppel, 2012; Ghitza, 2011, 2013; Peelle & Davis, 2012; Gross et al., 2013). Neural oscillations are important because they modulate the excitability of neural networks. As we will discuss in more depth in chapter 3, the intracellular peaks and troughs in a neural oscillation influence the probability with which neurons fire (Buzsáki, 2009). If we are to assume the theory of predictive coding (e.g. Friston, 2005), this means that the temporal regularities in speech should entrain the ongoing neural oscillations such that excitability aligns to the relevant points in time, simultaneously segmenting speech and facilitating speech processing.

2.6 Chapter summary

In the current chapter, we described how speech processing unfolds in three stages (i.e. an auditory stage, a pre-lexical stage and a lexical stage) which are computationally complex due to the variability problem and the segmentation problem. Additionally, we discussed three models that have attempted to present how listeners deal with these problems in speech processing during speech comprehension.

In the Cohort model, speech processing is serial and feed-forward, with no pre-lexical analysis of phonological information (Marslen-Wilson & Welsh, 1978; Wilson, 1990). That is, the incoming phonetic information is not mapped onto phonological entities, but, instead, directly activates lexical candidates with similar onsets.[11] Thus in the Cohort model speech compre-

---

[11] Note that, in this sense, the Cohort model seems to align with the exemplar framework, but, on the other hand, the lexical candidates are represented as discrete nodes, which is more in line with the abstractionist framework.

hension involves two processes, i.e. lexical access and lexical evaluation. During lexical access, word representations that match the initial input are activated into the cohort in parallel. As speech unfolds, representations that no longer match the input are removed from the cohort in the evaluation stage up until only one word remains in the cohort and can be recognized (i.e. at the uniqueness point). The model therein heavily relies on word onsets and on uniqueness points, which poses a problem in the recognition of embedded words and does not do justice to human ability in the restoration of missed word onsets.

TRACE is a computationally implemented model of speech processing which implements lexical activation in a three layered (i.e. features, phonemes and words) connectionist architecture (McClelland & Elman, 1986). That is, as opposed to Cohort, TRACE includes a pre-lexical stage to word recognition and accounts for the activation of embedded words in allowing parts of the speech input other than the onset to activate lexical representations as well. In this architecture, phonemic features extracted from the speech input initially activate the nodes in the feature layer, which in turn spreads to the corresponding nodes in the phonemic and word layers. However, lexical analysis is completed on the entire network of interconnected feature, phoneme and lexical representations, which are additionally activated or inhibited at each new time step, making TRACE biologically unrealistic and unreasonably inefficient.

Finally in Shortlist, the speech stream is segmented by relying on pre-lexical segmental and supra-segmental cues to word boundaries rather than looking for words (Norris et al., 1995; Norris & McQueen, 2008). Shortlist is made up of two layers, i.e. the input layer and the word layer. Upon speech input, a search is performed to find a small set of best matching words, which are subsequently 'shortlisted'. To create this 'shortlist', the model incorporates PWC and MSS to constrain word hypotheses. This means that, in Shortlist, pre-lexical cues—including metrical structure—are taken into consideration, allowing for a more efficient model to speech processing. However, while segmentation and word recognition based on metrical structure may be a successful strategy in languages with lexical stress, it is arguably less efficient in languages in which the domain for metrical rules is not the lexical word.

In French, stress is held to apply to the phrase and not to the lexical words. Furthermore, French is often described as a syllable based language with fairly homogeneous metrical weight on syllables. Consequently, it is held that the French metrical structure is defined by the syllable and that the syllable is used as the basic unit for segmenting speech (Mehler et al., 1981; Cutler et al., 1986; Content et al., 2001a; Dumay et al., 2002). However, it was argued that speech may provide multiple signals for the listener to use when confronting the segmentation problem, which the listener can make use of on an as-needed basis (e.g. Cutler et al., 1986; Wauquier-Gravelines, 1999; Dumay et al., 2002).

Moreover, in this thesis, it is hypothesized that IA and FA are phonologically encoded and metrically strong (*cf.* Di Cristo, 1999), thus providing two lexical entries (i.e. IA at the left lexical boundary and FA at the right boundary of the word). If IA and FA contribute to the metrical organization of French and apply to the domain close to the word, French accentuation could play a more prominent role in speech processing than is currently acknowledged, and allow listeners to segment both on the syllable and on the stress. Indeed, both primary FA *and* secondary IA have been shown to guide French listeners in the segmentation of speech (e.g. Rolland & Lœvenbruck 2002; for use of FA, see Banel & Bacri 1994; Bagou et al. 2002; for use of IA, see Welby 2007; Spinelli et al. 2007, 2010), although, due to the alleged domain of French accentuation, it was assumed that the accents helped listeners segment the accentual phrases (AP; Jun & Fougeron, 2000), and not the lexical word. Moreover, in those studies, neither IA nor FA were considered to aid the listener by means of their metrical weight. That is, IA was not considered a metrical stress, but rather a "loose boundary marker", and FA only functions as a correlate to the phrase boundary, concomitant to the intonation contour.

We argued listeners may be segmenting the speech stream over multiple time-windows in parallel, which recent developments in neuroscience seem to confirm (e.g. Poeppel et al., 2008; Giraud & Poeppel, 2012; Ghitza et al., 2012). For French listeners, this means that they may be segmenting speech into syllables *and* phrases—***and***, if, as we argued, IA and FA hold metrical weight and are encoded at the lexical level, also in words. That is, the metrical structure provided by FA and IA allows for prediction, and prediction impacts the ease of information processing by selectively directing attention in an effort to maximize encoding resources at the following stressed syllable and plan decoding to proceed during time windows with less important information (DAT and ABH; Large & Jones, 1999; Pitt & Samuel, 1990), providing a new perspective on the role of the French accents in speech comprehension, wherein IA and FA alternate to mark boundaries at the lexical level (*cf.* Astésano, 2017).

In the next chapter, we will discuss evidence supporting the notion of parallel temporal segmentation through brain oscillations which track regularities in speech. Indeed, evidence suggesting auditory networks to entrain to speech in order to segment the speech stream into bite-sized portions for analysis is accumulating rapidly, additionally linking the mechanism to functions in, for instance, grouping, intelligibility and gating (e.g. Sohoglu et al., 2012; Zion Golumbic et al., 2013; Doelling et al., 2014; Chait et al., 2015; Ding et al., 2016a; Mai et al., 2016; Cope et al., 2017; Keitel et al., 2017). However, the mechanism is seldom associated with prosodic processing, and when prosody is considered at all, researchers will usually attempt to suppress the prosodic information from the speech signal.

More advances on the predictive value of prosodic structure in guiding attention during speech processing is provided by studies using the method of Event-Related Potentials (ERP). Indeed, results from studies using this method (i.e. the brain's neural correlate for anticipation and prediction error, presented in section 3.2) show metrical regularity to facilitate lexical processing (e.g. Rothermich et al., 2010, 2012; Bohn et al., 2013), presumably through the interactions between prosody and attention, and their combined effect on neural excitability.

BOTH METHODS are presented in more depth in the next chapter, but note now that, while speech processing has been difficult to model computationally for many decades, the new interdisciplinary approach wherein linguist, neurolinguist and psycholinguistist researchers relate metrical structure to cognitive attentional control and their effect on neural excitability provides a promising outlook on uncovering how listeners take advantage of the temporal dynamics in speech during comprehension.

3 Neural alignment to speech rhythm

In the previous chapters, we explained the value of metrical and predictable stress in the process of speech perception. Metrical stress can be used as an attentional guide through the speech stream. Indeed, we presented metrically stressed syllables as perceptually stable, cues to words onset, and 'attention-grabbing' both through their acoustic prominence and through their predictability.

In the current chapter, we will discuss how the models presented previously (i.e. MSS, ABH and DAT) are in agreement with recent developments on speech perception from the field of neuroscience. According to these new theoretical frameworks, neural excitability aligns to the rhythmic and metrical structure in speech, such that processing is optimized at the crucial time points in speech and comprehension is facilitated.

3.1 Neural excitability and attentional sampling

The excitability states of neuronal ensembles have been found to fluctuate rhythmically in what is called a **neural** or **brain oscillation**. This means that the phase of an oscillation influences the probability with which a neuron will fire. More specifically, when an event occurs during the extracellular troughs (i.e. intracellular peaks) of the oscillation, there will be a greater neural response than when the event occurred during the extracellular peaks (i.e. intracellular troughs) (Buzsáki, 2009).

delta (δ) $\rightarrow \sim 0-4$ Hz
($> \sim 250$ ms)

theta (θ) $\rightarrow \sim 4-8$ Hz
($\sim 125-250$ ms)

alpha (α) $\rightarrow \sim 8-12$ Hz
($\sim 83-125$ ms)

beta (β) $\rightarrow \sim 12-30$ Hz
($\sim 33-83$ ms)

gamma (γ) $\rightarrow \sim 30-$ Hz
($< \sim 33$ ms)

Table 3.1: The temporal excitation of neurons is correlated with fluctuations of oscillatory EEG activity in different frequency bands: delta, theta, alpha, beta and gamma.

Consequently, the timing of an event (and the phase in the oscillation at that moment) can influence its subsequent processing (e.g. Luck, 2005; Rohenkohl & Nobre, 2011; Rohenkohl et al., 2012; VanRullen, 2016). That is, when an event occurs during high excitability, it will be better analyzed than when it occurs during states of low excitability. This means that the phase of intrinsic, endogenous neural oscillations predicts subsequent perceptual performance (e.g. VanRullen et al., 2011; Rohenkohl et al., 2012; Cravo et al., 2013; VanRullen, 2016). In fact, in a series of studies on visual perception, the phase of the oscillation was found the *determining* factor on whether a stimulus presented at or near the perceptual threshold was perceived *at all* (e.g. VanRullen et al., 2011, see VanRullen 2016 for a review). This led VanRullen to propose that perception is discrete as opposed to continuous and that neural oscillations are behind the sensory mechanism of attentional (and temporal) sampling (see also VanRullen & Koch, 2003).

Note, however, that, where in the studies presented above, the detection of near-threshold visual stimuli depended on their (random) timing within an *intrinsic* (alpha) cycle (i.e. an endogenous oscillation which was not set in motion through prediction), 'random' pre-stimulus oscillatory phase is less likely to have an effect on auditory perception in general, and in particular on the perception of speech (Zoefel & Heil, 2013, see also VanRullen et al. 2014; Zoefel & VanRullen 2017). This is because, where visual information is presented spatially, auditory information presents itself temporally, requiring a more intelligent and dynamic processing mechanism, which tracks the input over multiple frequencies, based on predictions derived from rhythmic regularities.

According to theoretical frameworks of predictive coding (such as DAT), such a mechanism involves the endogenous oscillations to phase-align or synchronize with external rhythmic (and, thus, predictable) events. This mechanism is also called neural entrainment. Indeed, aligning the high excitability phase of oscillations to the most relevant temporal windows of external sensory information, ensures an optimized processing of the input (Schroeder & Lakatos, 2009; Herbst & Landau, 2016). This means that, if neural excitability can peak on the time points most relevant for the task at hand, this should lead to a behavioral advantage. Crucially, such a mechanism relies heavily on the ability to predict when a relevant event is going to occur and to focus attention and neural resources accordingly. That is, the input must display some temporal regularities that the mechanism can use to derive predictions. There is ample evidence for temporal sampling across different sensory domains (for example, motor, e.g. Morillon et al. 2014; visual, e.g. Rohenkohl & Nobre 2011; auditory, e.g. Henry & Obleser 2012; Schmidt-Kassow et al. 2009, 2013; Ten Oever et al. 2015, 2017, cross-sensory e.g. Arnal et al. 2009). In these studies, a rhythmic (isochronic) presentation of external stimuli phase-reset the oscillations in the corresponding frequency band,[1] which, in turn, led to a behavioral advantage.

Note, however, that in the studies cited above, the regularity or rhythm at which stimuli were presented was simple, isochronic and one-dimensional. Still, there is also evidence for neural tracking of more complex rhythms (e.g. Jones, 1976; Large & Jones, 1999; Jones et al., 2002; Henry et al., 2014). In Henry et al. (2014), listeners were asked to detect near threshold gaps in auditory stimuli that was modulated along two frequency bands. Detection was optimal when the troughs of the two neural frequency bands coincided. This provides strong evidence for DAT and suggests that the degree of attentive engagement with auditory input fluctuates rhythmically and over multiple temporal scales. This indicates that, when a task or stimulus presents a hierarchy of temporal regularities, multiple (nested) oscillations can guide attention towards the task relevant time points.

Speech is inherently rhythmic. The rhythms in speech are hierarchical in nature, with different layers in the prosodic hierarchy conveying information along multiple time-scales (e.g. Rosen, 1992). Speech presents slow temporal fluctuations in the $0-3$ ($\sim>$ 333 ms) and $4-8$ ($\sim 125-250$ ms) frequency range that represent syllabic and prosodic properties. As we have seen in the previous chapters, syllabic and prosodic cues are essential in speech comprehension, however, it is only recently, that it has been suggested that ongoing neural oscillations take advantage of this characteristic by adjusting their phase to match the rhythmic structure in speech (Poeppel et al., 2008; Ghitza, 2011; Peelle & Davis, 2012; Giraud & Poeppel, 2012; Ding & Simon, 2014; Ghitza, 2016).

---

[1] Note, that is not clear whether there was a phase-reset of the endogenous oscillation or whether the oscillation was set in motion, or induced, by the rhythmic presentation of the stimuli.

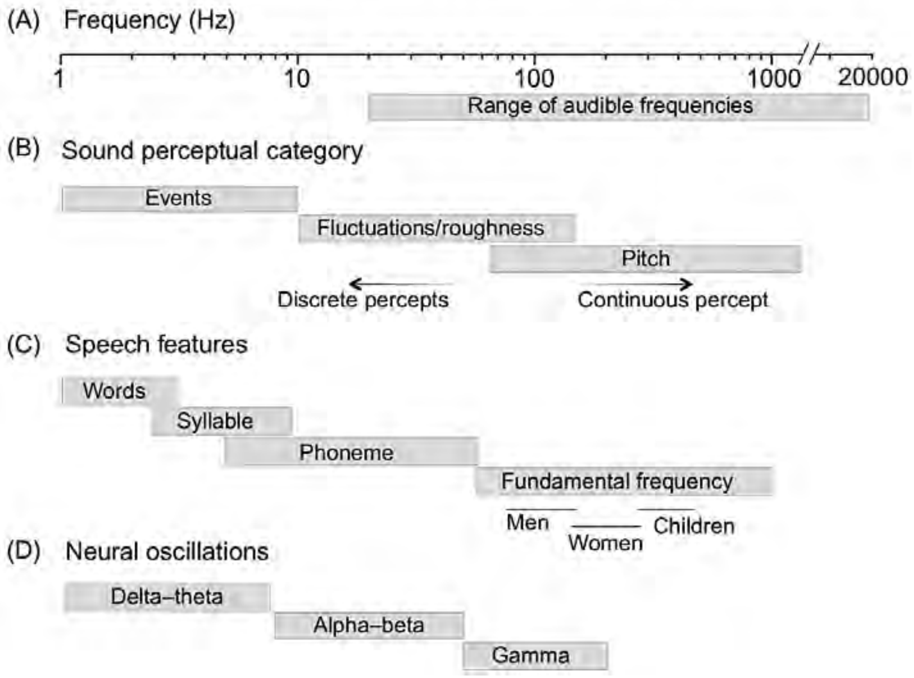


Figure 3.1: (A) Scale of perceived temporal modulation. (B) Relevant psychophysical parameters (perceptual changes) of the spectrogram reflect the temporal constraints that superimpose on the structure of linguistic signals. (C) Temporal structure of linguistic features. (D) The length of linguistic features remarkably matches the frequency of oscillations that are observed at rest in the brain. Figure taken from Hickok & Small (2015).

Modulation of neural activity within specific frequency bands (see table 3.1) has been linked to tracking linguistic information delivered at corresponding timescales: prosodic cues are tracked by low delta oscillators (0.5—2 Hz; e.g. Power et al., 2013; Bourguignon et al., 2013; Kayser et al., 2015; Ding & He, 2016; Molinaro et al., 2016; Goswami et al., 2016; Ghitza, 2016; Goswami, 2017), syllable structure is tracked by theta oscillators (4—8 Hz; e.g. Ding & Simon, 2012a,b, 2013; Doelling et al., 2014; Chait et al., 2015; Zoefel et al., 2015; Ding et al., 2016a), and phonetic features or formant transitions are tracked by higher beta/gamma frequencies ($\sim$ 25—35 Hz; e.g. Rosen, 1992; Gross et al., 2013; Martin, 2016).

Particularly in the analysis of speech, such a mechanism can serve many purposes. Clearly, it ensures high neural sensitivity at the most relevant parts in the speech signal, but it can also simultaneously process and segment speech at multiple time-scales (e.g. Poeppel et al., 2008; Giraud & Poeppel, 2012), as such contributing to the solution for the segmentation problem discussed in the previous chapter. Furthermore, nesting amongst neural oscillators of different frequencies can have a functional role in the temporal integration of the speech events (e.g. Sohoglu et al., 2012; Giraud & Poeppel, 2012; Chait et al., 2015; Mai et al., 2016; Cope et al., 2017). Finally, nesting has been linked to the interplay between phonological processing (in delta and theta) and other cognitive processes such as memory processing or motor (in beta) and lexical retrieval (in gamma) (see Mai et al., 2016, for an elaborate discussion).

The primary criticism to entrainment to speech is that the interval durations between speech events are not strictly isochronic. Speech is only *quasi*-periodic.[2] It is however possible to have oscillators track the different metrical layers in rhythm based on prediction, while being reset by salient acoustic events (e.g. Large, 2008; Large & Jones, 1999; Giraud & Poeppel, 2012). In fact, DAT distinguishes between two types of attention, one that serves 'future oriented attending' and one that serves 'analytic attending' (Jones & Boltz, 1989; Large & Jones, 1999; Henry et al., 2014). In future oriented attending, temporal predictability drives oscillators to guide attention towards the points anticipated to be most relevant, while in analytic attending the acoustic input is monitored rather closely. If the oscillatory mechanism involves resetting the periods of oscillators in order for them to time-lock to inter-event intervals in the speech signal, then the system does not depend on strict periodicity and allows for inter-event intervals of (slight) different durations (see also Turk & Shattuck-Hufnagel, 2013).

Of course these two types of attending do not need to represent two distinct processes and may be somewhere along a continuum, possibly additionally reflecting the relative dominance between oscillators/prosodic domains (e.g. Henry et al., 2014). For instance, high frequency oscillations (i.e. gamma) have been shown to phase shift as a function of a spoken word, while the lower frequency oscillations (delta and theta) do not always match the temporal modulation in the acoustic signal, but represent top-down attentional modulations instead (e.g. Kösem et al., 2016, see also Schroeder & Lakatos 2009; Zion Golumbic et al. 2013). That is, delta oscillations have been associated with attentional and predictive modulations during the processing of auditory information (Schroeder & Lakatos, 2009; Zion Golumbic et al., 2013) as they can be elicited even in the absence of acoustic cues delimiting speech units signal (Ding et al., 2016a).

Moreover, as was argued in chapter 1, the timing of speech events does not solely depend on phonological rules within one prosodic domain. Instead, the timing of events is determined by the united constraints imposed by each layer along the entire prosodic hierarchy. Recall, that the higher layers in the prosodic hierarchy (e.g. the phrase or AP) co-determine the temporal location of the events in lower layers of the hierarchy, such as the stressed syllables (Cummins & Port, 1998; Port, 2003, see also).

---

[2] Syllables, for example, vary in duration by 150 ms or more, and these variations occur continuously through ongoing speech, while, likewise, interstress intervals can vary between $300-700$ ms (e.g. Fant et al., 1991; Hickok & Poeppel, 2015; Ding et al., 2016b).

In chapter 1, we saw that work from Cummins and Port, for instance, showed that stressed syllables within a phrase were not placed at random locations but were biased towards the harmonic of the phrase period (Cummins & Port, 1998; Port, 2003; Tajima & Port, 2003). This suggests that speech is temporally organized according to hierarchical principles wherein the lower levels are nested under the higher levels with integer relations, i.e. the window of the phrase constrains the possible locations for stress. Similarly, the Principle of Rhythmic Alternation, wherein rhythm plays a role in the phonological processes which determine the distribution and alternation of strong and weak syllables, is constrained within the phrasal domain (e.g. Fox, 2000; Frota, 2012; van der Hulst, 2014). These top-down constraints mean that even if stress intervals are not strictly isochronic, they may still be temporally predictable through the longer window that determines their presentation in time.

Indeed, as is suggested in Ghitza & Greenberg (2009), auditory perception, and entrainment to speech, may be less dependent on strict isochrony and more on inherent temporal constraints on the interval durations between speech events (see also Greenberg et al., 2003; Greenberg & Arai, 2004; Ghitza et al., 2012; Ghitza, 2013; Doelling et al., 2014). That is, besides the constraints on the timing of stressed syllables imposed by the phrasal domain, Ghitza & Greenberg (2009) demonstrated that phoneme perception depends on the duration of the syllable (see also Giraud & Poeppel, 2012; Hyafil et al., 2015). In their study, the authors found that while compressing syllables rendered speech unintelligible, inserting periods of silence to bring the syllable back to its normal duration, restored intelligibility. These results led the authors to propose that speech (or phoneme) perception is not a continuous process, but critically depends on the delivery of discrete samples (or "packets"; Ghitza, 2014) within—in this case—the theta range. That is, they suggest that there may be biological limitations in the speech system which mandate phoneme-clusters to be presented in time-windows that leave sufficient time for the decoding of the segmental information.

So, even though speech is not strictly periodic, it seems that it is—both in production (Cummins & Port, 1998; Port, 2003) and in perception (Greenberg et al., 2003; Ghitza, 2013; Giraud & Poeppel, 2012)—regulated through coupled oscillators that impose constraints on the time windows of speech events to allow for prediction and optimal processing.

If there is a hierarchy among the oscillators tracking the speech stream, then the most dominant oscillator would arguably be the one underlying the metrical beat of the speech rhythm. Indeed, as we have seen countless times before throughout this text, there is a certain privileged processing of speech events occurring in beat position (e.g. Cutler, 1976; Jones, 1976; Pitt & Samuel, 1990; Large & Jones, 1999; Jones et al., 2002; Quené & Port, 2005; Schmidt-Kassow & Kotz, 2008; Schmidt-Kassow et al., 2009; Rothermich et al., 2010, 2012; Cason & Schön, 2012; Falk & Dalla Bella, 2016; Harding, 2016). Given the status of prosody in general, and accentuation in particular, in controlling the perception of complex rhythms such as speech and providing the temporal context in which speech segments are organized, it can be expected that more attentional effort is assigned to the beat positions. This proposition is in line with the theoretical frameworks discussed in section 1.4, suggesting the interplay between prosody and attention (and their combined effect on neural excitability) to be at the heart of an efficient and biologically plausible speech analyzing mechanism.

Most work on neural entrainment (as the mechanism behind the modulations of neural excitability) to speech has, however, concentrated on the temporal limits in processing the levels lower in the prosodic hierarchy. That is, work has predominantly investigated syllabic encoding in the theta band ($4-8$ Hz) and phonemic encoding in the low-gamma band ($25-35$ Hz). The role of entrainment in processing higher levels of linguistic abstraction corresponding to longer time windows (e.g. stress patterns, lexical words, prosodic phrases) has remained relatively understudied (although see for instance Power et al., 2013; Bourguignon et al., 2013; Kayser et al., 2015; Goswami et al., 2016; Meyer et al., 2017, for notable exceptions).

Indeed, even though languages' metrical structures have long been recognized to play a vital role in speech comprehension, with prosodically organized syllables demarcating increasingly larger prominence structures, thereby lending language much of its predictable properties, neural alignment to the temporal regularities conveyed by the prosodic hierarchical structure has yet to be empirically tested. Instead, most evidence for the perceptual benefit of a rhythmic beat during speech comprehension, has used the technique of Event-Related Potentials (ERP), presented in the next section.

ERPs are averaged time-locked brain signals, elicited by external input. They are different from entrained neural oscillations, but they may be related. Where entrained neural oscillations represent the activity of synchronized neuronal ensembles that are intrinsically coupled and coupled to a common, external and regular event, ERPs are responses to the presentation of an external stimulus resulting in an increase of neural activity often modulated by prediction. Note that this means that in many of the studies either discussed or cited above, it is not clear whether the results reflect actual neural entrainment[3] or the superposition of a train of evoked-related potentials.

Indeed it has frequently been argued that the observed results, interpreted as entrainment, in reality are a series of evoked potentials elicited by the rhythmic presentation of the stimuli (see e.g. Haegens & Zion Golumbic, 2017, for an excellent critical review). The series of evoked potentials, also called steady-state evoked potential (SSEP), surface as increased neural power at the frequency with which the stimuli were presented. This critical observation does not negate the possibility of neural entrainment as mechanism underlying optimal coding of rhythmic events, but justly points out that the results can also be explained through 'simpler' superpositions of event-related potentials.[4]

Note, however, that the behavioral benefit or enhanced sensitivity observed for events occurring in time-periods predicted by pulse or meter, and in fact rhythmic Gestalt itself, do point to rhythmically enhanced top-down modulation of neural excitability, considering pulse and meter are not directly present acoustically. Recall from section 1.3 that pulse and meter are percepts and not part of the auditory signal itself (Large, 2008). Additionally, the perceptual illusion of a salient, metrical accent, even when its phonetic correlates are suppressed—such as with the French initial accent (IA) in the study of Jankowski et al. (1999)—similarly suggests increased neural sensitivity during this time window, possibly modulated by oscillations although top-down prediction also explains the phonological percept.

REGARDLESS, clearly the evidence reported above demonstrate a perceptual advantage for rhythmic events, which likely results from the interplay between the bottom-up acoustic salience and top-down guided attention, two mechanisms which are not mutually exclusive. In speech, the perceptual advantage should be greatest at the metrical rate which underlies its rhythm, but the oscillatory involvement has, for as far as we know, not yet been empirically tested (see also Beier & Ferreira, 2018). Most work on the facilitatory effect of beat on speech perception is provided by studies using event-related potentials, a method which, indeed, is well equipped to demonstrate the interplay between bottom-up properties and top-down attentional processes.

---

[3] Endogenous neural oscillations (i.e. oscillations in the absence of external stimuli) being modulated (either phase-reset or otherwise changed in their oscillatory magnitude) by external stimulation.

[4] Note that, conversely, modulations in ERP amplitudes could reflect underlying neural entrainment.

3.2 Event-Related Potentials and predictive coding

In the current section, we will focus on the method of ERPs (e.g. Luck, 2005; Duncan et al., 2009), specifically in the investigation of language processing. ERPs generally reflect the mismatch between sensory input and expected input, and, as such, are tightly linked to the theoretical framework of predictive coding. As presented in section 2.2.1, central in predictive coding is that perception is not only based on a passive and bottom-up analysis of sensory input, but additionally involves a comparison with top-down derived expectations. That is, listeners are held to continuously generate predictions about abstract anticipated properties of the upcoming speech signal and violations of these predictions are evident in early cortical processing, i.e. mismatches between the sensory information and the top-down prediction result in so-called prediction errors which require additional processing. It is then this additional effort that is reflected or measured in the modulation of ERP components.

CRUCIALLY, this interpretation of ERPs as it pertains to language processing, underlines that speech perception constantly relies on online prediction (e.g. Tavano & Scharinger, 2015), generating prior assumption based on context and/or long-term memory.

The method of ERP has proven a valuable tool in the study of metrical stress processing during speech comprehension and presents several advantages over more traditional methods such as behavioral measures. For instance, ERPs provide a continuous measure of cognitive processing between the stimulus and the neural response (e.g. Luck, 2005). Because ERP components are measured online and have a high temporal resolution, it is possible to infer which processing stage was affected by the experimental manipulation. But also, ERPs are more sensitive than behavioral measures and help to detect difficulties in processes as subtle and automatic as metrical processing. Such high sensitivity is especially useful when not the *legality* but the ***probability*** of the stimuli are manipulated, such as in the current work.[5] Moreover, because the online measure does not require behavioral responses, attention

---

[5] Recall that in the current work, we seek to determine to what extend presenting words without their hypothesized bipolar stress template disrupts word level processing during speech comprehension, by phonetically reducing IA or FA, which, indeed, does not create a stress violation in French.

can be directed away from the metrical manipulation, and processing costs or facilitatory effects can be monitored covertly. That is, ERPs are especially sensitive to expectancy violations, and a mismatch between the expectation and the linguistic input automatically results in a larger or later ERP.

Furthermore, components may differ in their timing and location depending on the type of expectation violation, which allows for the detection of expectancy advantages that are crossed with other linguistic processes (e.g. Kutas & Hillyard, 1980; Luck, 2005). That is, the method of ERP can be used to show that the brain detects violations of expected stress patterns in speech, together with the subsequent hindrance on specific linguistic processes such as lexical access, semantic retrieval or syntactic analysis (e.g. Böcker et al., 1999; Magne et al., 2007; Domahs et al., 2008; Schmidt-Kassow & Kotz, 2008; Rothermich et al., 2010, 2012; Bohn et al., 2013; Harding, 2016).

Below, we will review four components relevant in the current work: the MMN, PMN, N325 and N400 (e.g. Näätänen et al., 2007; Bentin et al., 1985; Connolly & Phillips, 1994; Böcker et al., 1999; Steinhauer & Connolly, 2008; Brown & Hagoort, 1993; Kutas & Federmeier, 2011). These four components all have in common that their modulation reflects a mismatch between an expectation based on long-term memory representations or established phonological/linguistic representations and the violation in the experimental setting and, as such, they each reflect the outcome of predictive coding (Friston, 2005) and allow for inferences on the time course (i.e. processing stage) and anticipated phonological representations (with which the input mismatches) during speech comprehension (see also Scharinger et al., 2016).

3.2.1 MisMatch Negativity (MMN)

A prime example of the ERP-component as the outcome of prediction mismatching input, is the Mismatch Negativity component (MMN; e.g. Näätänen et al., 2007, see also Garrido et al. 2009; Winkler et al. 2009; Denham & Winkler 2017). The MMN is a pre-attentive, fronto-central negative deflection peaking around 250 ms after the detection of a regularity violation. The amplitude of the MMN typically reflects the magnitude of the deviance from what was expected. Such deviance can be purely acoustic (bottom-up) or it can be a deviance from a top-down derived prediction which is based on long-term memory representations (e.g. Winkler et al., 2009; Garrido et al., 2009). In the latter case, the MMN can thus index the strength of memory traces.

MMNs are typically investigated in an oddball paradigm wherein a low-probability stimulus (the oddball, or deviant) occurs within a train of high-probability stimuli (Näätänen et al., 2007). The frequently occurring standard stimuli are assumed to develop predictions that are subsequently violated by the infrequently occurring deviant stimulus. The standard and deviant stimuli will usually be very similar acoustically, contrasting only on the phonological property of interest in the investigation (e.g. phoneme or stress pattern). The MMN is then obtained by subtracting the ERP elicited by the standard from the ERP elicited by the deviant. Therefore, the MMN represents the difference between the neural response to the frequently occurring standard stimulus and the infrequently occurring deviant stimulus, i.e. the MMN is then a 'difference wave' that reflects the status of the manipulated phonological feature.

Importantly, whereas an MMN may be elicited by a purely acoustic difference, many studies will additionally switch the position of the deviant stimulus and the standard stimulus, such that they have another condition, wherein the deviant is presented frequently, while the (formerly) standard stimulus is presented infrequently (e.g. Honbolygó & Csépe, 2013; Astésano et al., 2013; Aguilera et al., 2014; Scharinger et al., 2016). If the standard and deviant stimuli differ only acoustically, the MMNs in both conditions should be similar. Often, however, MMN amplitudes will differ, presumably due to a more established representation for one type of stimulus over the other. That is, repeatedly presenting a stimulus with a firm phonological representation, only builds on its probability leading to a large mismatch response when its anticipation is violated. In the reverse situation, when a train of improbable standards is interrupted by a more probable deviant, the violation, and thus the mismatch response, is much smaller.

The MMN has proved a valuable tool in investigations of underspecification of phonemic representations (e.g. Eulitz & Lahiri, 2004; Näätänen et al., 2007; Winkler et al., 2009; Deguchi et al., 2010; Ylinen et al., 2016; Scharinger et al., 2016, 2017), as well as the phonological representation of stress patterns (e.g. Ylinen et al., 2009; Honbolygó et al., 2004; Honbolygó & Csépe, 2013; Astésano et al., 2013; Aguilera et al., 2014; Honbolygó et al., 2017; Garami et al., 2017). For instance, Honbolygó et al. (2004) investigated processing difficulties of stress patterns in Hungarian participants. The standard in their oddball study was a disyllabic word with trochaic stress, the typical stress pattern in Hungarian, while the deviant carried an iambic stress pattern. The deviant elicited two different MMNs: one in response to the lack of the typical and expected stress on the first syllable, and another to the atypical additional stress on the second syllable. In a follow-up study, the trochaic and iambic stress pattern served both as standards and deviants in two separate blocks (Honbolygó & Csépe, 2013). Again, the results indicated that the deviant with an iambic stress pattern elicited two consecutive MMNs, however, when the trochaic patterns had been the deviant, no MMN followed. The authors argued that the unfamiliar iambic stress pattern mismatched both the short and long-term memory representations, and, therefore, elicited the MMNs, while the typical (and thus expected) trochaic stress pattern did not elicit any MMN because it did not mismatch the long-term memory representation of word stress in Hungarian. These findings provide evidence that processing of stress pattern changes relies on language-specific long-term memory representations (see also Ylinen et al., 2009, for similar results).

In a study addressing the phonological status of the French initial accent (IA), Aguilera et al. (2014) showed that IA is not only perceived (recall that French listeners are allegedly 'deaf to stress'), but anticipated by listeners as belonging to the abstract representation of the word (see also Astésano et al., 2013). The authors manipulated the phonetic realization of IA on trisyllabic words in an oddball paradigm. When the oddball had been presented without IA, a clear MMN emerged, which was however significantly smaller, when the oddball was presented with IA, suggesting a long-term representation of the accent and underlining the preference and expectation for stress templates with IA. We will return to this study in section 6.1 where it will be discussed in more detail and where we will additionally present the results of two follow-up studies in which the oddball paradigm was used to examine the phonological representation of the French final accent.

3.2.2 Phonological Mapping Negativity (PMN) and N325

The Phonological Mapping Negativity (PMN) is a fronto-central negativity peaking between 250 and 350 ms post-stimulus onset (Connolly & Phillips, 1994; Newman & Connolly, 2009). The PMN is held to reflect the pre-lexical processing cost of expectation violating phonemic information. It is similar to the N400, presented below, in the sense that it is context dependent (Newman & Connolly, 2009; Steinhauer & Connolly, 2008; Kujala et al., 2004), typically being most prominent on the last word of a sentence or list. PMN is however argued to precede the semantic processing response (i.e. N400) to sentence-ending words that are primed by the context. For instance, when PMN and N400 responses were recorded to sentence ending words whose phonological and semantic features were manipulated, the results showed that words that were semantically incongruent, but had the expected initial phoneme, elicited a N400, whereas semantically congruent words starting with an unexpected initial phoneme, elicited only a PMN (Connolly & Phillips, 1994).

A component related to the PMN, but held to more specifically reflect difficulties in metrical processing, is the N325 (Böcker et al., 1999). This component was first observed in a study by Böcker and colleagues wherein Dutch speakers either passively listened to, or activily discriminated between, a series of bisyllabic words marked with either the Dutch dominant trochaic stress template (88% of Dutch words are trochaically stressed) or the less frequent iambic stress template (only 12% of the words are iambically stressed) (Böcker et al., 1999). That is, in the study, lists of four words were presented ending in a word that either matched in stress template with the preceding three words or not. The infrequent, and therefore less expected, iambic template elicited a larger negativity at the frontal sites and around 325 ms post-stimulus onset: the N325. This negativity was interpreted to reflect the ease with which the stress template was extracted from the acoustic signal, presumably due to the more established (and, consequently, expected) phonological representation of the Dutch dominant stress pattern.

Crucially, the component was argued to be different than either PMN or N400, because it was enhanced in the active metrical discrimination task compared to the task wherein participants listened passively. Moreover, even though the N325 was enhanced in incongruent as opposed to congruent conditions, Böcker et al. (1999) obtained a more ample N325 to iambic than to trochaic stress templates regardless of whether the template was congruent in the list and regardless of whether the word was presented list final. That is, the difference in N325 to iambic compared to trochaic stress templates was independent of context and therefore held to more specifically reflect listeners' general metrical preferences.

We will return to this study in section 6.2 where we will additionally present the results of two lexical decision studies in which we examined the phonological and metrical preference for both the French initial accent and final accent and there respective interactions with the different stages in speech processing.

3.2.3 N400

The N400 is a centro-parietally located negativity that peaks around 400 ms after the detection of a semantic discrepancy. It is often elicited in paradigms such as the semantic priming paradigm (wherein a target-word directly follows a word or image to which it is semantically related or not), or the semantic anomaly paradigm (wherein sentences are presented with a target-word that is semantically congruent or incongruent within the sentence context).

The negativity is considered an adept indicator of obstructed speech comprehension, with amplitude modulations or delayed latencies revealing difficulties in speech processing. Still, the precise nature of the N400 remains a topic of considerable debate. That is, it is unclear, when observing N400 modulations, precisely which stage in speech processing was affected, and whether modulations are restricted to semantic information or whether the N400 can additionally be modulated by mismatching phonological information, such as metrical patterns.

One commonly held belief on the nature of the N400, is that it results from hindered contextual integration (van den Brink et al., 2001; Brown & Hagoort, 1993; Boulenger et al., 2011). In this view, the N400 indicates difficulties in the post-lexical stage of speech comprehension, i.e. the stage after initial pre-lexical activation and lexical access have been completed (see section 2.1), and is unlikely to be influenced by phonological processes. Another stance, however, considers the N400 to reflect the degree of lexical pre-activation. In this view, higher levels of pre-activation (as a results of, for instance, supporting prior semantic information or word frequency) facilitate lexical access and reduce N400 amplitude (Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; DeLong et al., 2005). This stance then takes the N400 to reflect predictive, anticipatory processes that need not exclusively be of semantic nature, but can be phonological as well (DeLong et al., 2005; Lau et al., 2008).[6] Indeed, a number of studies have shown misguided phonological expectations in healthy subjects (e.g. Praamstra & Stegeman, 1993; Dumay et al., 2001; DeLong et al., 2005) or impaired phonological analysis in patients (Robson et al., 2017) to interfere with subsequent semantic evaluation and modulate the N400.

Metrical information has also been found to interplay with lexico-semantic processing (e.g. Magne et al., 2007; Rothermich et al., 2010; Marie et al., 2011; Rothermich et al., 2012; Bohn et al., 2013). In a series of studies, Rothermich and colleagues manipulated the metrical regularity in German jabberwocky or semantically anomalous sentences (Rothermich et al., 2010, 2012; Rothermich & Kotz, 2013). In the studies, the words preceding the target word were either all trochaically stressed or iambically, leading to a clear expectation of where to expect stress on the target word.

Listeners were shown to direct their attention to syllables that were expected to be stressed, i.e. metrical regularity facilitated semantic ambiguity resolution, as indicated by an amplitude modulated and earlier N400. These results are related to the Dynamic Attending Theory and to the Attentional Bounce Hypothesis by the authors, and demonstrate that presenting speech with a regular underlying beat, allowed listeners to a priori direct their attention from one stressed syllable to the next (in their words) “island of reliability”, which in turn facilitated semantic processing.

Finally, in a previous ERP study investigating the relationship between metrical structure and late speech processing in French, metrical violations were found to obstruct semantic processing (Astésano et al., 2004; Magne et al., 2007). Recall that, in the study, participants listened to sentences in which semantic and/or metrical congruity was manipulated. Semantic congruity was manipulated by presenting sentences in which the last word was incoherent with the semantic context of the sentence, while metrical congruity was manipulated by lengthening the medial syllable of the last word, an illegal stress pattern in French. The metrical violation resulted in an increased N400, even when the sentences were semantically congruent. This indicates that phonological information and metrical expectancies can also modulate the (typically semantic) N400.

We will return to these studies in section 6.3, where we will additionally

---

[6] Note that while the post-lexical integration theory *may* reject anticipatory processes and consider the N400 to index exclusively post-lexical processes initiated upon perceiving the target word, it not necessarily *needs* to; one can easily imagine integration processes to also benefit from successful (semantic) anticipation based on prior contextual information (as is pointed out by Yan et al. 2017, see also Kuperberg & Jaeger 2016 and Nieuwland et al. 2018).

present the results of an N400 study wherein we examined the interplay between the French initial accent and lexico-semantic processing.

3.3 Chapter summary

In the current chapter, we described the relationship between accentuation and attention, particularly in how it pertains to neural excitability. It was explained in what way the Attentional Bounce Hypothesis (ABH; Pitt & Samuel, 1990) and Dynamic Attending Theory (DAT; Large & Jones, 1999) envision accentuation to guide neural excitability such that speech comprehension is facilitated. Perceiving regularities has been found beneficial for understanding acoustic information because it induces temporal expectations for upcoming events which are subsequently preferentially processed. In other words, assuming the attentional theories, we presented how metrical stress is expected to modulate neural excitability and benefit speech processing. We showed that DAT and ABH provide biologically plausible theories in their account on the perceptual advantage of metrical structure in speech comprehension.

In section 3.1, we saw that the excitability states of neuronal ensembles have been found to fluctuate rhythmically in what is called a neural or brain oscillation. This means that the phase of an oscillation influences the probability with which a neuron will fire (Buzsáki, 2009). That is, although time is typically perceived as continuous, studies in many cognitive domains (e.g. visual cognition, auditory cognition) show that the neural system chunks time; sampling, integrating and analyzing perceptual information in discontinuous time windows (VanRullen & Koch, 2003; Ghitza & Greenberg, 2009; VanRullen, 2016). Recent studies have discovered that when auditory information displays temporal regularities, it is processed through a dynamic oscillatory mechanism, which tracks the input over multiple frequencies, based on predictions (e.g. Jones, 1976; Large & Jones, 1999; Jones et al., 2002; Henry et al., 2014). Such a dynamic oscillatory mechanism, and the consequential behavioral benefit, had been demonstrated across sensory domains (Engel et al., 2001; Arnal et al., 2009; Schmidt-Kassow et al., 2009; Jensen & Mazaheri, 2010; Rohenkohl & Nobre, 2011; Henry & Obleser, 2012; Schmidt-Kassow et al., 2013; Morillon et al., 2014; Ten Oever et al., 2015, 2017).

There is no real reason why the mechanism should not be employed during the analysis of speech—an inherently rhythmic sound.

According to the oscillation based functional model, intrinsic oscillations align to an incoming speech signal along different frequency bands (Poeppel et al., 2008; Ghitza, 2011; Peelle & Davis, 2012; Giraud & Poeppel, 2012; Ding & Simon, 2014; Ghitza, 2016). This interaction is thought to act as a mechanism of sampling and packaging of the input, such that oscillators are phased-locked to generate an hierarchic organization of temporal windows. Particularly in the analysis of speech, a computationally challenging process, such a mechanism can serve many purposes. The mechanism ensures high neural sensitivity at the time points most important for speech comprehension and, additionally, presents a possibly universal account on how listeners simultaneously chunk speech over different timescales. Indeed, metrical structures are known to be essential to speech perception, and the oscillatory account is compatible with this view.

Most work on neural alignment to speech has, however, concentrated on the temporal limits in processing the lower levels in the prosodic hierarchy, i.e. the gamma band, corresponding to phonemes, and the theta band corresponding to syllables. Higher levels of linguistic abstraction that correspond to longer time windows (stress, words, prosodic phrases) often fall 'outside the scope' of this work. Moreover, when there is evidence of delta alignment to speech, suggesting guided attentional and predictive modulations in speech processing, it is typically not related to prosodic structure, but to abstract online syntactic parsing (e.g. Ding et al., 2016a; Meyer et al., 2017). This is unfortunate, since prosody plays an important role in speech segmentation and comprehension processes. Prosodically organized syllables become part of larger structures of prominence networks and give language much of its predictable properties, for instance, by structuring an utterance so that prominent events lie at privileged phases of a higher-level prosodic unit (e.g. Cummins & Port, 1998).

Indeed, even though languages' metrical structures have long been recognized to be crucially involved in speech comprehension, most work on the perceptual benefit of metrical predictability has used the method of Event-Related Potentials ERP. Section 3.2 presented the method of event-related potentials (ERP), specifically as it pertains to the investigation of the role of rhythm and prediction during speech processing. ERPs are averaged time-locked brain signals, elicited by external input and typically reflect the mismatch between sensory input and anticipated input. Mismatches between the sensory information and the top-down prediction result in so-called prediction errors which require addition processing and is reflected or measured in the modulation of ERP components.

With their high temporal resolution, ERP offers a direct measure of the activation in the neural networks underlying speech comprehension. Moreover, ERP components offer the possibility to observe rather subtle effects of perceptual regularities and metrical stress on speech comprehension that may not be evident behaviorally. We presented four components: the MMN, PMN, N325 and N400 (e.g. Näätänen et al., 2007; Bentin et al., 1985; Connolly & Phillips, 1994; Böcker et al., 1999; Steinhauer & Connolly, 2008; Brown & Hagoort, 1993; Kutas & Federmeier, 2011), which are employed in the current work wherein we investigate the interplay between metrical expectation and different level of speech analysis.

THAT IS, we use the method of ERP to infer on the phonological representation of the French initial accent (IA) and final accent (FA), as well as on their real-time interactions with the stages in speech processing.

Part III

RESEARCH CONTRIBUTION

4 Research Questions and Hypotheses

As should be clear by now, the current dissertation sets out to determine whether there is metrical stress in French. More specifically, we take a functional approach to examine whether the French initial and final accents (i.e. IA and FA, respectively) are phonologically encoded in bipolar stress templates underlying the representation of the lexical word, and whether presenting words without their stress patterns hinders word-level processing.

Phonological representation of French metrical stress

In a first effort, we will attempt to establish both whether French listeners have a phonological deafness for their accentual patterns, and whether the French initial and final accents hold a phonological representation, by manipulating the accents in an oddball ERP study (section 6.1).

⇒ We hypothesize that French listeners will readily perceive the accentual manipulation and that if words are encoded with both accents underlying their phonological representation, then manipulating the presence of the accents in an oddball setting should result in asymmetrical MMN difference waves, such that presenting deviants without the accents results in a more ample MMN than presenting them with the accents (*cf.* Aguilera et al., 2014; Astésano et al., 2013).

Metrical stress in word recognition

Pre-lexical stress templates serve to access the mental lexicon, so if, as we assume, both the initial accent and final accent are phonologically encoded and attached to the representation of the word, then French accentuation provides two lexical entries. This suggests a functional role for the accents in word-level processing. Therefore, the second effort sets out to determine to what extent accentuation contributes to the process of word recognition in French. We manipulate the presence of both the initial accent and the final accent in two lexical decision studies, such that we may observe the interactions of both accents within the process of lexical access (section 6.2).

⇒ We hypothesize that if word recognition involves the pre-lexical extraction of metrical information, then presenting words without their expected stress templates, whether IA or FA, should hinder lexical access.

French stress in lexico-semantic processing

Because speech processing has been shown to unfold in a cascading manner, obstructed lexical access by presenting listeners with words without their expected stress patterns, should continue to additionally hinder the later post-access stages in speech comprehension. Previous perception studies have shown the initial accent to be a reliable marker of lexical structure, even more so than FA (Astésano et al., 2007, 2012; Garnier et al., 2016; Garnier, 2018), suggesting the initial accent to have an especially prominent role. For this reason, in this final effort, we investigate the interplay between the French initial accent and the later processing stages during which semantic access and contextual integration take place. We manipulate the presence of the initial accent orthogonally to semantic congruity in an ERP N400 study.

⇒ We hypothesize that if IA is linked to the phonological representation of words and cues lexical access, then presenting words without IA should not only hamper word recognition but continue to interact with later semantic processing and modulate the N400.

5 Experimental Strategy

In order to investigate the questions just presented, we manipulated the presence of the initial accent and final accent on trisyllabic words in ERP paradigms. In a selection of these studies, we additionally collected behavioral measures. Below, we will present how we manipulated the presence of the accents to obtain our stimuli, after which we will explain and motivate the choices we took to analyze both our behavioral and EEG data.

5.1 Stimuli creation

5.1.1 Corpus

We used two different corpuses for the studies presented in the current work. The first corpus was used for our oddball studies (presented in section 6.1) and lexical decision studies (presented in section 6.2), wherein words and pseudowords were presented in isolation. The corpus consisted of sentences spoken by a naïve female speaker of French. In the sentences, target words were placed in a single intonational phrase to increase the probability of clear IA and FA marking (as discussed in Astésano et al., 2007). Target words were extracted to create our stimuli. They were all trisyllabic lexical nouns, or trisyllabic pseudowords that were phonologically similar to the lexical nouns (all consonant-vowel (CV) structures) but had no lexical content in

French. An example of a sentence in the corpus is presented below, with the target word in **bold**:

> "Le principe de cette nouvelle émission, dit-elle, et sa **diffusion**, pourraient être très mal accueillis par le public"
>
> *(English: The concept of this new program, she said, and its* ***broadcast****, may not be well accepted by the audience)*

The second corpus was used in our N400 study (presented in section 6.3) wherein we crossed metrical pattern with semantic congruity. The corpus originally consisted of 512 experimental sentences that all ended with a trisyllabic noun and were spoken by a native male speaker of standard French and recorded in an anechoic chamber using a digital audiotape (sampling at 44.1 kHz) (see also Magne et al., 2007).[1] All sentences were spoken in a declarative mode, and the pitch contour was always falling at the end of the sentence. Among the 512 sentences, 256 ended with semantically congruent words (+S) and 256 ended with semantically incongruent words (−S).

Semantically incongruous sentences were built by replacing the final congruent word with a word that shared similar acoustic and phonological characteristics, but did not make sense in the sentence context. Moreover, semantically congruent and incongruent target words all had CV syllable structures and were matched for word frequency (92.38 and 91.36 occurrences per million, respectively), using the LEXIQUE2 French lexical database (New et al., 2001, in Magne et al. 2007). Examples of a sentence-pair in the corpus is presented below, with the target words **bold**:

> "La greffe de moelle soigne la **leucémie**" (+S)
>
> *(English: Bone marrow transplant treats* ***leukemia****)*

> "La greffe de moelle soigne la **densité**" (−S)
>
> *(English: Bone marrow transplant treats* ***density****)*

---

[1] We wish to thank Mireille Besson for allowing us to reuse this corpus.

5.1.2 Stimuli manipulation: Initial Accent

5.1.2.1 Stimuli selection

For the initial accent, stimuli selection was based on the presence of a marked and natural IA in the original corpus. Because the primary phonetic parameter of IA is a rise in f_0 (Astésano, 2001, see also table 1.3), this meant that only sentences in which the target words started with a rise of f_0 of at least 10% on the first syllable compared to the preceding f_0 value on the (unaccented) determinant (Astésano et al., 2007; Ladd, 2008) were admitted as stimuli.

Note that for our stimuli for the N400 study, the selection additionally necessitated that the natural IA should be similar across semantic conditions. That is, only sentences were admitted where there was a pronounced and similar IA both on the semantically congruent target word and on the semantically incongruent target word, as judged both on a acoustic analysis and on the judgment of two phonetic experts (e.g. 'leucémie' *and* 'densité', see above).

5.1.2.2 Sound manipulation

The metrical condition (±IA) was created by lowering the f_0 value on the first vowel of the target-words near the f_0 value on the preceding (unaccented) determinant in order to remove the natural +IA and create the −IA condition. This manipulation was achieved using a customized quadratic algorithm adapted from Aguilera et al. (2014) in PRAAT (Boersma & Weenink, 2016) which progressively modified the f_0 values while allowing for microprosodic variations to be maintained in order for the natural sound of the stimuli to remain intact.

That is, in a first step, pitch contours are calculated from the original sound files, i.e. the **original pitch contour**. Next, the **desired pitch contour** (i.e. target contour without IA) are created by adjusting the period between the midpoint of the first vowel and the beginning of the third vowel using a quadratic algorithm.

The transformation is defined by:

$$TargetF0(t) = SourceF0(t) * k(t)$$

Wherein,

- SourceF0: the original pitch contour
- TargetF0(t): the desired pitch contour
- $k(t)$: a time series of a multiplicative coefficient

$$\rightarrow (1 - kref)/(tdebV3 - tref)^2 * (t - tref)^2 + kref$$

Wherein,

- $kref$ = mean f_0 determinant vowel/mean f_0 first vowel $*1,05$
- $tref$: midpoint time of the first vowel
- $tdebV3$: onset time of the third vowel
- t starting at the beginning of the first vowel, ending at the beginning of third vowel

Importantly, the quadratic algorithm allows for maximum adjustment centered on the vowel of the first syllable (i.e. the syllable carrying IA), and ensures that, in the desired pitch contour, the f_0 at the midpoint of the first vowel is 5% higher than the mean f_0 of the determinant, which is lower that the perceptual threshold value given in Ladd (2008). The transformation of the source f_0 declines smoothly to a subtle effect at the end of the period (i.e. just before the third syllable) with the k coefficient changing quadratically between these two time points. The desired pitch contour is then fitted back on the original contour to keep the naturalness of the sound using the Overlap-Add method in PRAAT (Boersma & Weenink, 2016).

Finally, the +IA stimuli are forward and back transformed to equalize the speech quality between +IA and −IA stimuli. That is, because the quality of the speech signal needs to be similar between conditions, the PRAAT scripts were reversed such that the desired f_0 contour is fitted back to the original f_0 contour, resulting in our fourth pitch contour: **back contour**. A panel of three experts selected stimuli with the most natural IA (+IA) (see figure 5.1 for an example of the f_0 manipulation on the stimulus 'chibuté' for the lexical decision study presented in section 6.2.1).

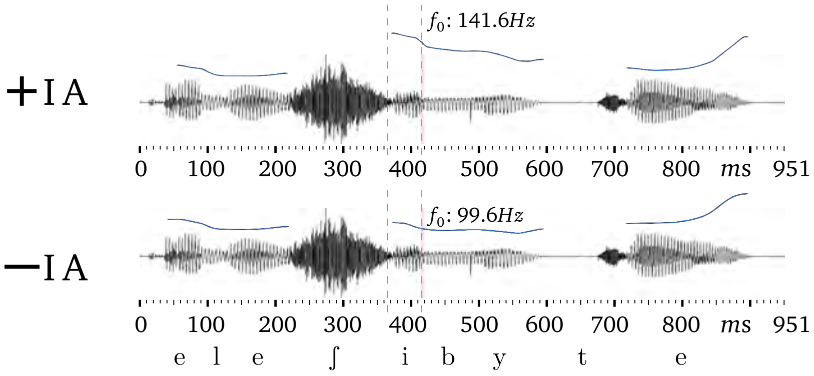


Figure 5.1: Example of f_0 resynthesis +IA (top) and −IA (bottom) of the stimulus eleʃibyte ('et les chibutés'), with quadratic interpolation from the f_0 value of the preceding determinant to the f_0 value at the beginning of the last stressed syllable for −IA targets.

Note that for our stimuli in the N400 study (presented in section 6.3), wherein the target words were embedded into a sentence, we slightly modified this manipulation.

$$k(t) = (1 - \text{kref})/(\text{tdebV3} - \text{ onset time determinant})^2 * (t - \text{ tref})^2 + \text{ kref}$$

First, we used different values to calculate $kref$, i.e. the ratio between mean f_0 of the determinant and the mean f_0 of the initial syllable times 1.05%. With PRAAT, we obtained the midpoint f_0 of the determinant and the maximal f_0 of the first syllable and used those values to determine $kref$.

$$kref = \text{midpoint } f_0 \text{ determinant vowel/max } f_0 \text{ first vowel} * 1,05$$

Second, we changed the period in which the f_0 is modulated (t). In order to avoid a sudden decrease at the midpoint of the first vowel, t now starts at the beginning of the determinant instead of at the beginning of the first syllable. To further have the modulation naturally decline up until the end of the target word t ends at the offset of the third syllable instead of at the third syllable onset.

$$k(t) = (1 - kref)/(tdebV3 - \text{onset time determinant})^2 * (t - tref)^2 + kref$$

The algorithm still centers on the midpoint of the first syllable, i.e. the effect on the f_0 is still greatest on the initial accent, and declines gradually both left and right of the initial accent (see figure 5.2 for an example of the f_0

resynthesis on matching semantically congruent and incongruent sentences for the N400 study presented in section 6.3).

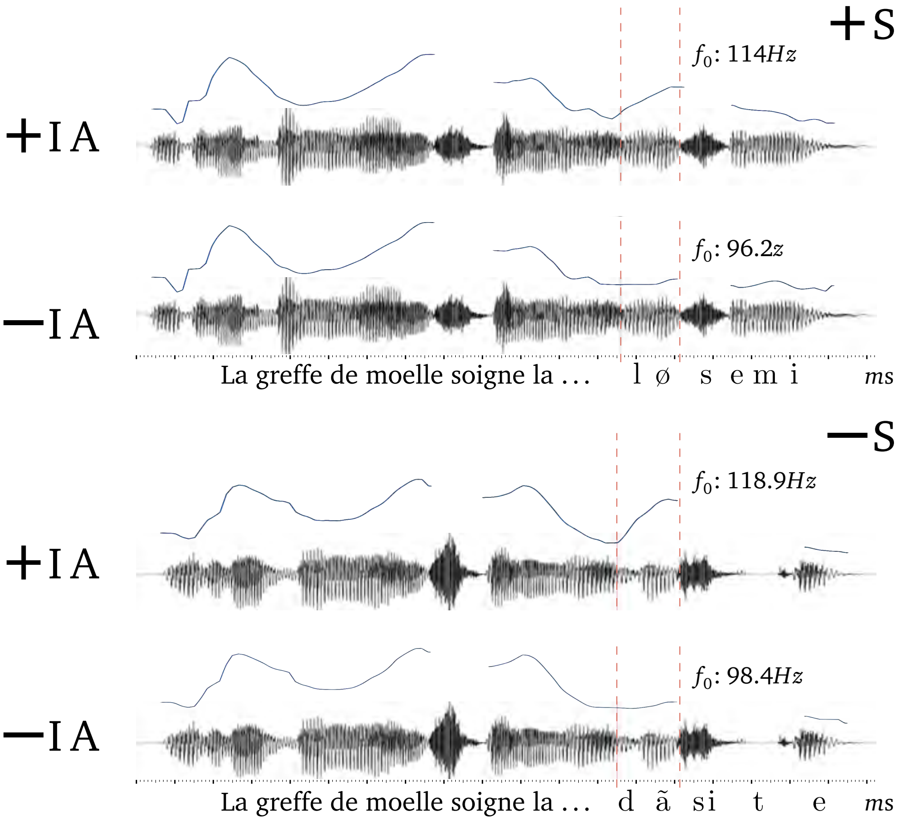


Figure 5.2: Example of f_0 resynthesis with (+IA) and without initial accent (−IA) on semantically incongruent (+S, top two) and semantically congruent (−S, bottom two) sentences with quadratic interpolation from the f_0 value of the preceding determinant to the f_0 value at the beginning of the last stressed syllable for +IA targets (visible in blue). The time window of ±IA is indicated by vertical red dashed lines.

5.1.3 Stimuli manipulation: Final Accent

5.1.3.1 Stimuli selection

Stimuli selection was based on the presence of a marked and natural FA in the original corpus. Because the primary phonetic parameter of FA is duration (Astésano, 2001, see also table 1.3), this meant that target words were only admitted in the current corpus if the third syllable was minimally 25% longer in duration than the preceding unaccented medial syllable (Astésano, 2001). In order to further optimize the sound manipulation, target words wherein the third syllable

- displayed a marked f_0 rise
- contained a consonant or schwa at their offset
- had a long consonant at the syllabic onset relative to the final vowel at their offset
- or, contained a nasal vowel

were eliminated from the corpus.

5.1.3.2 Sound manipulation

The metrical condition (±FA) was created by shortening the duration of the third syllable (FA) of the target word such that it was of equal duration to the medial, unaccented syllable and did not end in a final rise of f_0 (the two main phonetic signatures of FA, see table 1.3). This procedure was first done automatically using a customized script in PRAAT (Boersma & Weenink, 2016) which cut the waveform, and then fine-tuned manually to correct perceptual bursts.

In order to keep a natural sound to the stimuli, we additionally applied a fade-out by filtering the end of the sound files with the latter half of a Hanning window (see figure 5.3 for an example of the duration manipulation on the stimulus 'chibuté' for the lexical decision study presented in section 6.2.2). Because in a relatively small number of stimulus items ($< 5\%$), FA was still perceptible due to a small final rise in f_0, we additionally applied a modified version of the algorithm presented in section 5.1.2 to adjust this small but perceptible pitch movement. A panel of three phonetic experts selected stimuli with the most natural sound.

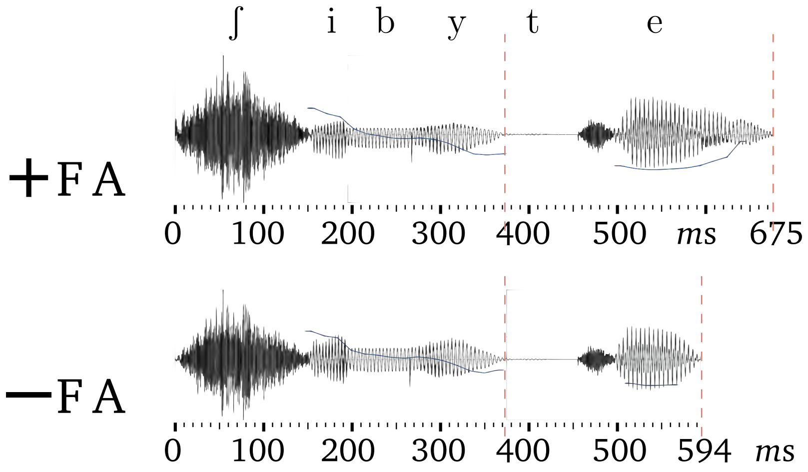


Figure 5.3: Example of duration manipulation +FA (top) and −FA (bottom) of the stimulus ʃibyte (chibuté). The duration of the final syllable is equal to the duration of the unaccented second syllable for the −FA targets.

Note that we used a slightly different procedure for the stimuli for the MMN oddball studies (presented in section 6.1). For these studies it was necessary to avoid MMNs reflecting purely durational differences between the stimuli (i.e. total word length −FA being shorter than the total length of words +FA), and make sure MMN had similar onset latencies. Therefore, durations were equalized between ±FA stimuli by shortening the first two syllables of +FA stimuli. To additionally avoid confounds from shortening the two initial syllables, these first two syllables were shortened below the perceptual threshold following Rossi (1972) and Klatt (1976).

Finally, to verify that the durational modulations on the first two syllables were not perceptible, we presented two independent French phonetic experts with an XO-task wherein they listened to word pairs that were either both manipulated on the first two syllables (25%), both without the durational manipulation (25%), or one with and the other without (50%). The listeners judged whether the two words were identical or different. Only stimuli with accuracy rates that were at or below chance-level were admitted in the current corpus (see figure 5.4 for an example of the duration manipulation on the stimulus 'paradis' for the oddball study presented in section 6.1.1).

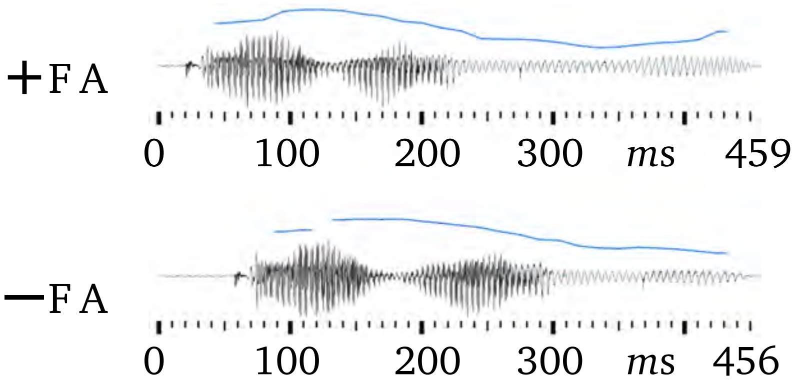


Figure 5.4: Example of durational manipulation for +FA (top) and −FA (bottom) of the stimulus paʁadi (*paradis*). The two waveforms and associated pitch tracks show how syllable duration was shortened substantially for the final syllable, and moderately for the initial two syllables.

5.2 Statistical procedures

5.2.1 Behavioral analysis

Behavioral data was analyzed with a Mixed Effects Model using the `lme4` package (Bates et al., 2012) in R (Team, 2014). Mixed models are a type of regression model that take into account both the variation in the dependent variable (i.e. the **outcome**) that is explained by the independent variables of interest (i.e. the **predictors** or **fixed effects**) and the variation that is not explained by the predictors (i.e. **random effects**). Different models were used for our different outcome types. For response latencies, a continuous measure, we used linear regression, and for the accuracy rates, binary logistic regression was used to analyze model fit. Note, however, that because both types of model included a 'mixture' of fixed and random effects, they are both mixed models. Below, we will explain why we chose to analyze our data with mixed models, particularly concerning the inclusion of random effects which is especially advantageous—if not necessary—in the analysis of repeated-measures datasets such as obtained in the current work. Simultaneously, the fixed effects and random effects specified in the studies will be presented, but note that they will be explicitly stated again for each individual study.

5.2.1.1 Fixed and random effects

Although one of the main difficulties in mixed regression analysis, is determining which predictors should be specified as fixed effects and which predictors should be included as random effects (e.g. Harrison et al., 2018), the decision was straightforward for the studies presented here. Fitting a variable as a fixed effect, or predictor, assumes all levels of interest within the predictor are represented in the data, that they are independent, and that they share a common residual variance. Predictors are generally the controlled manipulation in the experiments. In the studies presented in the current dissertation, two fixed predictor effects were included in the models: the metrical template (i.e. ±IA or ±FA) and the lexical accord (i.e. lexicality or semantic congruity). Additionally, the interplay between the stress templates and the lexical manipulation as reflected by the outcome (i.e. the interaction between the predictors) was also tested. For instance, to find out whether the presence of the French initial accent facilitates lexical decision making and whether such an effect depends on the lexicality of an item (i.e. word versus pseudoword), we fitted the data to a model with the response latencies as the outcome and the interaction of ±IA and lexicality as the predictors.

Fitting a variable as a random effect, assumes that only a subset of possibilities sampled from a complete set is represented in the data. That is, random effects describe those variables where only a sample of all possible levels, which have their own mean and variance, has been collected. Random variables are often clustered, non-independent observations that increase the error term (i.e. the deviations from the predictions) and cannot be controlled experimentally. Not only do random variables add variability to the data that is independent from the predictors, but, because they are often clustered and interdependent, they violate the assumption of independence, an important assumption in most statistical procedures (e.g. Baayen et al., 2008; Winter, 2013; Singmann & Kellen, 2017).

For instance, a given participant in an experiment may be slower than another, which adds variability to the data that is not due not the conditions under investigation in the experiment. However, the participant being slower will affect all *his* response latencies in a similar way, i.e. these different response latencies are interdependent (correlated), meaning that the participant's response in one condition is predictive of his response in another condition. As mentioned above, standard statistical procedures (e.g.

ANOVA or ordinary least-squares regression) are not robust to such violations of independence, and will produce increased Type I errors (i.e. false positives) or, more generally, overconfident results (e.g. Barr et al., 2013; Singmann & Kellen, 2017). In mixed models, individual differences can be modeled by assuming different random intercepts (baseline levels) for each participant, meaning that each participant is assigned an estimated intercept value (i.e. the displacement of each participant from the grand mean). Assuming a different baseline response latency for each participant in mixed models resolves the violation of independence.

Besides the random intercepts, models may additionally include random slopes. In random slopes, the potential dependencies in the data brought about by, for instance, the different participants with regard to the within-subjects fixed effects is taken into consideration. As explained above, the random intercept takes into account the differences between the levels in the random variables, however, it also assumes that the difference is equal for all predictor levels, which is not always the case. Thus, fitting only a random intercept allows, for instance, participant baselines to vary, but assumes a common slope for the fitted covariate (fixed effect), while fitting random intercepts *and* slopes allows the slope of a predictor to vary as well. Again, not adding random slopes, such as in simple fixed effects regression models, can inflate α and increase Type I error rates (e.g. Barr et al., 2013; Winter, 2013; Singmann & Kellen, 2017), which is undesirable.

Straightforward subsets or sample variables in the studies presented here are the listeners (i.e. we did not test the entire population of French listeners, and within the batch of selected listeners, some may be stronger than others at the task at hand or more sensitive to the manipulation) and the stimuli (i.e. only a selection of, for instance, all lexical words and pseudowords were presented in the experiments, some of which may be more obvious than others). Thus, including listeners and stimuli as random effects allows for control of interdependence by constraining the data from each individual listener and each individual item to have the same intercept and/or slope (e.g. Baayen et al., 2008) and additionally helps account for part of the error term.

In sum, mixed models take both fixed and random effects and provide so-called *partial pooling* by allowing for differences between the levels in the random variables while estimations are simultaneously based on group-level data that is assumed to be normally distributed. Therein, mixed models contrast with the more standard statistical procedures wherein all data-points are treated as independent observations, so-called *complete pooling*, which violates the independence assumption, while they also refrain from estimating effects for each individual level of the random variables, so-called *no pooling*, which decreases statistical power (i.e. increases Type II error rates). Estimation instead takes into account the factors that generate between-level variation, as well as within-level variation, allowing for more accurate estimates of the effects, improved statistical power, and minimized Type I errors.

It is often recommended to start the analysis with a **maximal model** which takes the maximal random effects structure (i.e. by-participant and by-stimulus random intercepts as well as random slopes for the fixed effects) (e.g. Barr et al., 2013; Singmann & Kellen, 2017). Note, however, that the maximal model is the model *hypothesized* to best account for the variability in the data. Whether the maximal model actually *is* the best fit to the data, or could be simplified, requires formal testing, which is discussed below.

5.2.1.2 Model selection

Once the maximal model is specified, it is important to both check model assumptions (i.e. normality, homoskedasticity, independence) as well as verify whether all predictors significantly contribute to the fit of the model. The (distributions of) residuals and model assumption are best inspected graphically,[2] meaning examining plots such as histograms, scatter-plots, quantile-quantile plots to check the distribution of the model residuals versus the fitted/predictor variables (e.g. Baayen et al., 2008; Winter, 2013; Singmann & Kellen, 2017; Harrison et al., 2018).[3]

Next, it is important to check if model complexity can be reduced. Model selection typically refers to establishing the best trade-off between the fit of a model and model complexity. This implies that there is usually no ‘per-

---

[2] One may also use more formal tests, such as Shapiro-Wilk to see if residuals deviate from Gaussian ideal. However, as such formal tests are basically significance tests (testing $H0$) and therefore often a function of sample size rather than of effect size of non-normality, we preferred visual inspection of diagnostic tests (such as QQ-plots) as they more reliably measure the degree of non-normality. From a visual inspection of the residual plots, we inferred whether a deviation from Gaussian ideal was small enough to allow for inferences.

[3] If assumptions are violated, this can sometimes be corrected by, for example, transforming the data (e.g. log-transform) or by taking the violation into consideration during model selection and interpretation. Because the datasets presented in the current dissertation tended to meet model assumptions, we will not go into detail on these steps here, but see e.g. Harrison et al. (2018) for recommendations.

fect method' to base model selection on, and, indeed, several methods of model selection are available (e.g. information theory, null-hypothesis testing), each of which optimizes the fit versus complexity trade-off differently.

In the current work, we chose null-hypothesis significance testing because we were less interested in finding the 'perfect' model that best accounts for the data, and more in establishing whether a predictor or fixed effect significantly accounted for the variability in the observed data. That is, we recognize that, for instance, response latencies can be co-determined by factors other than metrical pattern and lexical accord, and so the model that best accounts for the variability in response latencies could be quite complex, but we chose to keep the models simple and specifically evaluate whether our manipulation significantly improved the fit of the model and thus significantly predicted the outcome.

We had no specific, a priori hypotheses on the random structure underlying the data, which is why we used stepwise deletion from the maximal random structure to simpler structures to determine the most appropriate combination of random intercepts/slopes. We did have a priori hypotheses about our predictors, which is why, for the fixed effects, we specifically compared models containing the parameter of interest to the corresponding 'null' model without the parameter (but with the same random structure). We used likelihood ratio tests, which inform on whether adding a parameter to the model significantly improves model fit and therefore should be included. Restricted maximum likelihood (REML) was used for estimating variance components of random effects, while maximum likelihood (ML) was used to compare models with the same random structure but different fixed effects.The likelihood ratio test essentially informs on how much more likely the data is under a more complex model (including the predictor of interest) than under the simpler model (without that predictor). If the model comparison is significant, we conclude that the inclusion of a parameter is warranted in the model because it improves model fit, and that the parameter (i.e. our manipulation) significantly impacted the outcome.

5.2.2 EEG analysis

As described in chapter 3, neurons communicate by means of electrical signals that travel along neural pathways. These electrical signals, or the amplitude fluctuations they generate, can be recorded non-invasively using electrodes placed on the scalp (which is typically prepared with gel to improve conductance) in a measure called electroencephalography (EEG). The raw and continuous EEG recordings are typically quite noisy and requires multiple preprocessing steps before ERP analysis. The specific preprocessing steps in ERP analysis conducted in the present dissertation are described below.

5.2.2.1 Recording

EEG recordings typically require three types of electrodes: active electrodes, the ground electrodes, and reference electrodes. The ground electrode is connected to a ground circuit, which is attached to an amplifier so as to give the EEG voltage sufficient power to be digitized (at 2048 Hz sampling rate, in the current thesis) and displayed graphically. This ground circuit also records noise from the amplifier circuitry, which is canceled out by 'differential amplifiers'. Differential amplifiers measure the difference between active-ground voltage and reference-ground voltage, i.e. reference electrodes are used along with active and ground electrodes. Because outside noise will be the same for both voltages, it can be eliminated by subtraction.

Ground electrodes can be placed anywhere on the scalp, but reference electrodes are usually placed on neutral sites where they capture few to, ideally, no neural activity. As is explained below, for the studies presented in the current dissertation, the reference sites are usually the mastoids (i.e. the protruding area of the temporal bone that is located behind the ear), although, in some experiments, the signal captured at the mastoids contained too much noise and therefore required re-referencing to the average of the electrodes. For this reason, the reference will always be stated explicitly for each study.

Active electrodes are generally arranged according to the International 10−20 system (see figure 5.5). In this arrangement, electrodes are placed at 10% or 20% intervals along lines of latitude (left-right) and longitude (front-back). Each electrode has a letter to mark the lobe (F: Frontal; T: Temporal; C: Central; P: Parietal; O: Occipital) and a number which indicates the hemisphere location (even numbers: right hemisphere; odd numbers: left hemisphere; *z* (zero): midline). The larger the electrode-number, the further away its position from the midline. How many active electrodes are used in an experiment may vary, and in fact does so in the current dissertation, therefore, also the total number of electrodes will be stated explicitly per study.

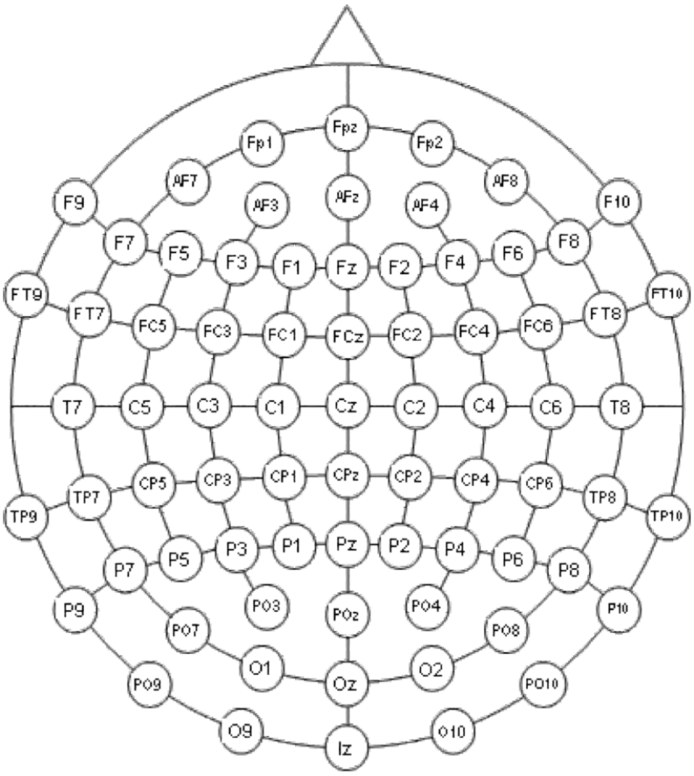


Figure 5.5: Active electrodes are arranged according to the International 10 – 20 system, wherein electrodes are placed at 10% or 20% intervals along lines of latitude (left-right) and longitude (front-back). Letters mark the lobe (i.e. F: Frontal; T: Temporal; C: Central; P: Parietal; O: Occipital) and numbers indicate the hemisphere location (ie.e. even numbers: right hemisphere; odd numbers: left hemisphere; *z* (zero): midline).

5.2.2.2 Preprocessing

Filtering Preprocessing started with filtering out voltage-fluctuations at frequencies unlikely to have a neural source. In a filtering process, the time-series signal is transformed to the frequency domain, such that amplitudes at specified frequencies can be attenuated, after which the signal is back-transformed to the time domain. High-pass filters attenuate the amplitude at frequencies below the specified cutoff frequency to, for instance, remove slow drifts, and, conversely, low-pass filters attenuate the amplitude at frequencies above the specified cutoff frequency to, for instance, remove noise from surrounding electric devices. Bandpass filters combine high-pass and low-pass filters to attenuate frequencies outside of the frequency band. The filters attenuate frequency amplitude along a slope which crosses the cutoff frequency. In the studies presented here, following the recommendations of Luck (2005), the EEG signal is FIR bandpass filtered between $0.01 - 30$ Hz.

Independent Component Analysis Artifacts such as oculomotor artifacts (i.e. eye-movements or blinks) can be removed by Independent Component Analysis (ICA) (e.g. Makeig et al., 1996). ICA decomposes the EEG signal, using an 'unmixing' matrix to find maximally temporally independent individual components which comprise the data, including their probable electrode locations. Important for ICA is to clean the data best as possible beforehand, in order to assure correct identification of independent components, which was done manually. Additionally, the data were high-pass filtered at 1 Hz for the independent component analysis *only* (Winkler et al., 2015).

The resulting component matrix was next transferred to the EEG set that was high-pass filtered at 0.01 Hz. Components reflecting the oculomotor artifacts were identified by means of visual inspection and removed, after which the EEG data was recomposed without the removed artifact-components. Furthermore, individual noisy electrodes that distorted a dataset with noise, were interpolated (i.e. reconstructed by averaging the weighted signal of neighboring electrodes) to keep data input statistically consistent across participants, while improving the signal-to-noise ratio.

Averaging the event-locked response The next step is to select trial epochs with the 'event' or 'time-mark' at the appropriate time point in the stimulus presentation. A baseline window is used to set a reference amplitude value for measuring the hypothesized ERP components. Baseline correction normalizes the amplitude of the ERP components between experimental conditions such that the mean amplitude in the baseline window is equal to zero μV (Luck, 2005). Typical baseline windows are 200 ms, which is long enough to avoid residual component amplitude differences based on small fluctuations due to noise.

After this step, it is often recommended to again inspect the the epochs for remaining artifacts. In the current work, this was done automatically as well as visually so as to fine-tune the rejection of data segments (at times epochs marked as artifacts may appear reasonable on visual inspection, while, conversely, the automatic rejection may also miss noisy epochs that should be rejected).

Finally, epochs are averaged per condition and per subject, to obtain per participant ERPs for statistical analysis (discussed next).

5.2.2.3 Statistical procedure

With its high temporal resolution, EEG provides a rich database to determine the exact latency of an effect. However, testing at all data points independently quickly leads to a multiple comparison problem where the risk of making Type I errors increases considerably. Because EEG measures are not independent, but in fact temporally and spatially correlated, a non-parametric t_{max} permutation test is used to analyze the data (Groppe et al., 2011; Luck, 2014).

In t_{max} permutation testing, the null distribution is estimated by repeatedly resampling the obtained data and calculating t-scores for each sample. The most extreme t-scores (t_{max}) are selected for the null distribution. Finally, the t-scores of the observed data are computed and compared to the simulated t_{max} distribution, just as in parametric hypothesis testing. Because with each permutation the chance of obtaining a large t_{max} increases, the test automatically becomes more conservative when making more comparisons. Also, because the actual, obtained data is used to estimate the null distribution, the test does not assume test independence, allowing for stringent control of Type I error without considerable decrease in sensitivity.

Below it will be described in more detail why we chose to analyze our data with non-parametric t_{max} permutation tests, particularly in comparison to more traditional statistical procedures such as the standard t-test or analysis of variance.

Conventional t-test or ANOVA In the conventional parametric approach to ERP statistical analysis, amplitude and latency measurements are treated as standard dependent variables. That is, amplitudes and peak latencies will enter a t-test or ANOVA in an ERP experiment the same way, for instance, reaction times in a behavioral experiment. However, ERP variables differ from behavioral variables in a number of ways.

The first difference is that ERP measurements tend to be noisier than behavioral measurements. Such noise can be problematic since, with a t-test or ANOVA, significance (i.e. p-values) not only depends on the difference in means between conditions, but also on the amount of variance within each of the conditions. Thus, the more variance there is in the data, the lower the power of the statistical procedure. A bigger and somewhat more complicated difference between ERP and behavioral data is the richness within ERP data. EEG data is made up of a two dimensional spatio-temporal structure.[4] The space dimension is characterized by the number of electrodes and the temporal dimension by the number of time points as specified by the sampling rate. This richness quickly leads to many comparisons. Multiple comparisons inflate α and therefore lead to a higher probability of making a Type I error (i.e. rejecting $H0$ when it is actually true).

A common recommendation to control α, is to use the Bonferroni correction. The Bonferroni correction is a very simple correction that provides strong control of Type I error probability. The correction takes the decided upon α (usually 0.05) and divides it by the number of tests carried out. The advantage of this correction is that α is maintained at 0.05 and therefore significant p-values are likely to be *real* effects (i.e. there is no increased Type I error probability). A disadvantage of applying Bonferroni correction is that it unnecessarily decreases sensitivity in EEG/ERP studies, that is, statistical power drops and Type II error probability increases. This is because Bonferroni makes a false premise about t-tests or comparisons in ERP studies by assuming that they are independent from each other, which is not the case. Tested time and location points in EEG data are close together and highly correlated (i.e. they are predictive of one another). So while traditional parametric procedures often lead to a multiple comparison problem wherein the probability of making Type I errors increases dramatically, correcting with Bonferroni increases the probability of making Type II errors. Non-parametric tests, such as $t_{\max}$ permutation test, provide a middle-ground in the trade-off between sensitivity and specificity of statistical tests.

---

[4] Actually EEG structures are three dimensional if we were to also consider their spectral dimension.

Non-parametric $t_{\max}$ permutation test Advances in statistical theory, together with the increased computational power of computers, provide alternative solutions for the multiple comparisons problem inherent to/in EEG data analysis; non-parametric mass-univariate hypothesis testing. One advantage of non-parametric analyses, is that they do not rely on assumptions on the probability distribution of the data. Another advantage is that non-parametric tests take into account the fact that the individual t-tests are not actually independent. Also, while with conventional parametric analysis each variable (i.e. latency or amplitude) is tested individually, with non-parametric analysis the complete spatio-temporal dimension can be tested with a single test statistic (e.g. Maris & Oostenveld, 2007; Groppe et al., 2011; Luck, 2014).

Similar to the Bonferroni correction, the permutation approach also provides strong control over the Type I error probability. With the permutation test, actual t-scores (t-scores derived from testing our HA) are compared to a distribution of t-scores under $H0$. This distribution is obtained by shuffling event codes, i.e. trials are randomly re-assigned to the experimental conditions. The largest t-score (i.e. t_{max}) computed from the shuffled data is a t-score that we now *know* belongs to $H0$. Repeating this process many times will lead to a distribution of t_{max} scores that can be expected to be obtained by chance (i.e. under $H0$, if there was no difference between the conditions).

Next, the test for significance is similar to conventional parametric tests. That is, t-scores obtained from the actual, unshuffled data have to be larger than the critical t_{max} scores in the simulated $H0$ distribution. The procedure cleverly corrects for multiple comparisons because the t_{max} distribution, from which the p-values of each comparison are derived, automatically adjusts to reflect the increased chance of false discoveries due to an increased number of comparisons. As more comparisons are made, there is an increased probability of obtaining more extreme observations by chance and the test becomes more conservative. This can however be a disadvantage as well, particularly there are no or few a priori hypotheses. In such cases, the number of tests increase, the statistical power of the test is weakened and, consequently, test sensitivity decreases (although still less dramatically than in Bonferroni correction). It is therefore advisable to additionally reduce the number comparisons, for example, by specifying strong hypotheses prior to testing. The prior expectations can be used to pre-select electrodes and time windows of interest, allowing for the number of comparisons to be kept minimal. Additionally, we will typically try to keep sampling rate low at the time of statistical testing, to further reduce the number of comparisons and increase power while maintaining appropriate Type I error rates.

6 Studies

In this chapter, we will present the five studies conducted in this dissertation. The studies are presented with an introduction wherein we explain precisely which question motivated the investigation and how we derived our predictions.

When appropriate, some of the studies are grouped in order to facilitate reading. For instance, section 6.1 presents two oddball studies that set out to investigate the phonological representation of French accentuation, i.e. FA in the first, and FA and IA mixed in the second. Similarly, in section 6.2, we present two lexical decision studies which were constructed to determine the interaction of the two French accents (IA in the first, and FA in the second) with the process of lexical access. Finally, in section 6.3, we present an investigation wherein in metrical expectancy and semantic congruity are manipulated orthogonally in order to establish whether difficulties in metrical processing also impact the later stages, e.g. semantic retrieval or semantic integration, in speech comprehension.

To further underline the natural coherence between the studies, with each study motivating the next, each section ends with an elaborate discussion of results wherein the findings are related to each other .

All in all, we will show how the results fortified and motivated our inquisition of metrical stress processing in French.

6.1 Phonological representation of French metrical stress

As was explained in section 1.5, in French, accentuation is not lexically distinctive and tightly intertwined with intonation. This has led to the language being described as *'a language without accent'* (Rossi, 1980) and to French accentuation being attributed a rather trivial role in speech processing. Indeed, as some authors have argued, if French does not know lexical stress, it is reasonable to assume that its speakers are confronted with stressed syllables too infrequently to be able to develop a sensitivity to accentual information—essentially leaving them *'deaf to stress'* (Dupoux et al., 1997). Because listeners can still readily decode speech, despite this presumed 'phonological deafness', according to the authors, it stood to reason that accentuation unlikely plays an important function in French comprehension processes.

However, we argued, that if one considers Di Cristo's model in which the metrical structure of speech plays a central role (Di Cristo, 2000), it becomes possible to envision stress templates underlying the cognitive representation of the lexical word. If stress templates are phonologically encoded at the level of the word, they may readily contribute to speech comprehension. Studies investigating the phonological status of French accentuation have all reported results in favor of a sensitivity to the metrical structure of words. Not only were metrical incongruences (stress on the medial syllable, a violation in French) found to slow down semantic processing (Magne et al., 2007), but a series of perception studies showed both the initial accent (IA) and final accent (FA) to be metrically strong, independent from phrase boundaries (e.g. Astésano et al., 2012; Garnier et al., 2016; Garnier, 2018). Furthermore, IA was perceived even when its phonetic correlates were suppressed or when its f_0 rise peaked further along on the word, again suggesting a metrical expectation for the accent (Jankowski et al., 1999; Astésano & Bertrand, 2016; Garnier et al., 2016; Astésano, 2017; Garnier, 2018).

Additional evidence against the notion of stress deafness in French, is provided by recent neuroimaging studies using the event-related potentials technique (ERP; presented in section 3.2). Results from a recent MisMatch Negativity (MMN) study confirmed a long-term memory representation and phonological preference for IA (Astésano et al., 2013; Aguilera et al., 2014). In the oddball study, trisyllabic words were presented either with IA (+IA) in deviant position or without IA (−IA) in deviant position (see section 5.1.2 for information on the sound manipulation). 30 listeners took part in the experiment, 14 of which listened to the condition wherein the stimulus +IA was in the standard position and the word −IA in the deviant position, while for the other 16 listeners, positions were reversed. All 30 listeners completed two tasks, one passive task during which they listened to the stimuli while attending a silent movie, and one active task during which the listeners were asked to respond as quickly and accurately as possible when they detected the deviant stimulus.

Results indicated that the listeners clearly distinguished between the trisyllabic words carrying IA and those that did not. This indicates that French listeners are in fact not deaf to stress, but readily perceive the accentual manipulation. Interestingly, the authors additionally observed an asymmetry between the MMN elicited by +IA deviants and the MMN following −IA deviants. That is, when the deviant had been presented without initial accent, a clear MMN component emerged, while this MMN was significantly smaller, when the deviant was presented with initial accent. Recall from section 3.2.1, that an oddball paradigm typically elicits an MMN when a low-probability stimulus (the oddball or deviant) occurs within a train of high-probability stimuli (e.g. Näätänen et al., 2007). Therefore, not finding an MMN when presenting the oddball with IA indicates a long-term representation for the initial accent. Indeed, it is plausible that, if IA is part of a preferred stress template, only rarely presenting the template might make it the deviant within the experiment, but it does not make the template improbable. In other words, in the condition in which the oddball was presented with IA, while atypical in the context of the test, the oddball was still the expected stress template. Therefore, no MMN emerged.

Indeed, the MMN was argued in section 3.2.1 to be the prototypical component for prediction mismatching input (e.g. Näätänen et al., 2007; Garrido et al., 2009; Winkler et al., 2009; Denham & Winkler, 2017). The MMN is pre-attentive (i.e. automatic) and its amplitude is held to reflect the magnitude of the deviance from what was expected (Sussman, 2007; Näätänen et al., 2007; Sussman et al., 2014; Sussman & Shafer, 2014). The frequently occurring standard stimuli are assumed to develop predictions that are subsequently violated by the infrequently occurring deviant stimulus. While such a deviance can be purely acoustic and bottom-up, switching the positions of the standard and deviant stimuli allows for more substantial inferences on the phonological or long-term memory foundation of the manipulated prosodic entity (e.g. Winkler et al., 2009; Garrido et al., 2009).

This means that, if the +IA and −IA stimuli were to differ only acoustically, the MMNs should have been similar between both conditions. If there is, however, an asymmetry in MMN amplitudes—such as observed in Aguilera et al. (2014)—the asymmetry is informative about the strength of underlying memory traces. The results of Aguilera and colleagues thus show that when the train of the (improbable) −IA standards was interrupted by the more probable +IA deviant, the prediction violation (and thus MMN) was smaller than in the reverse case wherein the repeatedly presented (probable) +IA stimulus (i.e. the stimulus with an apparent firm phonological representation) only increased the listeners confidence about upcoming words to be marked with an initial accent in their underlying stress templates, leading to a large mismatch response when the anticipation was violated.

In this view, the MMN is linked to the abstractionist theory (Eulitz & Lahiri, 2004; Cutler, 2010, see section 2.2.1). Recall that, according to the abstractionist theory, phonological entities are stored into abstract and general representational units which allow for efficient coding. That is, the abstract phonological representations allow for auditory input to be compared or generalized to anticipated speech units. Such a mechanism of efficient coding parallels the principle of predictive coding in neuroscience (Friston, 2005, see also Scharinger et al. 2012, 2016), which holds that perception is less concerned with the fine-grained analysis of sensory information, but instead crucially depends on the ability to generate global expectations and compare or generalize incoming, bottom-up information to those predictions. Bottom-up evidence that matches with the prediction is categorized with little effort, while only the mismatching information requires further processing, a cognitive effort that is reflected in a modulation of the MMN. In the study of Aguilera et al. (2014) the results thus demonstrate that IA is encoded and part of the abstract phonological stress pattern in French.

In order to further ascertain that the observed MMNs were independent from differences in acoustic processing, Aguilera and colleagues carried out an additional analysis wherein they compared the MMNs resulting from the difference wave between −IA–deviants and −IA–standards to the difference wave between +IA–deviants and +IA–standards (i.e. between participants comparison). Again, results indicated that the difference between stimuli without initial accent was significantly larger than the difference between stimuli with initial accent, allowing for the purely acoustic interpretation of the results to be ruled out.

Finally, the behavioral results from the active task confirmed the interpretation of the ERP results. That is, the deviant stimuli −IA were slower to detect than the deviant stimuli +IA, and generated more detection errors. Moreover, the behavioral results between the two blocks in the experiment, showed that reaction times and error rates improved in the second block but only for deviants that had been presented with initial accent, which shows a learning effect generally observed only for deviants that are stored in long-term memory (Astésano, 2017). Overall, Aguilera and colleagues thus not only show that stimuli without IA are noticed by listeners, but also that IA is anticipated and attached to the metrical template underlying the representation of words.

In the current study, we set out to build on these findings and investigated the phonological representation of the French final accent (FA). In section 1.5 we argued words to be encoded with bipolar stress templates underlying their representation, marking not only the left (IA) but also the right (FA) lexical boundaries. Indeed, in a study directly addressing the perception of FA, participants showed little difficulty recognizing whether or not words were marked with the primary stress (Michelas et al., 2016, 2018), again contradicting the notion of 'stress deafness' for French.

Here we sought to determine whether FA is phonologically represented, similar as IA, and manipulated the presence of FA on trisyllabic words in an auditory oddball paradigm. In a first study, we presented participants with an oddball paradigm wherein either the standard word was presented with final accent and the deviant was presented without, or vice versa. In a second study, standards were presented with their full bipolar stress templates, including both IA and FA, while deviants were either presented without FA or without IA. We expected that, if words are encoded with both accents underlying their phonological representation, then ±FA deviants should result in asymmetrical MMNs, similar as ±IA deviants.

6.1.1 MisMatch Negativity: Final Accent

6.1.1.1 Study Summary

Research Question: *Are French listeners sensitive to the French primary final accent (FA) and is the accent part of the French phonologically expected stress pattern?*

Procedure: *19 participants passively attended a silent movie in an oddball paradigm. In one condition, deviants were presented without (−FA) and standards with final accent (+FA), while in another condition, these positions were switched.*

Results: *We obtained asymmetric MMN waveforms, such that deviant −FA elicited a larger MMN than deviants +FA (which did not elicit an MMN). Additionally, the difference waveforms between identical stimuli in different positions within the oddball paradigms, indicated −FA stimuli to be disfavored whether they were the deviants or the standards.*

Conclusion: *French listeners are not deaf to the final accent. Instead, the results indicate the final accent to be phonologically encoded and attached to expected stress pattern underlying the representation of the lexical word.*

Procedure:

- Nr participants → 19 (8 dev −fa and 11 dev +fa)
- 986 standards, 106 deviants
- isi → 600 ms

Preprocessing:

- Nr electrodes → 64
- Reference → Mastoids
- Filter and down-sampling → 0.01 − 30 bandpass, 128 Hz
- Epoch length → −100 − 1000

Analysis:

- Time-window → 551 − 651 ms
- Electrodes → Fz, Cz, FC1, FC2, F3, F4, C3, C4

Results:

- mmn-deviant −fa marginally significant
- mmn-deviant +fa not significant
- Between mmn significant
- Within +fa significant
- Within −fa not significant

6.1.1.2 Methods: Final Accent

6.1.1.2.1 Participants

The study was conducted in accordance with the Declaration of Helsinki. 21 French native speakers, aged 19 − 31 (mean age 24.0), gave their informed consent and volunteered to take part in the study. All subjects were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Two subjects were excluded from the EEG-analysis due to excessive artifacts in the signal.

6.1.1.2.2 Speech stimuli

Two trisyllabic French nouns were used in the current experiment (‘casino’ ([kazino], *casino*) and ‘paradis’ ([paʁadi], *paradise*). The stimuli were extracted from sentences spoken by a naïve native speaker of French. Stimuli with the most natural FA (+FA) were selected by a panel of three experts and re-synthesized without FA (−FA) by reducing the length of the final (third) syllable to match the duration of medial unaccented syllable in Praat (Boersma & Weenink, 2016, see section 5.1.3 for more information on the manipulation).

In order to avoid MMNs reflecting purely durational difference between the stimuli (i.e. total word length −FA being shorter than the total length of words +FA) (e.g. Jacobsen & Schröger, 2003; Colin et al., 2009; Honbolygó et al., 2017) and make sure MMN had similar onset latencies, durations were equalized between ±FA stimuli by shortening the first two syllables of +FA stimuli. To additionally avoid confounds from shortening the two initial syllables, these first two syllables were shortened below the perceptual threshold following Rossi (1972) and Klatt (1976), and as judged by two independent French phonetic experts (see figure 6.1 for an example of the duration manipulation on the stimulus ‘paradis’ and table 6.1 for an overview of syllable durations for both words with and without FA). This led to total word durations of 503.3 ms and 500.8 ms for ‘casino’ +FA and −FA, respectively, and 459.0 ms and 456.5 ms for ‘paradis’ +FA and −FA, respectively, with third syllable durations of 233.3 ms and 110.7 ms for ‘casino’ +FA and −FA, and 225.9 ms and 148.3 ms for ‘paradis’ +FA and −FA.

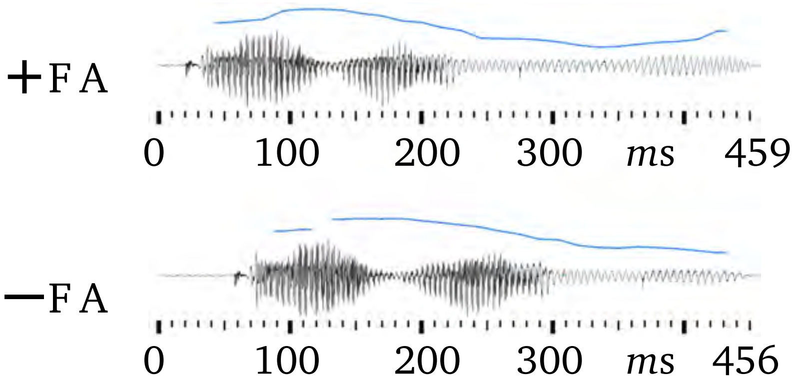


Figure 6.1: Example of durational manipulation for +FA (top) and −FA (bottom) of the stimulus paʁadi (‘paradis’). The two waveforms and associated pitch tracks show how syllable duration was shortened substantially for the final syllable, and moderately for the initial two syllables.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd syllable
	ms	f_0	*ms*	f_0	*ms*	f_0	*ms*	f_0	onset
CASINO									
+FA	503.3	117.8	112.9	125.8	157.1	122.2	233.3	111.0	269.0
−FA	500.8	119.0	146.9	125.1	178.1	122.3	178.8	110.7	325.0
PARADIS									
+FA	459.0	106.5	121.5	121.7	111.6	118.9	225.9	93.4	233.1
−FA	456.5	110.9	168.8	123.3	139.3	119.1	148.3	95.1	308.1

Table 6.1: Overview of durational and f_0 values, plus the timing of the third syllable (holding ±FA) onset for both ‘casino’ and ‘paradis’ with and without final accent.

Because in MMN studies which set out to investigate word processing, it is generally recommended to reduce stimulus variation between the standard and deviant as much as possible (Pulvermüller & Shtyrov, 2006; Honbolygó & Csépe, 2013), the oddball paradigm in the current study either presented only 'casino', or only 'paradis'. However, because we were interested in the phonological representation of FA, which should be similar between the two words, the data obtained from both versions are combined in the analysis (see below).

In both versions, there were a total of 1092 presentations, 986 standards and 106 deviants. The deviant could be either −FA with +FA as standard, or +FA as deviant and −FA in standard position. This means that there were a total of four versions of the oddball paradigm: (1) casino–deviant +FA, (2) casino–deviant −FA, (3) paradis–deviant +FA, and (4) paradis–deviant −FA.

6.1.1.2.3 Procedure

Each participant was comfortably seated in an electrically shielded and sound attenuated room. Stimuli were presented through headphones using Python2.7 with the `PyAudio` library on a Windows XP 32-bit platform. To ensure attention was diverted from the stimuli, participants watched a silent movie with no text (*Best of mr. Bean*).

Lists were assigned randomly: 4 participants listened to the casino–deviant +FA version, 3 listeners to the casino–deviant −FA version, 7 participants listened to paradis–deviant +FA and finally 5 participants had the paradis–deviant −FA version. This meant that data was obtained from 11 participants for the version in which +FA stimuli were in deviant position and −FA stimuli in standards, and from 8 participants for the version wherein ±FA positions were reversed.

Each participant listened to the complete list of 1092 stimuli (986 standards, 106 deviants) in one block, which lasted approximately 25 minutes. Deviants were interspersed randomly and online, while avoiding two consecutive occurrences and making sure that each list started with at least 25 standards. Finally, the inter-trial interval (ITI) consisted of stimulus duration plus inter-stimulus interval (ISI) (i.e. $\sim 475 + 600$ ms).

6.1.1.2.4 EEG recording and preprocessing

EEG data were recorded with 64 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958). The EEG signal was amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 2048 Hz.

The data were preprocessed using the `EEGlab` package (Delorme & Makeig, 2004) with the `ERPlab` toolbox (Luck et al., 2010) in Matlab (Mathworks, 2014). Each electrode was re-referenced offline to the algebraic average of the left and right mastoids. The data were band-pass filtered between $0.01 - 30$ Hz and resampled at 128 Hz. See section 5.2.2 for more details on the preprocessing of the EEG signal relating to, for instance, artifact rejection.

6.1.1.2.5 Analysis

The data were analyzed with the non-parametric t_{max} permutation test (Groppe et al., 2011; Luck, 2014, see section 5.2.2 for more information on this statistical procedure) using the `Mass Univariate ERP Toolbox` (Groppe et al., 2011) in Matlab (Mathworks, 2014).

We were interested in modulations of the MMN as elicited by the presence/absence of FA and therefore specifically tested for differences in the time-window between $551 - 651$ ms. Furthermore, because the MMN is a fronto-centrally located deflection we selected the fronto-central electrodes (Fz, Cz, FC1, FC2, F3, F4, C3 and C4) for the statistical analyses. Each comparison of interest was analyzed with a separate repeated measures, two-tailed t-tests, using the original data and 2500 random permutations to approximate the null distribution for the customary family-wise alpha (α) level of 0.05.[1]

6.1.1.3 Results

We obtained no significant MMN when the deviant had been +FA (critical t-score: ± 4.3095, $p = 0.8396$, ns). This indicates that even though the +FA stress template was rare in the experimental setting, listeners *still* expected

---

[1] In fact we used more than twice the number of permutations suggested for an alpha at 5% (Manly, 2006) so as to be even more certain of obtaining reliable results.

words to be marked with final accent. Presenting the deviant without final accent elicited a marginally significant MMN (critical t-score: ±4.2958, $p = 0.0652$). Visual inspection suggests the MMN was located at left frontal electrodes, starting 600 ms post stimulus onset (i.e. ∼ 300 ms post deviance detection). Furthermore, we observed an asymmetry between MMNs; the MMN was significantly more ample when the deviant had been presented −FA than when it had been presented +FA (critical t-score: ±3.1505, $p < 0.05$, see figure 6.2), indicating a phonological preference for FA.

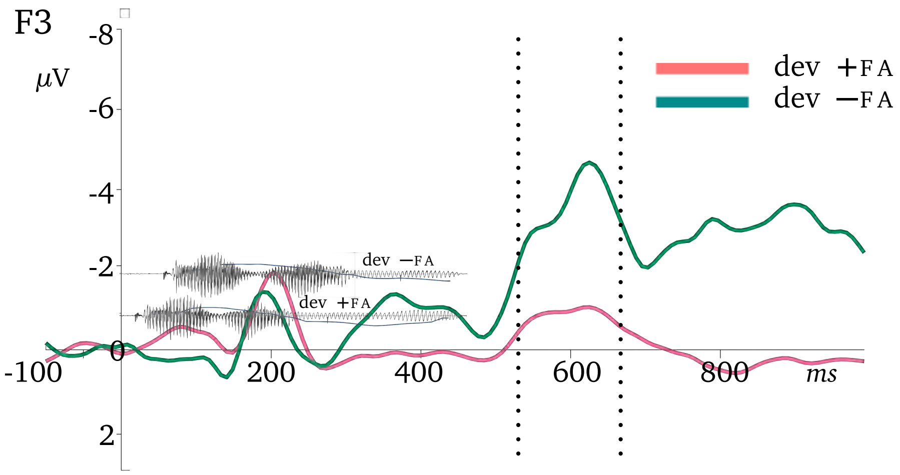


Figure 6.2: MMN components for +FA (in pink) and −FA (in green) deviants, recorded at the F3 (left frontal) electrode, with the oscillogram of the deviant stimuli [paʁadi] plotted in the background. Waveforms and oscillograms are temporally aligned to indicate the relation between the offset of the ±FA manipulation and the resulting stimulus-locked event-related potentials. The tested time-window is indicated by dashed vertical lines. For ease of presentation, ERP waveforms are low-pass filtered at 10 Hz and negativity is plotted as an upward deflection.

Finally, in the comparison between participants (i.e. comparing identical stimuli that differed in position within the oddball experiment) there was a significant difference between +FA in deviant position versus +FA in standard position at frontal (F4) and central (C4) electrodes during the whole time-window (critical t-score: ±3.7416, $p < 0.05$), while there was no such difference for stress templates −FA (critical t-score: ±4.394, $p = 0.84$, *ns*, see figure 6.3).

Note that the results presented here partially contradict those reported in Aguilera et al. (2014) in which IA had been manipulated (see also Astésano et al., 2013). In Aguilera et al. (2014), the between listeners analysis demonstrated a bigger difference between standards and deviants when stimuli had been presented −IA, than when they had been presented +IA. This discrepancy indicates differential processing between IA and FA, which is elaborated upon in the main discussion of the two experiments at the end of this section.

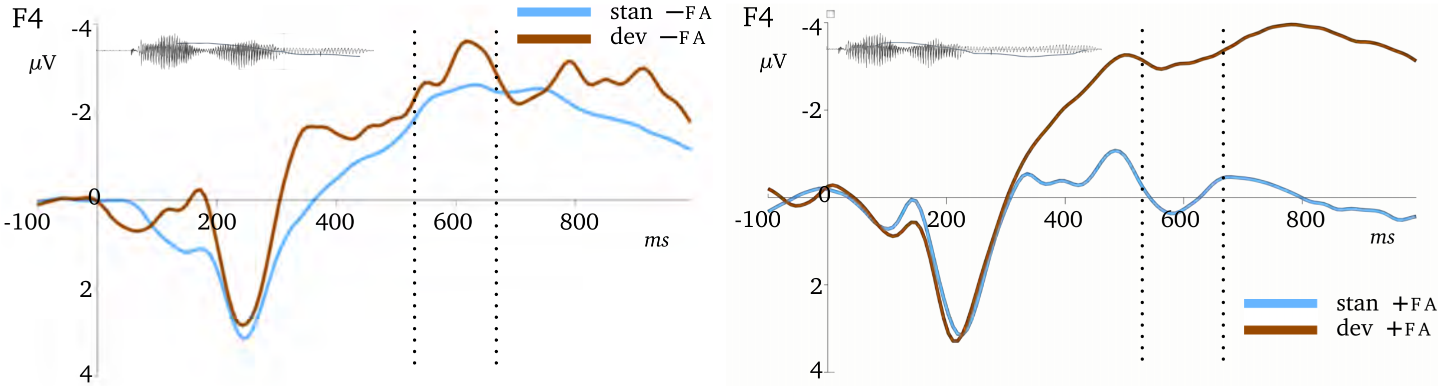


Figure 6.3: ERP components for −FA (left) and +FA (right) stimuli, recorded at the F4 (right frontal) electrode, with the oscillograms of [paʁadi] plotted in the background to indicate the relation between the offset of the ±FA manipulation and the resulting stimulus-locked ERPs. The tested time-window is indicated by dashed vertical lines. For ease of presentation, ERP waveforms are low-pass filtered at 10 Hz and negativity is plotted as an upward deflection.

6.1.2 MisMatch Negativity: Initial and Final Accent

6.1.2.1 Study Summary

Research Question: *Are the French final accent and secondary initial accent encoded similarly and do listeners have similar expectations between both accents?*

Procedure: 20 *participants passively attended a silent movie in an oddball paradigm. Standards were always presented with both the initial and final accent, while deviants were presented either without final accent (*−FA*) or without initial accent* −IA.

Results: *We obtained* MMNs *both to deviants* −FA *and to deviants* −IA, *although* −FA *deviants elicited a more ample* MMN.

Conclusion: *Both the initial accent and final accent were readily perceived by the French listeners, further arguing against the notion of stress deafness for French. The results, however, do suggest different respective roles for the final and initial accent.*

Procedure:

- Nr participants → 20
- 1000 standards, 100 −ia deviants and 100 −fa deviants
- isi → 600 ms

Analysis:

mmn-ia:

- Time-window → 201 − 301 ms
- Electrodes → Fz, Cz, FC1, FC2, C3, C4, F1, F2, F5h, F6h, FCz

mmn-fa:

- Time-window → 451 − 651 ms
- Electrodes → Fz, Cz, FC1, FC2, C3, C4, F1, F2, F5h, F6h, FCz

---

Preprocessing:

- Nr electrodes → 64
- Reference → Average
- Filter and down-sampling → 0.01 − 30 bandpass, 128Hz
- Epoch length → −200 − 1000

6.1.2.2 Methods: Initial and Final Accent

6.1.2.2.1 Participants

20 French native speakers, aged 19 – 45 (mean age 23.7; 14 female), took part in the study. None of the participants had taken part of the previous MMN study and all were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Each of the participants gave their written consent and was paid a small fee for their participation. The study was conducted in accordance with the Declaration of Helsinki.

6.1.2.2.2 Speech stimuli

The same two trisyllabic French nouns used in the previous study, were used in the current experiment (‘casino’ ([kazino], *casino*) and ‘paradis’ ([paʁadi], *paradise*). The stimuli were extracted from sentences spoken by a naïve native speaker of French. Stimuli with the most natural IA (+IA) were selected by a panel of three experts and re-synthesized without IA (–IA) using a customized quadratic algorithm in Praat (Boersma & Weenink, 2016). Using the same algorithm as (Aguilera et al., 2014), the f_0 value of the first vowel (i.e. IA) was lowered near the f_0 value of the preceding (unaccented) determinant, to de-accentuate the first syllable (i.e. remove IA; see figure 6.4 for an example of the –IA and –FA stimulus manipulation of the noun ‘paradis’). The algorithm progressively modified the f_0 values to reach the f_0 value at the beginning of the last (accented) vowel. This quadratic transformation allowed for micro-prosodic variations to be maintained, thus keeping the natural sound of the stimuli. The +IA stimuli were forward and back transformed to equalize the speech quality between +IA and –IA stimuli. See section 5.1.2, for more information on the manipulation of IA.

FA was manipulated similarly as presented in the previous study, i.e. by reducing the length of the final (third) syllable to match the duration of medial unaccented syllable in Praat (Boersma & Weenink, 2016, see section 5.1.3 for more information on the durational manipulation). In order to avoid MMNs reflecting purely durational difference between the stimuli (i.e. total word length –FA being shorter than the total length of words +FA) (e.g. Jacobsen & Schröger, 2003; Colin et al., 2009; Honbolygó et al., 2017), durations were equalized between ±FA stimuli by shortening the first two syllables of +FA stimuli (i.e. the standard +IA and +FA, and the deviants −IA but +FA). To additionally avoid shortening the two initial syllables confounding our results, these first two syllables were shortened below the perceptual threshold (Rossi, 1972; Klatt, 1976).

See figure 6.4 for an example of the sound manipulations on the stimulus ‘paradis’ and table 6.2 for an overview of stimuli properties for both words ±IA and ±FA.

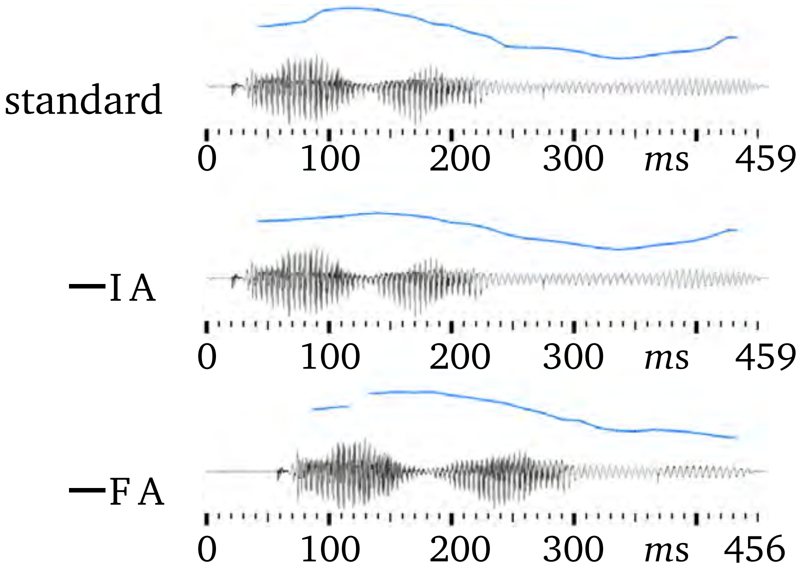


Figure 6.4: Example of the stimulus *paradis* [paʁadi]. At the top, the standard with both IA and FA, the deviant −IA (but +FA) in the middle, and the deviant −FA (but +IA) at the bottom.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd syllable
	ms	f_0	*ms*	f_0	*ms*	f_0	*ms*	f_0	onset
CASINO									
standard	503.3	117.8	112.9	125.8	157.1	122.2	233.3	111.0	269.0
dev−IA	503.3	116.0	112.9	120.5	157.1	121.6	233.3	109.9	269.0
dev−FA	500.8	119.0	146.9	125.1	178.1	122.3	178.8	110.7	325.0
PARADIS									
standard	459.0	106.5	121.5	121.7	111.6	118.9	225.9	93.4	233.1
dev−IA	459.0	104.1	121.5	114.4	111.6	114.4	225.9	93.4	233.1
dev−FA	456.5	110.9	168.8	123.3	139.3	119.1	148.3	95.1	308.1

Table 6.2: Overview of stimulus properties (durational and f_0 values, plus the timing of the third syllable onset) for both ‘casino’ and ‘paradis’ standards (with stress patterns +IA and +FA), deviants −IA and deviants −FA.

As in the previous study, we presented lists either only with ‘casino’, or only with ‘paradis’. However, because we were interested in the phonological representation of the accent (whether IA or FA), which should be similar between both words, the data obtained from both versions are merged in the analysis.

6.1.2.2.3 Procedure

Each participant was comfortably seated in an electrically shielded and sound attenuated room. Stimuli were presented through headphones using Python2.7 with the `PyAudio` library on a MacOS Sierra platform. Participants watched a silent movie to ensure their attention was diverted from the stimuli. Each participant listened to all 1200 stimuli (1000 standards, 100 deviants —IA, 100 deviants —FA) in one block, which lasted for approximately 25 minutes. Deviants were interspersed randomly and online, while avoiding two consecutive occurrences of the same deviant and making sure that each list started with 25 standards. Finally, the same inter-trial interval (ITI) was used as in the previous oddball study, and consisted of stimulus duration plus inter-stimulus interval (ISI) (i.e. $\sim 475 + 600$ ms).

6.1.2.2.4 EEG recording and preprocessing

EEG data were recorded with 64 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958). The EEG signal was amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 2048 Hz. The data were preprocessed using the `EEGlab` package (Delorme & Makeig, 2004) with the `ERPlab` toolbox (Luck et al., 2010) in Matlab (Mathworks, 2014). Each electrode was re-referenced offline to a common average reference. The data were band-pass filtered between $0.01 - 30$ Hz and resampled at 256 Hz. See section 5.2.2 for more details on the preprocessing of the EEG signal.

6.1.2.2.5 Analysis

The data were analyzed with the non-parametric t_{max} permutation test (Groppe et al., 2011; Luck, 2014, see section 5.2.2 for more information on this statistical procedure) using the `Mass Univariate ERP Toolbox` (Groppe et al., 2011) in Matlab (Mathworks, 2014).

We were interested in modulations of the MMN as elicited by the presence/absence of IA and FA. Therefore, we specifically tested for differences in the time-windows between 201 — 301 ms and 551 — 651 ms, respectively. Furthermore, because the MMN is a fronto-centrally located deflection we specifically tested the Fz, Cz, FC1, FC2, C3, C4, F1, F2, F5h, F6h and FCz electrodes in both time-windows. Each comparison of interest was analyzed with a separate repeated measures, two-tailed t-tests, using the original data and 2500 random permutations to approximate the null distribution for the customary family-wise alpha (α) level of 0.05.

6.1.2.3 Results

Both —IA and —FA deviants elicited a MMN, although the MMN was smaller, and only marginally significant, when the deviant had been —IA (critical t-score: ±3.368, $p = 0.06$) compared to when the deviant had been —FA (critical t-score: ±3.4322, $p < 0.05$) (see figure 6.5). This difference is interpreted in the main discussion of both oddball studies below.

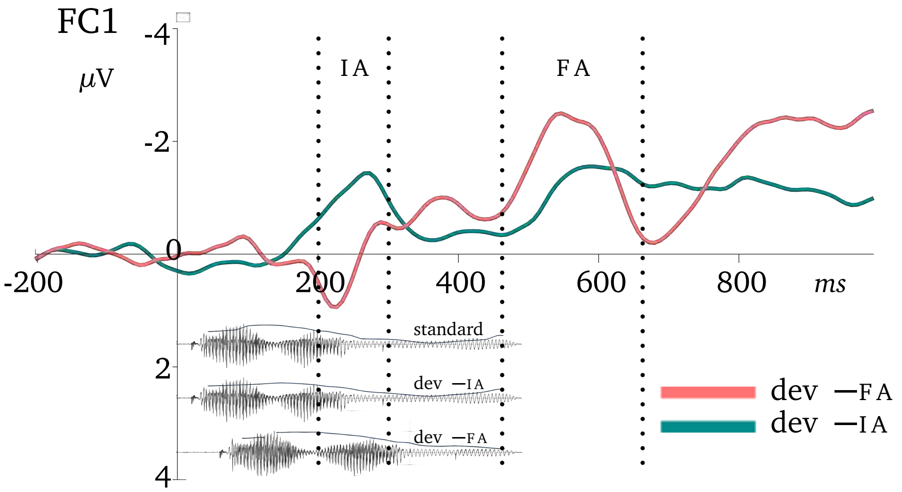


Figure 6.5: MMN components for —IA (in green) and —FA (in pink) deviants, recorded at the FC1 (left frontal) electrode, with the oscillogram of all stimuli type of [paʁadi] plotted in the background. Waveforms and oscillograms are temporally aligned to indicate the relation between the offset of the ±IA and ±FA manipulation and the resulting stimulus-locked event-related potentials. Tested time-windows are indicated by dashed vertical lines. For ease of presentation, ERP waveforms are low-pass filtered at 10 Hz and negativity is plotted as an upward deflection.

6.1.3 Discussion

In the current studies, we sought to investigate the phonological representation of French accentuation. We took advantage of the MMN component, which is held to index the strength of memory traces underlying phonological information. We based our expectations on the results presented in Aguilera et al. (2014), which had previously shown the French secondary initial accent (IA) to be encoded in long-term memory and to be expected by listeners.

We were first specifically interested in the representation of the French primary final accent (FA) and manipulated its presence on trisyllabic words in an auditory oddball paradigm. There were two versions of this paradigm; in one version the standards were presented with FA, while the deviants were presented without FA, and in the other version, positions were reversed. As we will discuss in more detail below, our results clearly showed a pre-attentive expectation for words to be presented with final accent and a general dispreference for words presented without the accent.

However, our results also partially deviated from those obtained in Aguilera et al. (2014), i.e. while the asymmetrical MMNs elicited by ±FA deviants are congruent to the results reported in Aguilera et al. (2014), the comparisons between participants differed. In order to better understand this deviance between IA and FA, in a follow-up study, we orthogonally manipulated the presence of both the final accent and the initial accent within the same paradigm. That is, in this second study, both −FA and −IA stimuli served as deviants, while the standard was consistently presented with both the final and initial accent. We obtained MMN difference waves to both −IA and −FA deviants. The amplitudes of the respective MMNs, however, differed in size, possibly reflecting a different functional role for the accents in their marking of the word.

Below, we will discuss our findings in turn: In section 6.1.3.1, we present the results obtained in the first oddball study that show FA to not only be readily perceived, but also to be expected by the listener and phonologically natural. In section 6.1.3.2, we discuss the differential processing of −FA and −IA deviants. We will interpret the results from an acoustic, exogenous point of view, as well as inspect the possibility for this difference to reflect more substantial, endogenous differences in the functions of the respective accents during word processing.

6.1.3.1 Phonological representation of the French final accent

In the first oddball study, wherein we had concentrated on the representation of the French final accent, we observed an asymmetry between MMNs elicited by deviants presented without final accent, compared to those elicited by deviants that had been presented with final accent. More specifically, the MMN was significantly more ample when the deviant had been presented −FA, than when it had been presented +FA (see figure 6.2). This asymmetry indicates that the final accent is encoded in long-term memory, where it underlies the representation of the word.

Our comparisons between participants corroborate with this interpretation (see figure 6.3). Presenting words without final accent elicited an ample ERP deflection, *irrespective* of the position of the stimuli within the experimental setting. That is, words without final accent appeared to require more cognitive effort, *regardless* of whether they had been the standards or the deviants in the oddball paradigm. This result shows stress templates without FA to be generally disfavored. Indeed, if there had been no preference for one stress pattern over the other, then repeatedly presenting words without final accent (i.e. when −FA is in standard position) *should* have made the stress pattern −FA the probable stress template. Clearly, it did not; even in standard position, the stress pattern without final accent remained unexpected. In other words, listeners continued to anticipate words to be marked with final accent, most likely due to its established phonological representation.

The comparison between standards and deviants presented with final accent points to the same conclusion. In this comparison, position within the experimental setting *did* matter. Recall that the MMN may reflect both a prediction error when anticipations based on established phonological representations are violated, as well as a mismatch within the experimental context. In the comparison between words +FA, interrupting a train of −FA stimuli with the sudden presentation of a template with FA, elicited a small prediction error, while no such prediction error followed the final accent when +FA stimuli were in the standard position. This finding again disproves the notion of stress deafness, i.e. listeners *readily notice* the accent when deviants +FA contrasted to the train of −FA templates. In other words, listeners detected FA when it mismatched the short-term anticipation established by the repeated stress templates −FA, negating their alleged phonological deafness. However, as was explained above, the mismatch did *not* result in a significant MMN (far from it, see figure 6.2), because presenting stimuli without final accent, even when congruent within the experimental setting (i.e. when −FA was in the standard position), remained unexpected due to the long-term phonological representation of the final accent.

In sum, we show that the final accent is readily perceived and elicits a small prediction error when it mismatches short-term memory, while stress patterns without final accent mismatch both short- *and* long-term memory representations and are thus not the expected metrical pattern in French.

6.1.3.2 Differential processing between the initial and final accents

While the asymmetrical MMNs elicited by ±FA deviants are congruent to the results reported in Aguilera et al. (2014), the comparisons between participants differed. Where Aguilera and colleagues obtained a bigger difference wave after words were presented without IA underlying their stress template, even when comparing acoustically identical stimuli in both standard and deviant position, we obtained results opposite to that (i.e. there was a bigger difference between +FA stimuli than between −FA stimuli). To better understand this incongruence, in a follow-up study, we orthogonally manipulated the presence of both the final accent and the initial accent within the same paradigm. That is, in this second study, both −FA and −IA stimuli served as deviants, while the standard was consistently presented with both the final and the initial accent.

We obtained two consecutive MMN deflections, one reflecting the absence of IA, the other reflecting the absence of FA (see figure 6.5). The amplitudes of the MMNs were, however, different in size, with the MMN following deviants −FA being more ample than the MMN following deviants −IA. These results could inform us about differences in the strength of the memory representations between IA and FA, with the final accent holding a stronger memory trace and being anticipated to a greater extent by listeners than the initial accent. However, there are several alternative explanations which are also compatible, and, possibly, more likely explain the different MMNs: one reflecting a purely exogenous, acoustic interpretation, and the other involving a more substantial, endogenous difference in the accents' respective functions during speech processing. Both account are discussed below.

6.1.3.2.1 Exogenous interpretation

In the exogenous interpretation, the dissimilar MMN amplitudes between IA stimuli and FA stimuli reflect differences in acoustic processing. Indeed, the acoustic manipulations had not been the same between our ±IA and ±FA stimuli, the former involving exclusively a manipulation of the f_0 rise, and the latter involving mainly a durational change (see section 5.2). It is possible that French listeners are more sensitive to durational changes than to changes in pitch movement (see e.g. Partanen et al., 2011, for an MMN study showing just that for Finnish speakers, although also note that sensitivity to stress phonetic features is likely language specific).

Moreover, while the presence of IA was *only* manipulated in f_0, the durational change of FA led to the additional disappearance of the accent's final rise (see figure 6.4), the secondary phonetic characteristic of FA (see table 1.3). This means that stimuli without FA differed from stimuli with FA on two acoustic parameters, while ±IA stimuli differed only in f_0. Because MMN amplitudes are held to reflect the *magnitude* of the deviance between standard stimuli and deviants (Sussman, 2007; Näätänen et al., 2007; Sussman et al., 2014; Sussman & Shafer, 2014), these exogenous interpretations may at least in part explain the observed MMN differences between our −IA and −FA stimuli.

However, a purely acoustic interpretation less straightforwardly accounts for the different findings in the between participants comparisons observed in the current study versus those presented in Aguilera et al. (2014). Therefore, we consider it more likely that the dissimilar amplitudes reflect different respective roles for the accents during speech processing. Indeed, while the initial accent sits at the left word boundary and is argued, in the current thesis, to signal word onsets and cue listeners on when to initiate lexical access, the final accent, which is located at the right word boundary, likely holds different functions, such as marking the word's offset and cue listeners on when to finalize their analysis of the word. In this view, the respective MMNs then reflect different interactions between the accents and the stages in speech perception, which we will turn to next.

6.1.3.2.2 Endogenous interpretation

Speech perception unfolds in three stages: an acoustic stage, during which the speech signal is spectrally decomposed and distinguished from non-speech sounds, a pre-lexical stage, during which phonological information is assembled and matching lexical candidates are activated, and, finally, the lexical stage, wherein candidates compete and are evaluated up until one word can be selected for word recognition. In our view, the initial accent is more likely to interplay with the pre-lexical stage during which lexical hypotheses are derived and activated, while the final accent will presumable be more involved in the later lexical stage which ends in the recognition of the word. In terms of the Cohort model (Marslen-Wilson & Welsh, 1978; Wilson, 1990, see chapter 2), the initial accent (the word's earliest phonological information) activates similar lexical representations into the, so-called, cohort. As the speech signal continues, matching candidates are additionally activated while, when words without final accents seize to match the activated representations, these are disregarded from the cohort or lessened in activation levels. In other words, the initial accent plausibly has more effect on the *start* of the process of word recognition and on early lexical activation levels, while the final accent is more likely involved in the *outcome* or *wrap-up* of the lexical competition.

Note that, in this view, dissimilar MMNs elicited by ±IA versus ±FA stimuli are not only explained in terms of different interactions during the process of word recognition, but also in terms of the precision of the prediction to which the stress patterns are compared. Recall that according to the theory of predictive coding, predictions which are precise require less additional cognitive effort than predictions which more generic. Stimuli without IA differ from the prediction phonologically, i.e. the listener has a general phonological preference or expectation for words to be presented with both IA and FA in their underlying stress templates. When the deviance is however later in the word, as with −FA, the listener's prediction more pointedly concerns the phonological stress template marking the right boundary of the particular lexical item expected from the train of standards.

That is, one can imagine that, if FA cues the lexical offset, listeners could have imagined words without final accent to be part of, or embedded in, a longer word, therefore deleting the anticipated word boundary. Indeed, as presented in chapter 2, the segmentation problem is closely related to the embedding problem which holds that words can have other words partially or wholly embedded within them, such that the speech stream usually matches with multiple lexical candidates (e.g. 'paradis' is a word on its own, but can also be at the onset of, for example, 'paradisiaque' or 'paradigmatique'). When presented stress patterns mismatch the expected stress template, this can lead to wrongfully deleting a word boundary, similar to what was found in the juncture misperception studies presented in section 1.4, wherein English and Dutch listeners erroneously inserted a word boundary when encountering a strong syllable (for instance, "analogy" → "an allergy") or deleted a word boundary before a weak syllable (for instance, "my gorge is" → "my gorgeous") (e.g. Cutler & Butterfield, 1992; Vroomen et al., 1996).

In fact, French listeners have been found to segment speech on FA in ambiguous sentences (see e.g. Banel & Bacri, 1994; Bagou et al., 2002; Christophe et al., 2004, for studies wherein FA signaled the right phrase boundary, which, if one recalls the Strict Layer Hypothesis presented in section 1.2, coincides with the right word boundary). For example, Banel & Bacri (1994) found listeners to use the lengthened syllables as a right boundary cue and, consequently, segmented immediately after them. That is, when listeners were asked to interpret ambiguous speech sounds such as [bagaʒ] which may be segmented into two words 'bas + gage' (low + pledge) or can be interpreted as 'baggage' (luggage), listeners favored the former interpretation when the syllables were marked with a trochaic stress template (long—short), while conversely favoring the latter interpretation when the stress template had been iambic (short—long). That is, lengthened syllables encouraged a boundary to the right, while short syllables did not. Because, in French, prosodic descriptions do not include the lexical word, the boundary was attributed to the phrasal domain. However, note here that FA might have also cued the right lexical boundary in that study.

Similarly, in the study on the interaction between metrical structure and semantic processing, Magne et al. (2007) artificially lengthened the medial syllable. This metrical 'incongruity' was found to obstruct semantic processing, possibly because listeners segmented speech on the medial syllable and, thus, before the word's actual offset. In the current study, shortening the final syllable in the deviant position, may have led the deviant to not only mismatch with the anticipated phonological stress template, but change the predicted lexical item because it was missing its right boundary mark (e.g. "paradis" → "paradigmatique"). That is, listeners may have notices the acoustic mismatch (i.e. syllable length and f_0 movement), the phonological incongruence (i.e. ±FA), and the lexical difference ('paradis' → 'paradigmatique'). In other words, repeatedly presenting the same lexical item in the standard position, led to more specific anticipations and activations of lexical candidates, which, in turn, resulted in MMNs reflecting the concurrent detection of several deviances: (1) the acoustic deviance, (2) the phonological mismatch and, possibly, (3) the mismatch to the lexical prediction (see e.g. Pulvermüller & Shtyrov, 2006; Jacobsen et al., 2004; Honbolygó et al., 2004; Honbolygó & Csépe, 2013; Honbolygó et al., 2017; Ylinen et al., 2009; Garami et al., 2017; Zora et al., 2016, for oddball studies investigating obstructed processing due to mismatching stress templates on words and/or pseudowords in Hungarian, Finnish and Swedish).

However, if the differences between IA and FA reflect interactions with different stages during word recognition, then, while interesting, the oddball paradigm (and MMN) unfortunately is not well suited to observe them. Clearly, oddball paradigms provide a rather artificial listening situation, wherein it is not clear whether each word presentation (whether in standard position or as deviant) encourages a fresh attempt to lexical access. That is, arguably the repeated presentation of the same word may involve a process different from normal listening situations wherein listeners go through all three stages of word recognition. This problem is addressed in the next section, wherein two studies are presented employing the lexical decision paradigm. In a lexical decision paradigm, listeners have to decide whether a word is a real word or a pseudoword, forcing them to actively initiate and complete the process of lexical access. In the first of our two lexical decision studies, the presence of IA is manipulated both on lexical words and on pseudowords, while, in the other, it is the presence of FA which is manipulated, allowing us to better observe their individual contributions to the process of word recognition.

IN CONCLUSION, in this oddball study, we investigated the cognitive representation of the French final stress. Indeed, the French initial accent had previously been shown to not only be readily perceived but expected by French listeners as part of the stress pattern underlying the lexical word, indicative of a functional role in their analysis of speech. The results of the present study how that FA—just as IA—is not only perceived, but anticipated by listeners as belonging to the abstract representation of the word. Unlike the results reported in Aguilera et al. (2014), when the standard was presented without FA, it remained unexpected, despite its high probability within the experimental context. This result suggests that the deviant without FA remained improbable within the experimental setting, indicating a long-term representation of the accent and underlining listeners' expectation for words to be marked by stress templates which also include FA.

Moreover, we observed an asymmetry between deviants presented with FA and deviants presented without, with larger MMN amplitudes when the deviant had been presented without FA. In this respect, the results are congruent to the asymmetrical MMNs reported in Aguilera et al. (2014) in which IA had been manipulated, and, together, the results are in line with Di Cristo's model, and demonstrate a cognitive, phonological expectation for metrically strong syllables at both left and right lexical boundaries. Altogether, the results contradict the traditionally accepted view of French as a language without accent and, instead, suggest accentuation to have a functional role in word level processing.

6.2 Metrical stress in word recognition

As was discussed in chapter 2, the ability to understand spoken language is a fundamental and intriguing human skill, with word recognition being one of the hardest processes to model computationally. Word recognition unfolds in three stages, an auditory, pre-lexical and lexical stage. During these three stages, the speech system must solve two challenges: mapping variable sound structures to their canonical representation and detecting the boundaries between individual speech units.

The first problem is partially solved through abstraction (e.g. Eulitz & Lahiri, 2004; Cutler, 2010). This means that phonological units are encoded in the form of abstract, cognitive templates to which variable speech sounds can be generalized. In the previous section, we argued that the stress patterns containing IA and FA are an example of such an abstract, phonological template. The second major problem is partially (i.e. together with, for instance, phonotactic and contextual cues) solved by relying on prosodic or metrical information to cue word boundaries and indicate to the listener on when to initiate lexical access (e.g. MSS and ABH; Cutler, 1990; Pitt & Samuel, 1990, presented in section 1.4).

Indeed, according to the Metrical Segmentation Strategy (MSS), the segmentation of continuous speech is accomplished by relying on the dominant metrical pattern of the language (Cutler & Norris, 1988). In stress based languages such as English and Dutch, where the vast majority of lexical words start with a strong syllable (Cutler & Carter, 1987; Vroomen & de Gelder, 1995), listeners are thought to exploit that high prosodic probability and initiate lexical access at each stressed syllable. But, while this may be a successful strategy in languages with lexical stress, segmenting on strong onsets is arguably much less efficient in languages in which the domain for metrical rules is not the lexical word.

French is often described as a syllable based language with fairly homogeneous metrical weight on syllables. Consequently, it is held that the French metrical structure is defined by the syllable and that the syllable is used as the basic unit for segmenting speech (Mehler et al., 1981; Cutler et al., 1986; Content et al., 2001a; Dumay et al., 2002, see also section 2.3). This idea is further supported by the view that, in French, accentuation is post-lexical, demarcating boundaries not at the level of the word but at the level of groups of words. Recall from chapter 1, that the primary French accent, FA, is fixed on the last syllable of the phrase, marking its right edge, accompanied, when necessary, (e.g. in case of long stretches of unaccented syllables), by secondary IA that marks the left boundary of the phrase.

However, in this thesis, we assume both IA and FA to be phonologically encoded and metrically strong at the level of the word. This means that French accentuation provides not one, but two lexical entries; at the left boundary and at the right boundary of the word, and, as such, may also play a vital role in the solution to this second problem and help listeners during the processes of speech segmentation and word recognition. Indeed, both primary FA *and* secondary IA have been found to guide French listeners in the segmentation of speech (e.g. Rolland & Lœvenbruck 2002; and, for use of FA, see Banel & Bacri 1994; Bagou et al. 2002; for use of IA, see Welby 2007; Spinelli et al. 2007, 2010). These studies challenge the idea that French listeners adopt a syllable based segmentation strategy, instead favoring a strategy in which listeners rely on metrical stress patterns during speech comprehension. Note, however, that they do not challenge the view that IA and FA demarcate phrase boundaries, and still consider accentuation to apply to the level of the Accentual Phrase (AP; Jun & Fougeron, 2000) and not to the level of the word.

In the previous section, wherein we presented the results of three oddball studies, we showed asymmetric MMNs that indicate both IA and FA to be phonologically encoded and anticipated by the listener, in line with Di Cristo's conjecture of (latent) stress templates underlying the representation of the lexical word. If IA and FA are cognitively encoded at the level of the word, then they can be expected to interact with and facilitate lexical processing. We however also proposed that the roles may differ and that the accents may interact with different stages during lexical access. In the current section, we present two studies in which the representation of French accentuation is further investigated, specifically regarding its contribution to word processing. In the studies, we are particularly interested in modulations of the N325, a component assumed to reflect difficulties in the extraction of lexical stress (Böcker et al., 1999, see also section 3.2.2). Recall, that this component was first encountered in a study in which the authors presented Dutch participants with sequences of four bisyllabic Dutch words, which they either passively attended, or actively discriminated between. The words were either trochaically stressed (the Dutch dominant template), or iambically. Results showed that the less frequent stress template elicited a larger frontal negativity (the N325) than did the dominant stress template, particularly in the discrimination task. This led the authors to conclude that the N325 may reflect difficulties in the extraction of metrical stress during lexical access.

Indeed, Böcker and colleagues disambiguate the N325 from two related ERP components. That is, the N325 is held separate from the N400 because it has a different spatial distribution (fronto-central as opposed to the more centro-parietal distribution of the N400, see also section 3.2). Furthermore, while the N325 is more similar to the PMN (Phonological Mapping Negativity, again see section 3.2) a component linked to pre-lexical, phonological processing (Connolly & Phillips, 1994; Kujala et al., 2004; Newman & Connolly, 2009) with a similar temporal and spatial distribution, also elicited when phonological information violates the subject's phonological expectations, the PMN is dependent on task contextual congruency, while the N325 is more sensitive to the overall probability of stress patterns. That is, Böcker and colleagues obtained a bigger N325 after presenting a word with the infrequent stress pattern, even when the word was congruent in the task, which differs from the typical elicitation of the PMN and makes the N325 exceptionally well suited for the purpose of our investigation.

In the first study, we manipulated the presence of IA on lexical words and pseudowords in a lexical decision task. We expect that, if IA is linked to the phonological representation of words, and is, along with FA, the expected stress template in French, presenting words without IA should elicit a larger N325 than presenting words with IA. Further, if IA is pre-lexical and interacts predominantly with this early stage in word recognition, then we expect similar differences in N325 amplitude between nouns and pseudowords. In the second study, we presented listeners with the same task, only now not the presence of IA but the presence of FA had been manipulated. For FA, we expect similar results as for IA, except that, if, as we argued in the previous section, FA should interact with later stages in lexical access, then in this study we should observe an interaction between presence of FA and lexicality. In sum, we expect both accents to apply to the lexical domain, but to, non-redundantly, display different respective functions in their roles during lexical processing.

6.2.1 Lexical decision study: Initial Accent

6.2.1.1 Study Summary

Procedure:

- Nr participants → 23
- Nr stimuli/condition → 120
- Task → lexical decision

Preprocessing:

- Nr electrodes → 32
- Reference → Mastoids
- Filter and down-sampling → 0.01 − 30 bandpass, 125 Hz
- Epoch length → −200 − 2000

Analysis:

- Behavioral → Linear Mixed Effects Model in R (DV: RT; IV: IA/lex)
- eeg → $t_{\max}$ univariate permutation test in Matlab, 2500 permutations (DV: amplitude; IV: ±ia/lex)

p2:

- Time-window → 151 − 251 ms
- Electrodes → Fz, Cz, FC1, FC2, F3, F4, C3, C4

n325:

- Time-window → 201 − 431 ms
- Electrodes → Fz, Cz, FC1, FC2, F3, F4, C3, C4

Results:

- Behavioral → effect lexicality
- erp → effect lexicality for p2
- erp → effect ±ia for n325

Research Question: *Is the initial accent expected as underlying the representation of words and is it treated predominantly pre-lexically or does it also intervene in later stages in lexical access?*

Procedure: 23 *participants completed a lexical decision task, wherein lexicality (real French nouns versus pseudowords) and metrical stress template (* +IA *vs* −IA*) were orthogonally manipulated on trisyllabic stimuli.*

Results: *We obtained a main effect of* ±IA *on the amplitude of the* N325 *components, such that stimuli* −IA *elicited a larger* N325 *than stimuli* +IA*. Additionally, we obtained an effect of lexicality on the* P2 *component, which will be argued to reflect a temporal overlap between the* N325 *and the* P2*, suggesting our metrical manipulation affected processing early on during word recognition.*

Conclusion: *The data reveal both the automaticity of pre-lexical stress extraction and a preference for stress templates with initial accent during the early stages of lexical access.*

6.2.1.2 Methods: Initial Accent

6.2.1.2.1 Participants

26 French native speakers, aged 19 – 31 (mean age 25.4; 20 female), gave their informed consent and volunteered to take part in the study which was conducted in accordance with the Declaration of Helsinki. None of the subjects had taken part in the previous oddball studies, and all were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Three subjects were excluded from the EEG-analysis due to excessive artifacts in the signal.

6.2.1.2.2 Speech stimuli

The stimuli consisted of 120 trisyllabic French nouns (e.g. *chocolat* [ʃɔkɔla], 'chocolate') and 120 trisyllabic pseudowords (e.g. *chibuté* [ʃibyte]), extracted from sentences spoken by a naïve native speaker of French. In the sentences, the target words (lexical word or pseudoword) were placed at the beginning of a major phrase to increase the probability of clear IA and FA marking (Astésano et al., 2007, see section 5.1 for more information on the construction of the original corpus). A panel of three experts selected the stimuli with the most natural IA (+IA).

To create the condition without IA (−IA), target words were re-synthesized using a customized quadratic algorithm in PRAAT (Boersma & Weenink, 2016). The algorithm was similar to the one used in Aguilera et al. (2014) and is described in detail in section 5.1.2. The algorithm progressively lowered the f_0 value of the first vowel (i.e. IA) towards the f_0 value of the preceding (unaccented) determinant. This quadratic transformation allowed for micro-prosodic variations to be maintained, thus keeping the natural sound of the stimuli. Furthermore, in order to equalize the speech quality between +IA and −IA stimuli, +IA stimuli were forward and back transformed (see figure 6.6 for an illustration of the f_0 manipulation and table 6.3 for more information on mean syllable and word durations and f_0 values for each of the four conditions).

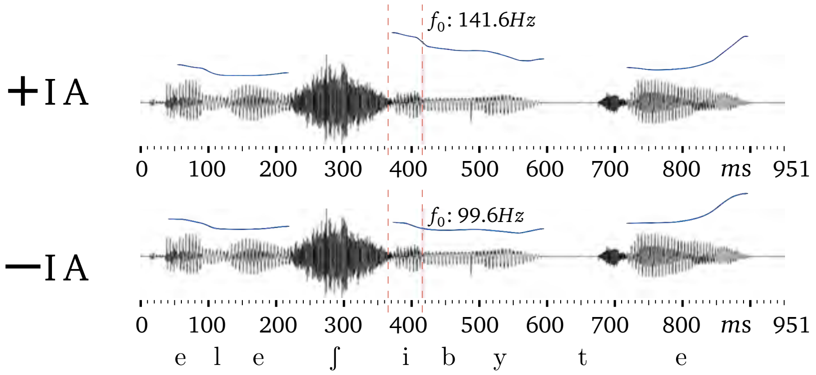


Figure 6.6: Example of f_0 resynthesis +IA (top) and −IA (bottom) of the stimulus ʃibyte (chibuté), with quadratic interpolation from the f_0 value of the preceding determinant to the f_0 value at the beginning of the last stressed syllable for −IA targets.

Table 6.3: Overview of mean stimulus properties for both lexical words and pseudowords ±IA (total duration, total first syllable and syllable-vowel durations, and first syllable-vowel and determinant-vowel f_0 values).

	Total duration		1st syllable ms		1st vowel ms		1st vowel f_0		Det vowel f_0	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
WORD										
+IA	647.7	62.5	176.6	36.1	79.0	22.4	263.1	19.2	199.4	11.7
−IA	647.7	62.5	176.2	35.8	79.0	22.4	217.6	13.4	196.7	11.8
PSEUDOWORD										
+IA	658.6	44.8	169.6	25.3	75.2	18.2	272.7	28.9	197.5	29.5
−IA	658.6	44.8	169.6	25.3	75.2	18.2	217.4	10.5	196.3	8.5

6.2.1.2.3 Procedure

Each participant was comfortably seated in an electrically shielded and sound attenuated room. Stimuli were presented through headphones using E-prime on a Windows XP 32-bit platform. Participants were allowed to adjust the volume to their individual preferences. They were instructed to judge as quickly and accurately as possible whether a word was a real word or a pseudoword by pressing the left or right button on a button-box. Button assignment was counter-balanced across participants.

To ensure participants understood the task requirements, the experiment began with a short practice phase. This phase consisted of 12 trials that were similar to the experimental trials, but not included in the analyses. Each participant listened to all 240 stimuli (60 words +IA, 60 words −IA, 60 pseudowords +IA and 60 pseudowords −IA), which were evenly distributed over four blocks, with block order balanced across participants.

In order to better control for eye-related EEG activity, each trial started with a 400 ms presentation of a white fixation cross at the center of a computer screen. The stimulus was presented immediately after the offset of the fixation cross. Participants were given a maximum of 2000 ms to give their answer. The ISI followed the participant's response and lasted until 2500 ms post stimulus onset. As a result, the duration of the ISI varied, while trial duration was fixed at $400 + 2500 = 2900$ ms. Total duration of the experiment, including the set-up of the EEG electrodes, was approximately 2 hours.

6.2.1.2.4 EEG recording and preprocessing

The EEG data were recorded with 32 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958). To detect blinks and eye-movements, four additional electrodes were placed around the eyes (HEOG: bipolar channel placed lateral to the outer corner of both eyes; VEOG: bipolar channel placed above and below the left eye). The EEG and EOG signals were amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 512 Hz.

The data were preprocessed using the `EEGlab` package (Delorme & Makeig, 2004) with the `ERPlab` toolbox (Luck et al., 2010) in Matlab (Mathworks, 2014). Each electrode was re-referenced offline to the algebraic average of the left and right mastoids. The data were band-pass filtered between $0.01 - 30$ Hz and epoched from -0.2 to 2 seconds surrounding the onset of the speech signal. Following a visual inspection, epochs containing EMG or other artifacts not related to eye-movements or blinks were manually removed. ICA was performed on the remaining epochs in order to identify and subtract components containing oculomotor artifacts from the data. Finally, data were averaged within and across participants to obtain the grand-averages for each of the four conditions (word +IA, word −IA, pseudoword +IA, pseudoword −IA). See section 5.2.2 for more details on the preprocessing of the EEG signal.

6.2.1.2.5 Analysis

Behavioral Response latencies were analyzed with mixed effects linear regression in R (Team, 2014) using the `lme4` package (Bates et al., 2012). The analysis tested whether lexicality and/or presence of IA influenced reaction times. That is, the model outcome was reaction times and the predictors were lexicality and presence of IA. Furthermore, listeners and stimuli were included in the model as random variables in order to control across-listeners and across-stimuli variability independent from our predictions. More specifically, for the random structure, we found intercepts for listeners and stimuli, as well as by-stimuli random slopes for the effects of metrical pattern and lexicality best accounted for underlying random variability.

Visual inspection of residual plots did not reveal any obvious deviations from homoskedasticity or normality (see appendix C.3). Likelihood ratio tests of the model with the effect in question against the model without the effect in question were used to obtain p-values. For more details on the workings of mixed model analysis, see section 5.2.1.

EEG The ERP data are analyzed using a non-parametric t_{max} permutation test (Blair & Karniski, 1993, see also Groppe et al. 2011; Luck 2014 and section 5.2.2 for a detailed description of this statistical test). In t_{max} permutation testing, the null distribution is estimated by repeatedly resampling the obtained data and calculating t-scores for each sample. As such, the test does not assume test independence, allowing for stringent control of Type I error without considerable decrease in sensitivity.

To further maximize power and reduce the number of comparisons, the data were down-sampled to 125 Hz and 13 out of the 32 electrodes were selected for testing (i.e. Fz, Cz, FC1, FC2, F3, F4, C3, C4, Pz, P3, P4, CP1, CP2), excluding for instance the occipital region more involved in visual processing. Finally, because we were mainly interested in modulations of the P2 and the N325, two time-windows were estimated following the method described in Böcker et al. (1999). For the P2, a 100 ms time-window was draped around the average peak latency at electrodes C3 and C4, while for the N325, the time-window was defined as the period between the peak latencies of the pre-lexical P2 and post-lexical 'N400'. This procedure resulted in a time window of $151-251$ for the P2 (for comparison, in Böcker and colleagues, the procedure led to a time-window of $171-271$ for the P2) and a time-window of $201-431$ for the N325 (in Böcker and colleagues this window was calculated at $221-431$).

Each comparison of interest was analyzed with a separate repeated measures, two-tailed t-test, using the original data and 2500 random permutations to approximate the null distribution for the customary family-wise alpha (α) level of 0.05.

6.2.1.3 Results

6.2.1.3.1 Behavioral results

Overall, performance on the lexical decision task revealed high accuracy (< 5% errors) with no differences between conditions, indicating listeners understood the task requirements and had little difficulty disambiguating pseudowords from lexical words.

The reaction times however did show that listeners were slower to respond to pseudowords than to lexical words; figure 6.7 and table 6.4 show that lexicality clearly affected response latencies with participants responding slower to pseudowords than to lexical words. The regression analysis confirms this effect and is significant at $p < 0.001$ ($\beta = 78.25,\ SE = 6.99,\ t = 11.2$, see table 6.5 for an overview of model fit).

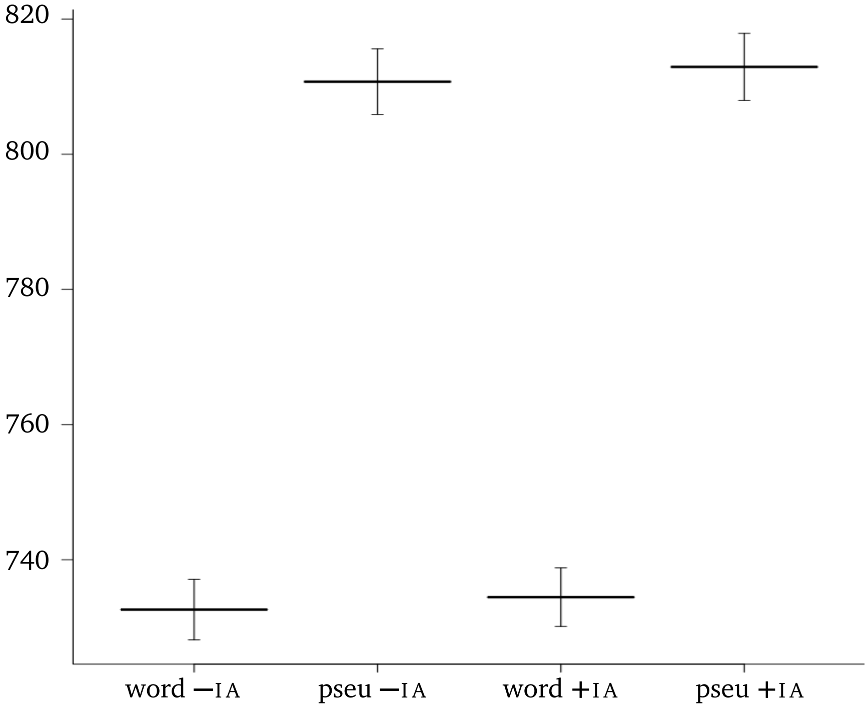


Figure 6.7: Error-bar plot of mean reaction times for all four conditions (word −IA, pseudoword −IA, word +IA, pseudoword +IA) showing a clear significant effect of lexicality and no effect of presence of initial accent, nor an interaction between both manipulations.

IA was also expected to facilitate lexical retrieval, however presence of IA had no impact on reaction times ($\beta = -1.82,\ SE = 8.64,\ t = -0.21,\ p = 0.83$, *ns*), nor did it interact with lexicality ($\beta = -0.5,\ SE = 14.03,\ t =$

−0.03, $p = 0.97$, *ns*).

Table 6.4: Reaction times per condition. Data analysis revealed a significant effect of lexicality, but no effect of ±IA and no interaction.

Response latencies	*m*	*sd*	maximum	minimum	% errors
WORD					
−IA	732.63	165.43	1774	325	1.2%
+IA	734.47	160.26	1679	306	0.5%
PSEUDOWORD					
−IA	810.73	179.13	1793	358	0.9%
+IA	812.92	182.79	1754	436	1.5%

6.2.1.3.2 ERP results

Table 6.5: Overview linear mixed models. The model fitting the data best takes lexicality as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent did not significantly contribute to the prediction of reaction times.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	772.37***	772.72***	655.34***	775.46***	658.15***	657.02***
	(17.61)	(17.99)	(20.67)	(22.17)	(23.19)	(39.18)
lexicality			78.25***		78.25***	79.00***
			(6.99)		(7.00)	(22.19)
±IA				−1.83	−1.87	−1.12
				(8.64)	(7.00)	(22.18)
±IA:lexicality						−0.50
						(14.03)
AIC	70588.80	70131.37	70026.59	70127.17	70022.79	70017.67
BIC	70608.61	70177.58	70079.41	70179.99	70082.22	70083.70
Log Likelihood	−35291.40	−35058.68	−35005.30	−35055.59	−35002.40	−34998.84
Num. obs.	5447	5447	5447	5447	5447	5447
Num. groups: subject	23	23	23	23	23	23
Var: subject (Intercept)	7027.58	7014.83	7017.44	7014.79	7017.39	7017.33
Var: Residual	24438.60	20923.99	20926.03	20923.97	20926.01	20925.98
Num. groups: stimuli		240	240	240	240	240
Num. groups: lex:stimuli		240	240	240	240	240
Num. groups: ±IA:stimuli		240	240	240	240	240
Var: stimuli (Intercept)		292.62	40.37	1016.38	173.44	210.20
Var: lex.stimuli (Intercept)		0.08	1126.62	146.17	1365.94	345.59
Var: ±IA.stimuli (Intercept)		0.05	787.50	919.59	345.11	1405.91

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

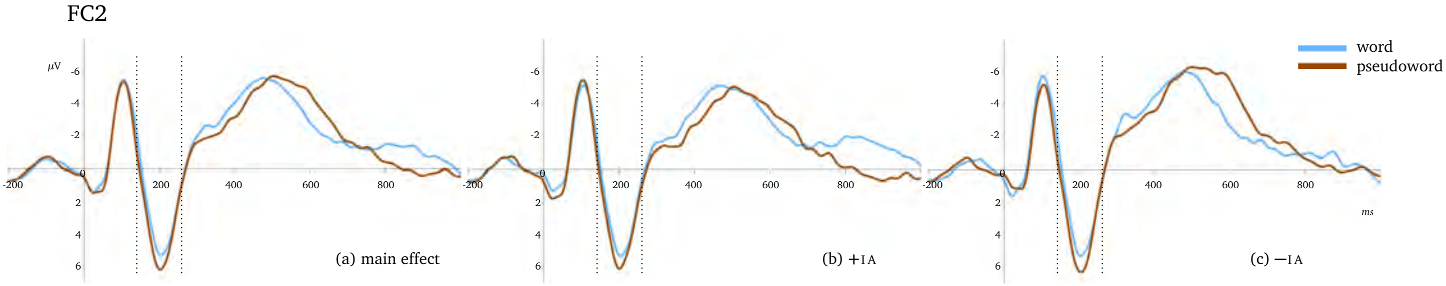


Figure 6.8: Grand average P2 in the lexical condition (pseudoword-condition in brown, word-condition in blue), recorded at the FC2 (fronto-central) electrode for: (a) main effect, (b) +IA, (c) −IA. The tested time-window is indicated by dashed vertical lines. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are low-pass filtered at 10 Hz.

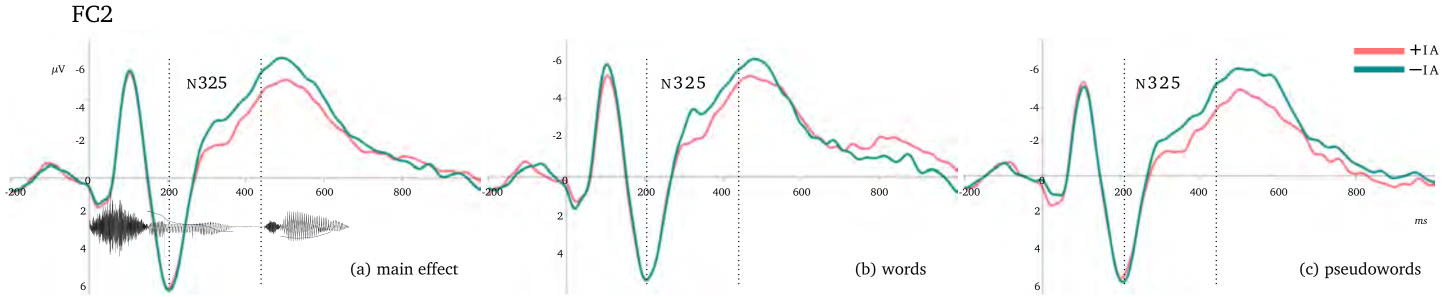
P2 As expected, presence of IA had no effect on P2 amplitude ($p = 0.77$, *ns*). Surprisingly however, lexicality did modulate P2 amplitude (critical t-score: ±3.5589, $df = 22$, $p < 0.05$); pseudowords elicited a larger P2 than words in the fronto-central region (FC2) peaking 182 ms after stimulus presentation (see figure 6.8).

The difference between words and pseudowords was also significant within the condition −IA (critical t-score: ±3.575, $df = 22$, $p < 0.05$). Within the condition +IA, this effect was not significant ($p = 0.4$, *ns*).

N325 In the N325 time-window, there was a main effect of presence of IA (critical t-score: ±3.6887, $df = 22$, $p < 0.05$). Compared to stimuli +IA, stimuli −IA elicited a larger negativity in the fronto-central region (FC2 and Cz) from $318 - 358$ ms after stimulus presentation (see figure 6.9). This indicates that processing of the stimuli was more demanding when there was no initial accent.

The effect was also significant within the lexical words condition (critical t-score: ±3.8546, $p < 0.05$); words −IA resulted in a larger negativity than words +IA. There was a similar trend in the pseudowords condition, but there was no interaction with lexicality. Lexicality did not modulate the N325 ($p = 0.14$, *ns*).

We will return to these results in the discussion of both lexical decision studies at the end of this section, at which point the results of the second study, wherein the presence of FA is manipulated, help interpret the results presented here.



(a) main effect (b) words (c) pseudowords

Figure 6.9: Grand average N325 in the ±IA condition (−IA-condition in green, +IA-condition in pink), recorded at the FC2 (fronto-central, bottom) electrode for: (a) main effect, (b) words, (c) pseudowords. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillogram and f_0 deflection of [ʃibyte] (*chibuté*) +IA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are lowpass filtered at 10 Hz.

6.2.2 Lexical decision study: Final Accent

6.2.2.1 Study Summary

Research Question: *Is the French final accent expected as underlying the representation of words, similar as the initial accent, or is it treated predominantly during the later stages in lexical access?*

Procedure: 20 *participants completed a lexical decision task, wherein lexicality (real French nouns versus pseudowords) and metrical stress template (* +FA *vs* −FA*) were orthogonally manipulated on trisyllabic stimuli.*

Results: *We obtained a main effect of* ±FA *on the amplitude of the* N325 *components, such that stimuli* −FA *elicited a larger* N325 *than stimuli* +FA*. Lexicality also modulated* N325 *amplitude, but only when stimuli had been presented with* FA*; with pseudowords eliciting a bigger* N325 *than words.*

Conclusion: *The* ERP *results reveal an interaction between the presence of the French final accent and lexicality. While pseudoword hindered processing whether or not they were presented with final accent, only words* −FA *obstructed lexical access. Listeners thus had a lexical expectation for the final accent, i.e.* FA *is natural to the listener and facilitates lexical access.*

6.2.2.2 Methods: Final Accent

6.2.2.2.1 Participants

20 French native speakers, aged 19 – 45 (mean age 23.7; 14 female), took part in the study. None of the participants had taken part in previous lexical decision study and all were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Each of the participants gave their written consent and was paid a small fee for their participation. The study was conducted in accordance with the Declaration of Helsinki.

6.2.2.2.2 Speech stimuli

Stimuli were selected and adapted from the corpus used in the previous lexical decision study and consisted of trisyllabic lexical nouns and pseudowords that were phonologically similar to the lexical nouns but had no lexical content in French. As is shown in figure 6.10 and table 6.6, −FA stimuli were created by shortening the duration of the third syllable (FA) and making sure it did not end in a final rise of f_0 (the two main phonetic signatures of FA, see table 1.3). This meant that words and pseudowords ending with a consonant or schwa, starting with an long consonant compared to the final vowel, or containing a nasal vowel were eliminated from the current corpus (see section 5.1.3 for more information).

The resulting corpus consisted of 80 trisyllabic French nouns and 80 trisyllabic pseudowords.

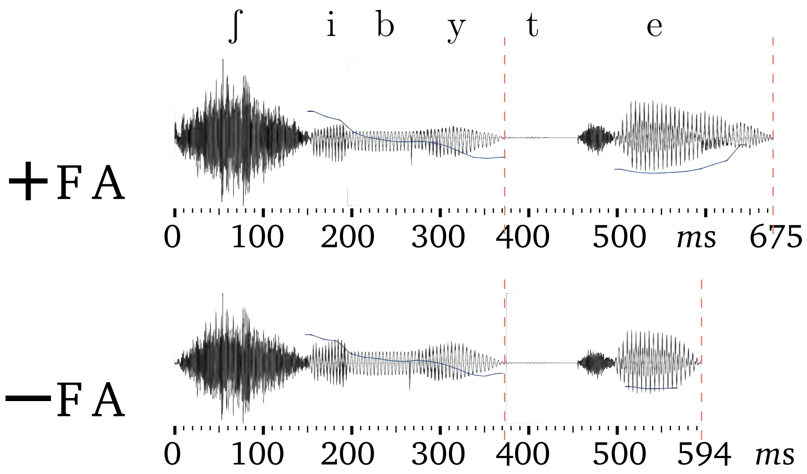


Figure 6.10: Example of duration manipulation +FA (top) and −FA (bottom) of the stimulus ʃibyte (chibuté). The duration of the final syllable is equal to the duration of the unaccented second syllable for the −FA targets.

Procedure:

- Nr participants → 20
- Nr stimuli/condition → 80
- Task → lexical decision

Preprocessing:

- Nr electrodes → 64
- Reference → Average
- Filter and down-sampling → 0.01 – 30 bandpass, 128Hz
- Epoch length → −200 – 1500

Analysis:

- Behavioral → Linear Mixed Effects Model in R (DV: RT; IV: fa/lex)
- eeg → $t_{\max}$ univariate permutation test in Matlab, 2500 permutations (DV: amplitude; IV: ±fa/lex)

p2:

- Time-window → 151 – 251 ms
- Electrodes → Fz, Cz, FC1, FC2, CPz, AFz, Fpz, F1, F2

n325:

- Time-window → 546 – 776 ms
- Electrodes → Fz, Cz, FC1, FC2, CPz, AFz, Fpz, F1, F2

Results:

- Behavioral → effect ±lex
- erp → effect ±fa for n325

Table 6.6: Overview of mean stimulus properties for both lexical words and pseudowords ±FA (total duration, syllable durations and timing of third syllable onset).

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd onset	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
WORD										
+FA	647.9	62.5	176.6	36.1	168.3	30.1	303.6	54.6	345.2	40.5
−FA	565.9	69.0	176.2	35.8	167.7	30.5	221.7	58.7	346.5	56.4
PSEUDOWORD										
+FA	658.9	44.8	169.6	25.3	175.3	31.3	314.0	47.2	345.0	31.9
−FA	564.0	49.0	169.6	25.3	175.3	31.3	219.0	47.2	345.0	31.9

6.2.2.2.3 Procedure

Each participant was comfortably seated in an electrically shielded and sound attenuated room. Stimuli were presented through headphones using Python2.7 with the `PyAudio` library on a MacOS Sierra platform. Participants were instructed to judge as quickly and accurately as possible whether a word was a real word or a pseudoword by pressing the left or right arrow-keys on a standard keyboard using their dominant, right hand. Key assignment was counter-balanced across participants.

To ensure participants understood the task requirements, the experiment began with a short practice phase. This phase consisted of 8 trials that were similar to the experimental trials, but not included in the analyses. Each participant listened to all 160 stimuli (40 words +FA, 40 words −FA, 40 pseudowords +FA and 40 pseudowords −FA), which were evenly distributed over two blocks using Latin square designs. Block order was balanced across participants.

Participants were allowed to give their answer as soon as they heard the stimulus up until 1500 ms post stimulus offset. The ISI was fixed at 600 ms. The experiment, including the set-up of the EEG electrodes, took approximately 1.5 hour.

6.2.2.2.4 EEG recording and preprocessing

EEG data were recorded with 64 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958). The EEG signal was amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 2048 Hz. The data were preprocessed using the `EEGlab` package (Delorme & Makeig, 2004) with the `ERPlab` toolbox (Luck et al., 2010) in Matlab (Mathworks, 2014). Each electrode was re-referenced offline to a common average reference, band-pass filtered between 0.01 − 30 Hz and resampled at 256 Hz.

Following a visual inspection, signal containing EMG or other artifacts not related to eye-movements or blinks was manually removed. ICA was performed on the remaining data in order to identify and subtract components containing oculomotor artifacts. Finally, data were epoched from −0.2 to 1.5 seconds surrounding the onset of the target word and averaged within and across participants to obtain the grand-averages for each of the four conditions (word +FA, word −FA, pseudoword +FA, pseudoword −FA). See section 5.2.1 for more details on the preprocessing of the EEG signal.

6.2.2.2.5 Analysis—behavioral and EEG

Behavioral Response latencies were analyzed with mixed effects linear regression in `R` (Team, 2014) using the `lme4` package (Bates et al., 2012). The analysis tested whether lexicality and/or presence of FA significantly impacted reaction times. That is, the model outcome was reaction times and the predictors were lexicality and presence of FA. Furthermore, subjects and stimuli as well as by-stimuli random slopes for lexicality were included in the model as random variable in order to control across-subject and across-stimuli variability independent from our predictions. Likelihood ratio tests indicated whether each model fitted the data significantly better than the null-model (without any fixed factors). Visual inspection of residual plots did not reveal any obvious deviations from homoskedasticity or normality (see appendix C.4). For more details on the workings of mixed model analysis, see section 5.2.1.

EEG The ERP data are analyzed using a non-paramatric $t_{\max}$ permutation test (Blair & Karniski, 1993, see also Groppe et al. 2011; Luck 2014 and section 5.2.2 for a detailed description of this statistical test). We chose this statistical test because it allows for stringent control of Type I error without considerable decrease in sensitivity. To further maximize power and reduce the number of comparisons, the data were down-sampled to 125 Hz and electrode sites and time-windows were selected on a priori literature/knowledge. Because metrical processing/stress extraction has a predominantly frontal distribution, 9 fronto-central electrodes were tested for an effect of the ±FA manipulation (Fz, Cz, FC1, FC2, CPz, AFz, Fpz, F1, F2). Time-windows were calculated based on the time-windows in the previous study, which resulted in a time window of $151 - 251$ for the P2 and a time-window of $547 - 776$ for the N325.

Each comparison of interest was analyzed with a separate repeated measures, two-tailed t-test, using the original data and 2500 random permutations to approximate the null distribution for the customary family-wise alpha (α) level of 0.05.

6.2.2.3 Results

6.2.2.3.1 Behavioral results

Overall, performance on the lexical decision task revealed high accuracy (< 5% errors) with no differences between conditions. As is evident from figure 6.11 and table 6.8, lexicality had

Table 6.7: Overview linear mixed models. The model fitting the data best takes lexicality as fixed factor and subjects and stimuli variability as random factors. Presence of final accent did not significantly contribute to the prediction of reaction times.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	1061.94***	1059.83***	1112.79***	1054.66***	1107.89***	1101.76***
	(24.46)	(25.47)	(26.07)	(25.85)	(26.44)	(26.77)
lexicality			−108.71***		−108.71***	−96.05***
			(13.82)		(13.84)	(16.34)
±FA				10.40	9.83	22.14
				(8.80)	(8.75)	(12.19)
±FA:lexicality						−25.42
						(17.51)
AIC	38591.28	38415.85	38358.03	38410.27	38352.59	38344.92
BIC	38609.08	38457.39	38405.50	38457.74	38406.00	38404.26
Log Likelihood	−19292.64	−19200.93	−19171.01	−19197.13	−19167.29	−19162.46
Num. obs.	2792	2792	2792	2792	2792	2792
Num. groups: subj	18	18	18	18	18	18
Var: subj (Intercept)	10389.67	10491.87	10566.46	10499.48	10573.56	10575.08
Var: Residual	57741.78	50200.34	50185.95	50185.00	50171.06	50157.33
Num. groups: stimuli		160	160	160	160	160
Num. groups: lex:stimuli		160	160	160	160	160
Var: stimuli (Intercept)		1.17	2103.46	7371.37	1394.45	1227.17
Var: lex.stimuli (Intercept)		7552.73	2208.74	209.28	0.00	3466.80

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

a significant effect on response latencies with participants responding slower to pseudowords than to lexical words ($\beta = -108.71$, $SE = 13.82$, $t = -7.87$, $p < 0.001$, see table 6.7).

Overall, these results are similar to those reported in the previous study. However, we did notice that listeners were on average slower to respond in this study compared to the study wherein IA had been manipulated. While this difference could indicate delayed response due to listeners being aware of the metrical manipulation (which occurred later for FA than for IA), we consider it more probable to reflect a difference in experimental set-up; in the lexical decision wherein IA was manipulated, listeners were able to use button-boxes (left and right thumbs) to give their answer, while in the current study they were asked to provide their decision using their right index- and middle-finger using the arrows on a standard keyboard.

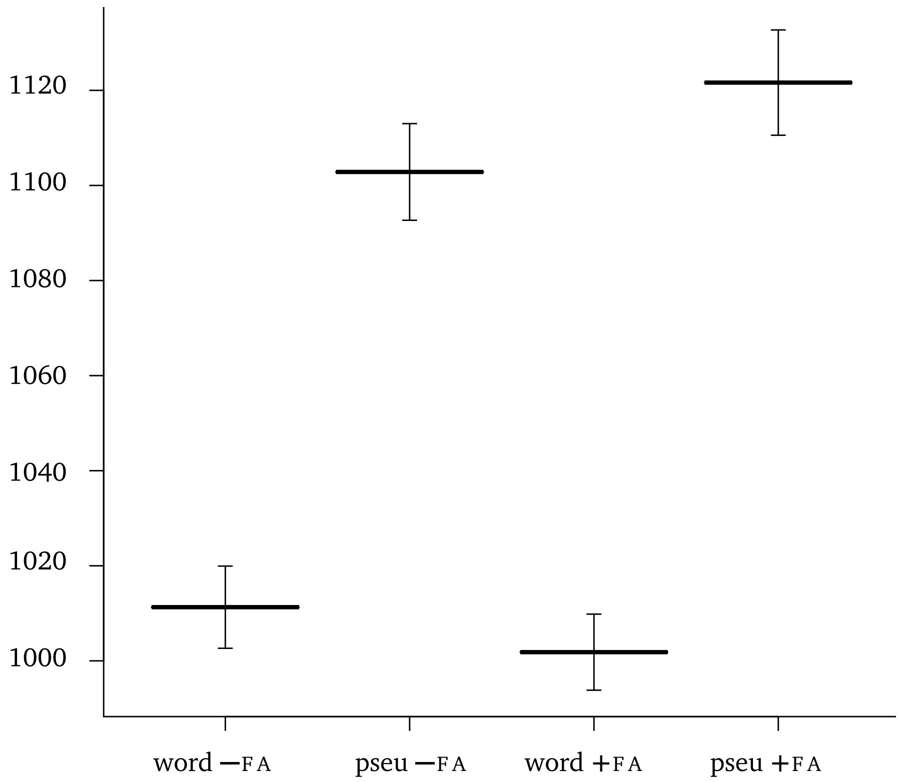
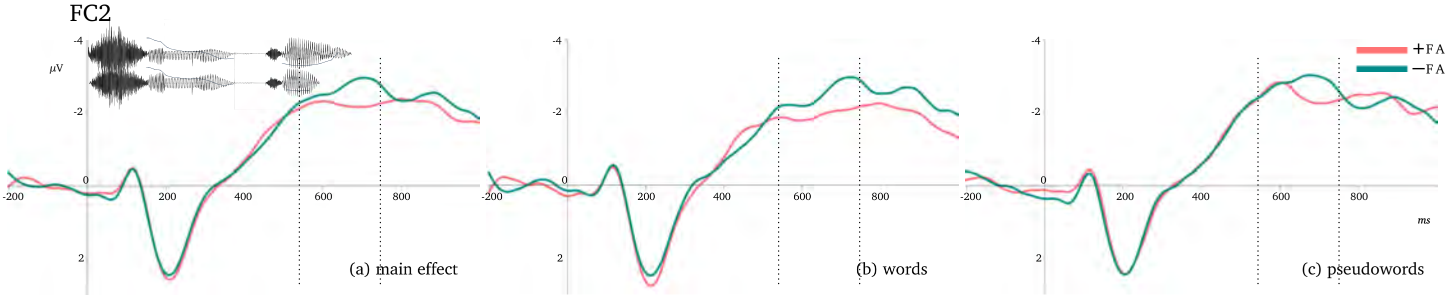


Figure 6.11: Error-bar plot of mean reaction times for all four conditions (word −FA, pseudoword −FA, word +FA, pseudoword +FA) showing a clear significant effect of lexicality and no effect of presence of final accent, nor an interaction between both manipulations.

Again similar to the previous study, the metrical manipulation had no effect on response latencies; FA did not impact reaction times ($\beta = 10.4$, $SE = 8.8$, $t = 1.18$, $p = 0.24$, *ns*) nor did it interact with lexicality ($\beta = -25.42$, $SE = 17.51$, $t = -1.45$, $p = 0.15$, *ns*).



(a) main effect (b) words (c) pseudowords

Figure 6.12: Grand average N325 in the ±FA condition (−FA in green, +FA in pink), recorded at the FC2 (fronto-central) electrode for: (a) main effect, (b) words (c) pseudowords. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [ʃibyte] (*chibuté*) +FA and −FA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are low-pass filtered at 10 Hz.

Table 6.8: Reaction times per condition. Data analysis revealed a significant effect both of lexicality, but no effect of ±FA and no interaction.

Response latencies	*m*	*sd*	maximum	minimum	% errors
WORD					
−FA	1011.29	224.25	2290.72	612.30	3.4%
+FA	1001.81	206.79	2444.67	651.06	2%
PSEUDOWORD					
−FA	1102.84	274.50	2461.48	623.52	1%
+FA	1121.36	297.45	2474.52	658.13	1%

6.2.2.3.2 ERP results

P2 Neither presence of FA nor lexicality modulated P2 amplitude (respectively, $p = 0.25$ and $p = 0.17$, *ns*).

N325 In the N325 time-window, there was a main effect of presence of FA (critical t-score: ±4.0428 ($df = 19$), $p < 0.05$). Compared to stimuli +FA, stimuli −FA elicited a larger negativity in the fronto-central region (FC2)[2] from 734−740 ms after stimulus presentation (see figure 6.12). This indicates an expectation for stimuli to be presented with final accent. The effect seems particularly prominent in the lexical word condition and, indeed,

---

[2] Note that this is the same electrode that was significant in the previous study wherein it was IA that was manipulated.

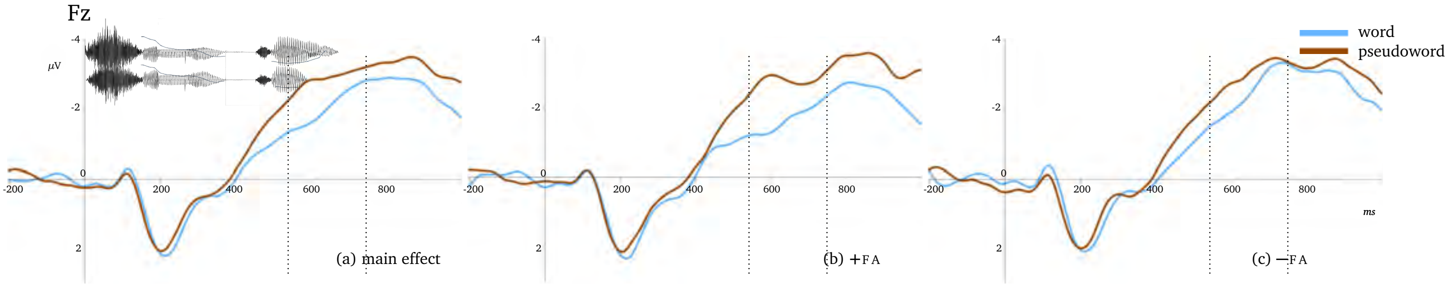


Figure 6.13: Grand average N325 in the lexicality condition (pseudowords in brown, words in blue), recorded at the Fz electrode for: (a) main effect, (b) +FA, (c) −FA. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [ʃibyte] (*chibuté*) +FA and −FA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are lowpass filtered at 10 Hz.

lexicality also significantly modulated the N325 in the frontal region (Fz and FC1) from $585 - 601$ (critical t-score: ±3.7571, $df = 19$, $p < 0.05$) with pseudowords eliciting a bigger N325 than words (see figure 6.13). The effect was significant within the +FA condition (critical t-score: ±3.7313, $df = 19$, $p < 0.05$), however it was not significant within the condition −FA ($p = 0.15$, *ns*). The effect of lexicality thus depended on the presence of the final accent, indicating an interaction between the lexical and metrical manipulations.

6.2.3 Discussion

Metrical stress is generally known to play an invaluable role in word level processing as it serves as gateway to the mental lexicon. In French, however, the contributions of stress to lexical processing are less well understood. French presents a language wherein stress is not lexically distinctive and held to apply to the level of the phrase. Consequently, most models of French prosody do not include the lexical domain in their description, and the role of French accentuation in word recognition has attracted rather little scientific interest.

Our results from the oddball studies presented in the previous section, however, suggested both the primary final accent (FA) and secondary initial accent (IA) to be attached to stress templates underlying the abstract representation of the word. That is, the results indicated both accents to be encoded in long-term memory and expected by the listener as a mark to both the left and the right lexical boundary. The results, therefore, hint to an important functional role for the accents in word level processing. Interestingly, our results also led us to propose that IA and FA may fulfill different purposes in this process, possibly interacting with different stages in the recognition of the word.

Here, we more specifically examined the interplay between metrical stress and the stages in lexical access in two lexical decision studies, one in which we manipulated the presence of IA, and the other in which the presence FA had been manipulated. Below, we will discuss the results of these studies in turn. First, in section 6.2.3.1, we present our behavioral findings which underline the value of the method of ERP in the investigation of pre-attentive, automatic processes such as metrical stress extraction. Then we present our ERP-data that demonstrate that presenting words with the French preferred stress template—with both the initial and final accent—facilitates lexical access.

In section 6.2.3.2, we discuss the results of our first study wherein the presence of the initial accent had been manipulated. In this study, we found the initial accent to *automatically* and *pre-lexically* be involved in lexical access. In section 6.2.3.3, we inspect our findings of the second lexical decision study, which show the final accent to be phonologically encoded as right marker of the lexical boundary and interact with lexical processing. In section 6.2.3.4, we discuss whether it is at all appropriate to conceive of FA and IA as interacting with specific stages in lexical access. That is, while the accents likely do play different roles in speech processing, cuing listeners on when to initiate lexical access versus when to wrap-up lexical competition, they both hold a pre-lexical status and, as such, are more probably involved in all processes of speech comprehension, i.e. metrical analysis has a *cascading* effect *throughout* speech comprehension.

In section 6.2.3.5, the findings of both studies are related to those presented in Böcker et al., a study wherein similar findings are reported on metrical processing in Dutch, a *stress based* language with *distinctive* and *obligatory lexical* stress. And, finally, we will conclude by pointing out the main effect of both our IA and FA metrical manipulations on the N325, which suggests a preference and expectation for the accents at the level of the *word*, *even if* stress is not lexically distinctive in French, *even though* the lexical word had no place in traditional descriptions, and *even though* attention had been explicitly diverted.

6.2.3.1 Behavioral results

In both the ±IA and ±FA studies, the reaction times showed a lexicality effect whereby listeners were slower to respond to pseudowords than to lexical words. This effect was expected and indicates participants had greater difficulty with lexical retrieval when words were absent from their mental lexicon. Still, also in both studies, performance on the lexical decision task revealed high accuracy (< 5% errors) with no differences between conditions, indicating a ceiling effect that is common in lexical decision tasks and often observed when a task is too easy.

While we expected the metrical manipulations to interplay with lexical processing, the advantage of the stress patterns was not reflected in our behavioral results. IA and FA were also expected to facilitate lexical retrieval, however neither presence of IA nor FA impacted reaction times, nor did they interact with lexicality. For the initial accent, it is possible that the unexpected absence of IA had been resolved at the time listeners responded. Conversely, for the final accent, the lexical decision likely had been made at the time of the metrical manipulation and resulting hindrance. Furthermore, the high accuracy scores suggest that the lexical task was not very demanding, resulting in only slight fluctuations in both accuracy rates and response latencies across the (task unrelated) metrical conditions.

Indeed, as was explained in section 3.2, behavioral measures may be unsuited for the detection of subtle metrical manipulations, since they do not only reflect the hindrance following unexpected stress templates, but also response selection, preparation, and execution (e.g. Rothermich et al., 2010). Because the method of ERPs provides a highly sensitive and *online* measure to obstructed processing, the ERP were more informative in the present studies and, indeed, showed the accents to differently interact with lexicality, as is discussed below.

6.2.3.2 The initial accent in pre-lexical processing

For the initial accent, we obtained a main effect of presence of IA, such that stimuli without initial accent elicited a larger negativity in the fronto-central region than stimuli with IA (see figure 6.9) indicating that processing was more demanding when there was no initial accent. Recall that Böcker and colleagues report similar findings after manipulating stress in Dutch, a language with lexically distinctive stress. In their study, listeners were asked to discriminate between the Dutch dominant stress template and a less frequent stress template. Words presented without the dominant stress pattern elicited a more ample N325. In our study, words presented without IA resulted in the larger N325 suggesting that, even though stress is not lexically distinctive in French, IA is part of the French expected stress pattern.

The effect appeared more pronounced for the nouns than for the pseudowords, but the metrical manipulation did not interact with lexicality in its modulation of the N325 amplitudes. This suggests that the initial accent predominantly interacted with the pre-lexical stage during word processing, wherein it cued listeners on when to initiate lexical access, a process which is facilitated by prediction (e.g. Pitt & Samuel, 1990; Large & Jones, 1999; Friston, 2005; Scharinger et al., 2016; Tavano & Scharinger, 2015).

This interpretation is confirmed by our results in the P2 time window. Not only was there no effect of ±IA on the P2 (a component generally associated with bottom-up extraction of purely physical/acoustical parameters Hillyard & Picton, 1987), showing our results to reveal a more controlled process in which stress is extracted in a top-down fashion, but lexicality *was* found to affect the P2 when stimuli were presented without IA. More specifically, pseudowords elicited a more ample P2 than did words when presented without initial accent. Because of the interaction with the metrical manipulation, and because of the early latency range of the P2, which is held to precede the lexical stage in word processing (Grosjean, 1980), the effect was interpreted to be the product of a temporal overlap between the P2 and the N325.

Indeed, the N325 was more negative for −IA stimuli than +IA stimuli and this difference was larger in the words condition than in the pseudowords condition. This means that, in the −IA condition, the overlap between the P2 and the N325 will be more evident for words than for pseudowords, while in the +IA condition the overlap will be smaller (as +IA stimuli elicited a smaller N325). Moreover, the lexicality effect failed to replicate in the lexical decision study wherein FA was manipulated, even though listeners had been presented the *same* items. Although, as we will discuss below, the final accent also modulated the N325, the N325 did not overlap with the P2 presumably because the final accent is later and, consequently, the N325 is temporally more distant from the P2.

Note that finding a temporal overlap between the N325 and the P2 in the

IA lexical decision study, implies that the process of stress extraction started before our predefined N325 time-window (201 − 431 ms) and during the P2 time-window (151 − 251 ms). Such an early latency then confirms that IA is anticipated by listeners and interacts with the early, pre-lexical stage in word processing. In fact, Böcker et al. (1999) report a similar overlap between the N325 and the P2 at the fronto-central sites, and consider it to reflect the interface of automated acoustic processing and controlled, top-down metrical analysis. For this reason, they argue the N325 to potentially index difficulties in processes that are involved in pre-lexical speech segmentation and the initiation of lexical access on the basis of rhythm and metrical stress. In that view, the N325 directly measures the role of metrical stress in speech processing as proposed in MSS and ABH (Cutler & Norris, 1988; Pitt & Samuel, 1990). Metrical stress helps listeners to a priori guide their attention towards word onsets, where the perceptually stable and prominent syllables cue listeners on when to commence their search in the mental lexicon.

In our study, the N325, and its overlap with the P2, demonstrate that word recognition crucially involves the automatic and pre-lexical extraction of the French initial accent. Upon hearing the initial accent, listeners infer that a new word has started and begin the process of lexical access. More broadly, the results show the value of metrical stress processing in French and give extra weight to the call in Astésano & Bertrand (2016) for metrical stress to be given a more prominent place in the descriptions of French prosody.

6.2.3.3 The final accent in its interaction with lexical selection

In the second lexical decision study wherein the final accent had been manipulated, stimuli −FA elicited a larger negativity, compared to stimuli +FA, in the fronto-central region (FC2) (i.e. the same electrode that was significant in the study wherein IA had been manipulated, see figure 6.12). In this study it is unlikely that FA should serve to indicate listeners on when to start formulating lexical hypotheses, the final accent being at the word's offset and thus rather late. Instead, the accent is held to confirm the strongest activated lexical item in the cohort and thus interact with later stages in lexical processing.

Indeed, the metricality effect was particularly prominent in the lexical word condition, and, moreover, lexicality also modulated the N325, but only when nouns and pseudowords had been presented with final accent (see figure 6.13). The effect of lexicality thus depended on the presence of the final accent, indicating an interaction between the lexical and metrical manipulation for FA. The ERP-figures suggest that while words with final accents are easy to process, pseudowords *and* words without final accent are difficult, presumably because they are similarly unexpected.

That is, pseudowords require a re-analysis when lexical access fails, because the word does not match any lexical representation in the mental lexicon. Words presented without their expected right boundary marker cuing the word's offset, may require a similar re-analysis, while nouns presented with their expected metrical pattern (i.e. with final accent) do not need to be re-analyzed because they match their canonical word representation more readily. These results therefore corroborate our findings in the oddball studies presented in the previous section, and demonstrate that the final accent is *phonologically natural* and *encoded* in the mental lexicon where it underlies the representation of the word.

6.2.3.4 Cascaded processing of metrical information

Because of the latency of the N325 to ±FA stimuli, the lexicality effect on the metrical N325, could also have been an overlap between the N325 and the later lexico-semantic N400, suggesting FA to additionally interact with access to meaning or semantic integration. In fact, a visual inspection of the ERP components in the lexical decision study wherein IA had been manipulated, also suggests that the presence of the initial accent affected the later lexico-semantic or semantic integration stages of word processing, as there seems to be an amplitude difference in the latency range typically associated with the N400 (see figure 6.9), similar as in the study wherein FA was manipulated. It appears that both IA and FA are pre-lexical, but IA sooner indicates listeners that a new word is upcoming thereby cuing lexical access, while FA *may* also cue lexical access but simultaneously indicates to the listener that the word is about to finish and s/he should complete the lexical competition (no new information will be provided). In this view, IA and FA are both encoded and attached to the cognitive representation of the word and anticipated by listeners, so that it is not unreasonable to assume that they should also impact later speech processes such as semantic retrieval.

It is however unclear whether these late amplitude modulations observed in the current studies really reflect difficulties in the post-lexical processes typically associated with the N400 (see section 3.2.3). Modulations of N400 amplitude are most commonly observed when a given word is presented in a semantically conflicting context, either through semantic priming (coffee-tea versus chair-tea) or in a semantic anomaly paradigm (I like my coffee with cream and sugar/socks) (Lau et al., 2008). In the current lexical decision paradigms, words were however presented in isolation, i.e. without semantic context. Moreover, the N400 has been shown relatively insensitive in experimental settings wherein semantic access is repeatedly discouraged (such as in lexical decision paradigms wherein many presented words lack semantic content) (Yan et al., 2017). Finally, the N400 is typically maximal over centro-parietal sites (Brown & Hagoort, 1993; Kutas & Federmeier, 2011, see also section 3.2.3), while the reported ERPs in the current study have a fronto-central distribution.

Although, there have been reports of phonological N400s with a more frontal distribution (e.g. Böcker et al., 1999; DeLong et al., 2005; Lau et al., 2008; Rothermich et al., 2010, 2012), combined with the lack of context, we are reluctant to interpret the negativities observed in the current studies as instances of the N400. We therefore follow the interpretation proposed in Böcker et al. (1999), where the authors encountered a similar late fronto-central amplitude difference as a result of their metrical manipulation, and interpreted it to reflect N325 residue. Still, the modulations in this late time-window and their possible implication for cascaded processing of metrical stress throughout speech comprehension, are intriguing and motivated an additional study more adapt to determine if accentuation also affects the later stages of speech processing, which is presented in the next section.

6.2.3.5 Present findings in relation to Böcker et al. (1999)

The results presented in the current section indicate that French listeners have a metrical anticipation for stress templates including both IA and FA as demonstrated by the modulation of the N325, a component which indexes difficulties in stress extraction during speech processing. Note that this finding is exceptional in demonstration the value of metrical processing during speech comprehension, particularly in the study of French.

Indeed, similar findings, wherein metrical manipulations modulated the N325 during word-level processing, have, to the best of our knowledge, only been reported in Böcker et al. (1999). Importantly, in that study, the language under investigation was Dutch, a *stress based* language with *distinctive* and *obligatory* lexical stress (Böcker et al., 1999). In their study, words presented without the dominant stress pattern elicited a more ample N325. Obtaining comparable effects in our investigation of metrical processing in a language traditionally held syllable based, wherein stress is not only not lexically distinctive, but, additionally, held to apply to the level of the phrase, underscores the value of metrical stress in speech processing.

The amplitude modulations of the N325 were small (between $1 - 2.5$ μV), but robust as revealed by our conservative non-parametric statistics (see section 5.2.2). In fact, finding a relatively small difference in amplitude was expected and comparable to the amplitude difference reported in Böcker et al. (1999). Similar to Böcker and colleagues, we did not manipulate the *legality*, but rather the *probability* of the presented stress templates. That is, while, in French, we show there to be a preference for words to be marked with their underlying metrical patterns, the accents do not depend on an explicit surface realization to be legal in French (see section 1.5). Recall, that when a word is embedded into a phrase, final accents within the phrase may be phonetically reduced, or de-accented, to favor a more prominent marking of the phrase boundary. Similarly, the initial accent is held to only surface when rhythmically necessary, e.g. to balance out a long stretch of syllables pronounced without FA. So, while French listeners may expect and prefer words to be marked with their underlying metrical pattern, words without explicit surface realization of the accents do not exceedingly hamper lexical processing.[3] Still observing a processing cost to our metrical manipulation

---

[3] As was additionally evidenced by the behavioral data, i.e. the high accuracy rates ($< 5\%$ errors) in both the ±IA and ±FA studies.

therefore underlines the phonological status of the accents.

Finally, while, in the study of Böcker and colleagues, the N325 effect was especially pronounced when participants *explicitly* attended the metrical structure of the stimuli, in our study, attention was *diverted* from the stress manipulation using a lexical decision task. That is, the N325 to dispreferred stress patterns observed in Böcker and colleagues, was particularly prominent in a metrical discrimination task that required stress patterns to be processed explicitly, and reduced in a task wherein participants were instructed to passively listen to the presented words. In our study, not only did we not alert listeners to the metrical manipulations, but we preoccupied them with a task requiring them to actively search their mental lexicon. Still finding a robust modulation of the N325, provides *even stronger* evidence that word processing *naturally* and *pre-attentively* engages the extraction and processing of the accentual information during lexical access in French.

IN SUM, in the current investigation on metrical processing in French, both the initial and the final accent modulated the fronto-central N325; a larger N325 emerged when either had been omitted. This suggests that, even though stress is not lexically distinctive, held to apply to the phrase, and allegedly not the metrical unit in French, listeners appear to have a metrical expectancy for both IA and FA and use the accents in word level processing. Upon speech input, listeners immediately extract the metrical information which they automatically compare to their anticipation. The facilitatory effect of metrical anticipation during this process then starts early and during pre-lexical analysis when lexicality had little influence (as was evident from the findings obtained in our ±IA lexical decision study) and continues even after lexical access is completed, but now metrical structure interacted with lexicality (as is evident in the lexicality effect for only stimuli with final accent). The results, therefore, confirm our findings presented in the previous section and indicate that the initial and final accent are phonologically encoded and as such contribute to word-level processing.

We additionally observed an amplitude modulation in a later latency range that could suggest hindered metrical processing to continue to obstruct post-lexical analysis during speech comprehension. As mentioned above, in the next section, we will present a study which more explicitly set out to examine the interplay between metrical anticipation and late lexico-semantic processing. In the study, we orthogonally manipulated the presence of IA in semantically congruent or incongruent sentences. Indeed, the initial accent has been found highly involved the marking of lexical structure in several perception studies, even more so than FA (Astésano et al., 2007, 2012; Garnier et al., 2016; Garnier, 2018), pointing towards a strong association between IA and word demarcation.

Also, in manipulating IA, we avoid the temporal overlap between ERP components that can be expected by manipulating FA and can make interpretations difficult. Furthermore, the study allows us to observe whether the initial accent, which, in the current study, was found to be crucially involved in the early pre-lexical stage of word recognition, also interacts with later, post access stages in speech processing.

Finally, in the study presented next, ±IA words are embedded in sentences, which allows us to confront a common critique according to which IA on words presented in isolation (as in the studies presented in the current and in the previous section) is utterance initial such that the observed facilitatory effects by metrical expectations could still be argued to apply to the post-lexical domain. In embedding ±IA words within congruent or incongruent semantic contexts, we may take a step further towards determine whether French metrical stress is encoded at the level of the lexical word, anticipated by listeners and functionally imperative in French speech comprehension.

6.3 French stress in lexico-semantic processing

In the current ERP-study, we investigate the interaction between the French initial accent and lexico-semantic processing. Although the initial accent is thought of as an optional secondary accent in French, and mainly recognized for its rhythmic balancing function and role in emphasis placement, in the current dissertation, IA is argued to have a lexical representation (*cf.* Rossi, 1980; Di Cristo, 1999, 2000; Astésano et al., 2007; Astésano, 2017), which our results presented in the previous sections seem to confirm. As such, the initial accent is likely to play a much more prominent role in speech comprehension.

Indeed, lexical stress is known to have invaluable functions in word processing because it is phonologically encoded and attached to the cognitive representation of the word, i.e. stress is part of the lexical entry, due to its place in the mental lexicon (e.g. Eulitz & Lahiri, 2004; Cutler, 2010).

In stress based languages, such as English or Dutch, the functional value of stress in speech comprehension is well established. As was presented in section 1.1, lexically stressed syllables are perceptually stable, acoustically salient, temporally predictable, and they signal word boundaries. And, indeed, the initial accent draws in attention with its rise in pitch (and secondary lengthening of syllabic onset) and is located word-initially, such that it may indicate the onsets of words and cue when to initiate lexical access. In fact, previous perception studies have shown IA to be a reliable cue to word boundaries to be perceived as prominent at both phrasal and lexical levels (Astésano et al., 2007, 2012; Garnier et al., 2016; Garnier, 2018).

Also, French accentuation is argued in the current work to be temporally predictable (i.e. separated by approximately 550 ms, Fant et al. 1991) thus forming a metrical framework which guides the listener's attention (Pitt & Samuel, 1990; Large & Jones, 1999; Astésano, 1999, 2001). Indeed, the initial accent is regularly perceived even when its pitch rise is weak or peaks further along in the word (e.g. Jankowski et al., 1999; Astésano et al., 2012; Garnier, 2018), indicating a strong metrical expectation for IA.

Furthermore, IA has been found to guide speech segmentation (e.g. Banel & Bacri, 1994; Bagou et al., 2002; Rolland & Lœvenbruck, 2002; Christophe et al., 2004; Welby, 2007; Spinelli et al., 2010) in spite of French often being described a syllable based due to the fairly homogeneous metrical weight on syllables. Note, however, that as we remarked previously, in those studies, segmentation was not considered lexical but presumed phrasal, i.e. listeners are assumed to adopt a *prosodic segmentation strategy* in which intonational and accentual patterns function to segment *prosodic groups* from the speech signal (Wauquier-Gravelines, 1999), a view that stems from traditional descriptions of French as 'a boundary language' (Vaissière, 1991), according to which stress is acoustically merged with intonational boundaries (Rossi, 1980).

The ERP studies presented in the previous sections, suggest a lexical representation for IA and underline a role of IA in French speech processing. Furthermore, the results of the lexical decision studies suggested the metrical manipulations elicited an N400, as there appeared to be a negativity in the latency range typically associated with the N400. We were however cautious to interpret this negativity as an N400, because words were presented in isolation, i.e. without semantic context. Presenting word in isolation had, as additional consequence, that IA was always in utterance initial position. As mentioned in the previous section, because words had been presented as independent utterances, they may have been processed as individual phrases (see section 1.2, for a discussion). Hence, it can not be ruled out that the templates—and the processing cost when IA was omitted—applied to the phrase level and not to the level of the lexical word. In the current N400 study, we manipulated the semantic congruity of the sentences, allowing us to better determine whether IA is represented at the level of the word, and to additionally observe whether metrical manipulations affect the later processing stages in speech comprehension.

Indeed, while the N400 is predominantly associated with purely post-lexical manipulations, a number of studies have shown misguided phonological expectations in healthy subjects (e.g. Praamstra & Stegeman, 1993; Dumay et al., 2001, 2002; DeLong et al., 2005) or impaired phonological analysis in patients (Robson et al., 2017) to interfere with subsequent semantic evaluation and modulate the N400 (see also section 3.2.3). Furthermore, metrical information has also been found to interplay with lexico-semantic processing.

In a series of studies, Rothermich and colleagues manipulated the metrical regularity in German jabberwocky (Rothermich et al., 2010) and semantically anomalous sentences (Rothermich et al., 2012; Rothermich & Kotz, 2013) by presenting words either with a metrically regular or irregular beat and showed metrical regularity to facilitate semantic ambiguity resolution, as indicated by a modulated and earlier N400, which, unlike its usual centro-parietal distribution, appeared to be more frontally located. The authors relate their findings to neural entrainment, and suggest metrically predictable stress to provide a metrical framework to which endogenous oscillations can align in an effort to optimize speech comprehension (*cf.* Pitt & Samuel, 1990). In their work, regular (i.e. predictable) linguistic meter allowed for a facilitated lexico-semantic integration.

Finally, in a previous ERP study investigating the relationship between metrical structure and late speech processing in French, metrical violations were found to obstruct semantic processing (Astésano et al., 2004; Magne et al., 2007, see also section 3.2.3). In the study, participants listened to sentences in which semantic and/or metrical congruity was manipulated. Recall that semantic congruity was manipulated by presenting sentences in which the last word was incoherent with the semantic context of the sentence, while metrical congruity was manipulated by lengthening the medial syllable of the last word, an illegal stress pattern in French. Furthermore, listeners completed two different tasks, one in which they attended semantic congruity, and one in which they judged metrical congruity. This allowed Magne et al. (2007) to determine whether metrical and/or semantic processing proceeds automatically or depends on the direction of attention.

Behavioral results showed listeners to make more errors when either meter or semantics was incongruent. Furthermore, listeners made the most errors when meter was incongruent, but semantics was congruent, indicating that metric incongruities disrupt semantic processing. This interpretation was corroborated by their results from the ERP data. Not only did Magne and colleagues obtain a larger N400 to metrically incongruous words than to metrically congruous words in the metric task, but, interestingly, the metrical violation resulted in an increased N400, also in the semantic task (i.e. independent from attention), and even when the sentences were semantically congruent (see also Astésano et al., 2004). These results indicate that accentual patterns, also in French, affect the later stages of speech comprehension, during which access to meaning and semantic integration takes place.

In the study of Magne et al. (2007), however, the processing cost resulted from presenting an illegal stress pattern, with metrical weight on the medial syllable, and it remains unclear whether semantic processing also suffers when words are presented with metrical structures that, while legal, deviate from the expected stress pattern. Or, put more concretely, if IA is linked to the phonological representation of words and is, along with FA, the expected stress template in French, we anticipate that presenting words without IA modulates the N400.

6.3.1 Study Summary

Research Question: *Does the pre-lexical French initial accent interplay with later processing stages in speech comprehension such as access to meaning and contextual integration?*

Procedure: 18 *listeners participated in the current* N400 *study. They judged semantic congruity of target words that were either semantically congruent or incongruent, and presented with initial accent or without. The target word were embedded in sentences the bias of boundary marking and investigate the interplay between accentuation and later speech processing stages.*

Results: *Results reveal a significant interaction effect: the fronto-centrally located* N400 *was larger for words presented without* IA*. Furthermore,* *we found an effect of semantic congruity, on the centro-parietal region (the traditional region for* N400*), which was bigger for word* −IA *than for words* +IA*. Finally, we observed an interaction such that* ±IA *continued to modulate* N400 *amplitude, but only in the sentences that were semantically incongruent. Furthermore, as participants attended to the semantic content of the sentences, the finding underlines the automaticity of stress processing.*

Conclusion: *Our data confirm our hypothesis that the initial accent is encoded at a lexical level and anticipated by listeners. Presenting words without the accent hinders lexical access and cascades down the process of speech comprehension to additionally obstruct semantic processing. In sum, we demonstrate accentuation to be crucially involved in speech comprehension in French.*

6.3.2 Methods

6.3.2.1 Participants

20 French native speakers, aged 19 − 47 (mean age 24.2), gave their written consent and volunteered to take part in the study which was conducted in accordance with the Declaration of Helsinki. Subjects had foreign language skills at high-school level or less, they were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Due to excessive artifacts in the EEG signal, two participants are excluded from the EEG analyses.

6.3.2.2 Speech stimuli

Stimuli were selected and adapted from the corpus described in Magne et al. (2007). This corpus consisted of French carrier sentences ending with a trisyllabic target noun that either made sense in the semantic context of the

Procedure:

- Nr participants → 18
- Nr stimuli/condition → 20
- Task → semantic anomaly judgment

Preprocessing:

- Nr electrodes → 64
- Reference → Mastoids
- Epoch length → −200 − 1000

Analysis:

- Behavioral → Linear Mixed Effects Model in R (DV: RT; IV: ±ia/ ±s)
- eeg → t_{max} permutation test in Matlab, 2500 permutations (DV: amplitude; IV: ±ia/ ±s)

p2:

- Time-window → 181 − 281 ms

Metrical n400:

- Time-window → 351 − 451 ms
- Electrodes → Fpz, FCz, Fz, AFz, Fp1, Fp2, FC1, FC2, F1, F2, AF3, AF4

Semantic n400:

- Electrodes → Fz, Cz, FC1, FC2, P1, P2, C3, C4, Pz, P3, P4, CP1, CP2
- Time-window → 450 − 600 ms

Results:

- Behavioral → effect ±ia
- Behavioral → effect ±s
- erp → effect ±ia for metrical n400
- erp → effect ±s for semantic n400
- erp → interaction effect

sentence (semantically congruent, +S) or was nonsensical with its preceding context (semantically incongruent, −S) (see figure 6.14 for an example of the item +S and −S, with target words +IA and −IA). Congruent and incongruent target words were acoustically and phonologically similar and had been matched in word frequency and word and syllable duration (see table 6.9, a more detailed account on the construction of the sentences can be found in Magne et al. 2007).

Stimuli selection was based on the presence of a marked and natural IA in the original corpus in both semantic conditions. Because the primary phonetic parameter of IA is a rise in f_0 (Astésano, 2001, see also table 1.3), this meant that only sentences in which the target nouns in both semantic conditions started with a rise of f_0 of at least 10% on the first syllable compared to the preceding f_0 value on the (unaccented) determinant (Ladd, 2008; Astésano et al., 2007) were admitted in the current corpus. 160 stimuli met this criteria; 80 carrier sentences with 80 +S target nouns and 80 −S target nouns.

The metrical condition (±IA) was created by lowering the f_0 value on the first vowel of the target-words near the f_0 value on the preceding (unaccented) determinant in order to remove the natural +IA and create the −IA condition (see figure 6.14). This manipulation was achieved using a customized quadratic algorithm Aguilera et al. (2014) in PRAAT (Boersma & Weenink, 2016) which progressively modified the f_0 values while allowing for micro-prosodic variations to be maintained such that the natural sound of the stimuli remained intact. Further, the +IA stimuli were forward and back transformed to equalize the speech quality between +IA and −IA stimuli (see section 5.1.2 for a detailed description of the sound manipulation).

The resulting 320 stimuli over the four experimental conditions (+S +IA, −S +IA, +S −IA, and −S −IA) were divided over four lists, such that each participant was presented with 80 unique sentences, i.e. 20 sentences per condition.

Figure 6.14: Example of f_0 resynthesis with (+IA) and without initial accent (−IA) on semantically incongruent (+S, top two) and semantically congruent (−S, bottom two) sentences with quadratic interpolation from the f_0 value of the preceding determinant to the f_0 value at the beginning of the last stressed syllable for +IA targets (visible in blue). The time window of ±IA is indicated by vertical red dashed lines.

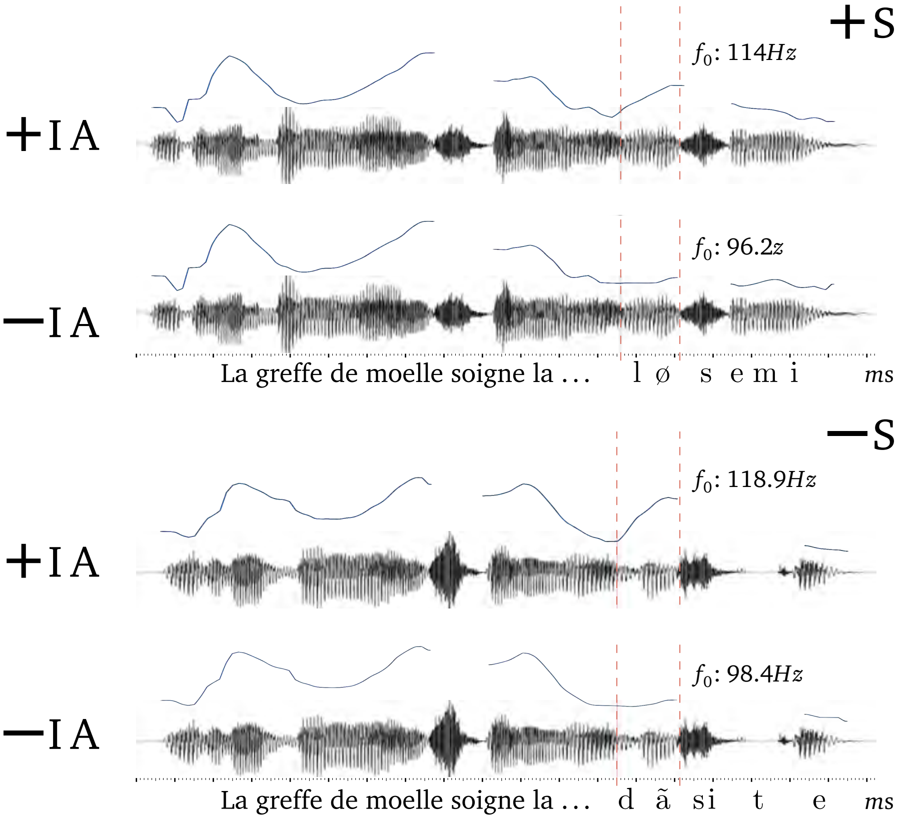


Table 6.9: Overview of mean stimulus properties within the semantically congruent and incongruent conditions ±IA (total sentence and target-word duration, first syllable and syllable-vowel durations, and first syllable-vowel f_0 values).

	Sentence *ms*		Target word *ms*		1st syllable *ms*		1st vowel *ms*		1st vowel f_0	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
SEMANTICALLY CONGRUENT										
−IA	2097.07	402.81	552.88	96.98	157.23	28.76	72.16	25.9	116.56	11.73
+IA	2092.07	402.81	552.88	96.98	157.23	28.76	72.16	25.9	126.38	12.2
SEMANTICALLY INCONGRUENT										
−IA	2122.59	411.72	583.45	61.72	160.57	32.16	77.86	27.23	123.02	42.18
+IA	2122.59	411.72	583.45	61.72	160.57	32.16	77.86	27.23	140.28	44.26

6.3.2.3 Procedure

Each participant was comfortably seated in an electrically shielded and sound attenuated room. Stimuli were presented through headphones using Python2.7 with the `PyAudio` library on a Windows XP 32-bit platform. Participants were instructed to judge as quickly and accurately as possible whether a sentence was semantically congruent or incongruent by pressing the left or right arrow key on a standard keyboard using their dominant, right hand. Arrow key assignment was counter-balanced across participants. The ISI was fixed at 600 ms. Participants were allowed to give their answer from the start of the target word until 1500 ms post stimulus offset. To ensure participants understood the task requirements, the experiment began with a short practice phase, consisting of 10 trials that were similar to the experimental trials, but not included in the analyses.

Each participant listened to a complete list of all 80 stimuli. Using Latin square designs, the four conditions (+S +IA, −S +IA, +S −IA, and −S −IA) were evenly distributed over two blocks, with block order balanced across participants. Total duration of the experiment, including the set-up of the EEG electrodes, was approximately 1.5h.

6.3.2.4 EEG recording and preprocessing

EEG data were recorded with 64 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958). The EEG signal was amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 2048 Hz. The data were preprocessed using the `EEGlab` package (Delorme & Makeig, 2004) with the `ERPlab` toolbox (Luck et al., 2010) in Matlab (Mathworks, 2014). Each electrode was re-referenced offline to the algebraic average of the left and right mastoids. The data were band-pass filtered between $0.01 - 30$ Hz and resampled at 256 Hz.

Following a visual inspection, signal containing EMG or other artifacts not related to eye-movements or blinks was manually removed. ICA was performed on the remaining data in order to identify and subtract components containing oculomotor artifacts. Finally, data were epoched from -0.2 to 1 seconds surrounding the onset of the target word and averaged within and across participants to obtain the grand-averages for each of the four conditions (+S +IA, +S −IA, −S +IA, −S −IA). See section 5.2.2 for more details on the preprocessing of the EEG signal.

6.3.2.5 Analysis—behavioral and EEG

6.3.2.5.1 Behavioral

The behavioral data (i.e. accuracy rates and response latencies) were analyzed in R (Team, 2014) with the `lme4` package (Bates et al., 2012). Visual inspection of residual plots did not reveal any obvious deviations from homoskedasticity or normality (see appendix C.5).

For the accuracy rates, binary logistic regression was used to analyze the two predictors semantic congruency and presence of IA. That is, the model tested how well semantic congruency and presence of IA predicted the proportion of errors. For response latency (a continuous variable), a linear mixed effects model was used to analyze the effect semantic congruency and IA had on reaction times.

Similar as for accuracy rates, the model additionally included participants and stimuli as random variables. More specifically, for the random structure, we had found intercepts for listeners and stimuli, as well as by-stimuli random slopes for the effects of metrical pattern and semantic congruity best accounted for underlying random variability. p-values were obtained by likelihood ratio tests of the model with the effect in question against the model without the effect in question. For more details on the application of mixed models for data analysis, see section 5.2.1.

6.3.2.5.2 EEG

The EEG data was analyzed with the non-parametric t_{max} permutation test, which allows for correction of multiple comparisons, while remaining statistically powerful (Groppe et al., 2011; Luck, 2014). To further maximize statistical power and reduce the number of comparisons, data were downsampled to 128 Hz.

Because, while the N400 resulting from semantic incongruities is typically maximal in the centro-parietal region of the brain (Brown & Hagoort, 1993; Kutas & Federmeier, 2011), violations in metrical/phonological expectancies more commonly result in a N400 that is more frontally located (e.g. Böcker et al., 1999; DeLong et al., 2005; Lau et al., 2008; Steinhauer & Connolly, 2008; Rothermich et al., 2010, 2012; Yan et al., 2017), we selected fronto-central and centro-parietal electrodes (Fpz, FCz, Fz, AFz, Fp1, Fp2, FC1, FC2, F1, F2, AF3, AF4, Cz, P1, P2, C3, C4, Pz, P3, P4, CP1, CP2).

Furthermore, because the phonological/metrical N400 has been reported to precede the semantic N400 temporally (e.g. Magne et al., 2007; Steinhauer & Connolly, 2008; Rothermich et al., 2010, 2012) we tested two separate time-windows; $351-451$ ms for the metrical N400 and $450-650$ ms for the semantic N400.

Finally, to make sure that modulations in our N400 time-windows would not reflect P2 residue due to differential acoustic processing on our ±IA stimuli, we also tested this time-window from $181-281$ ms.

6.3.3 Results

6.3.3.1 Behavioral data

6.3.3.1.1 Response accuracy

There was a significant main effect of ±IA with participants making more errors when stimuli had been presented −IA than when they had been presented +IA ($\beta = 1.58,\ SE = 0.63,\ t = 2.51,\ p < 0.05$, see table 6.10). The semantic condition was revealed a marginal predictor of error rate, with more errors when sentences were semantically congruent, than when they were semantically incongruent ($\beta = 1.73,\ SE = 0.94,\ t = 1.85,\ p = 0.06$). Interestingly, the error rates reported here are similar to those reported in Magne et al. (2007), with most errors on sentences that were semantically congruent, but metrically unexpected (note that the metrical manipulation actually created an *illegal* pattern in Magne et al. (2007)). Presence of IA and semantic congruency did not interact ($\beta = -0.3,\ SE = 1.26,\ t = -0.24,\ p = 0.81,\ ns$).

6.3.3.1.2 Reaction times

As can be seen in figure 6.15, both IA and semantic congruity affected response latencies (see also table 6.11).

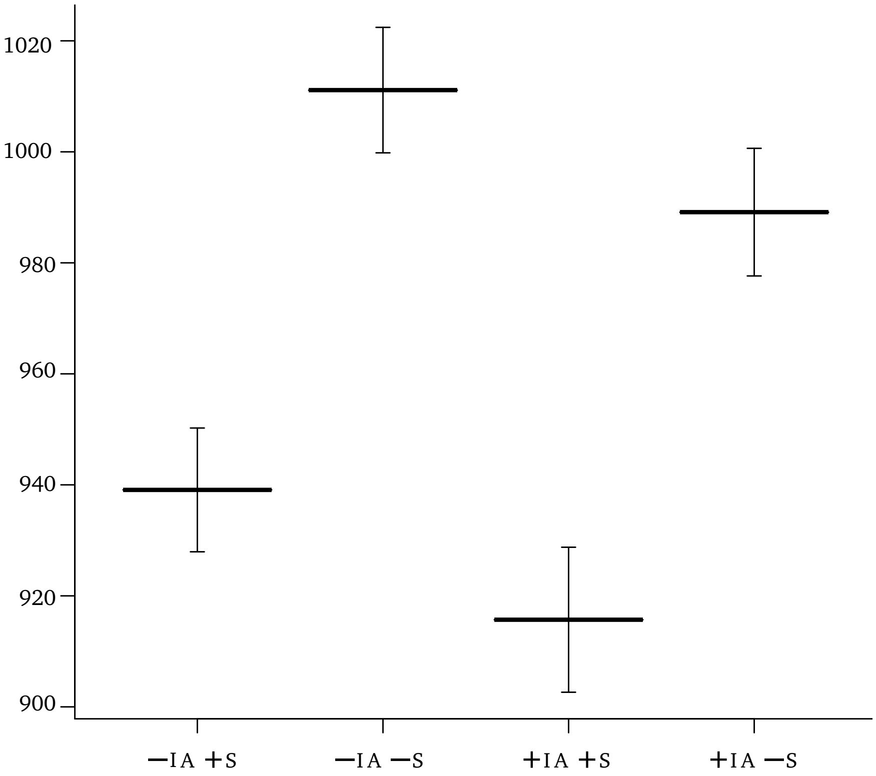


Figure 6.15: Error-bar plot of mean reaction times for all four conditions (−IA +S, −IA −S, +IA +S, +IA −S) revealing a significant effect of both ±IA and of ±S, with no interaction between the two experimental manipulations.

Table 6.10: Overview linear mixed models. The model fitting the data best takes semantic congruency as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent significantly contributes to the prediction of reaction times when it is the only fixed effect and marginally contributes when entered together with semantic congruency.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-3.00^{***}	-8.40^{***}	-9.50^{***}	-10.07^{***}	-10.52^{***}	-10.72^{***}
	(0.17)	(0.98)	(1.23)	(1.24)	(1.43)	(1.70)
congruency			1.73		1.69	1.93
			(0.94)		(0.98)	(1.41)
±IA				1.58^{*}	1.49^{*}	1.73
				(0.63)	(0.60)	(1.19)
±IA:congruency						−0.30
						(1.26)
AIC	539.70	412.20	405.14	400.87	399.81	401.76
BIC	550.10	443.39	441.53	437.26	441.41	448.55
Log Likelihood	−267.85	−200.10	−195.57	−193.44	−191.91	−191.88
Num. obs.	1338	1338	1338	1338	1338	1338
Num. groups: subj	20	20	20	20	20	20
Var: subj (Intercept)	0.16	1.10	1.31	1.40	1.31	1.32
Num. groups: congr:stimuli		160	160	160	160	160
Num. groups: ±IA:stimuli		160	160	160	160	160
Var: congr.stimuli (Intercept)		34.81	29.15	42.47	33.43	33.61
Var: ia.stimuli (Intercept)		0.00	3.72	2.20	2.11	2.11

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Response latencies	*m*	*sd*	maximum	minimum	% errors
SEMANTICALLY CONGRUENT					
−IA	939.09	202.87	1494.07	359.01	11%
+IA	915.70	239.43	1561.71	25.01	7%
SEMANTICALLY INCONGRUENT					
−IA	1011.12	209.89	1511.31	572.25	4.5%
+IA	989.14	206.93	1564.43	573.43	3%

Table 6.11: Reaction times per condition. Data analysis revealed a significant effect both of ±IA and of ±S, with no interaction between the two conditions.

When stimuli had been presented −IA, participants were slower to respond than when they had been presented +IA ($\beta = 21.0$, $SE = 9.37$, $t = 2.24$, $p < 0.05$, see table 6.12 for an overview of the regression analyses). This again indicates semantic ambiguity resolution was facilitated when words were presented with initial accent.

Furthermore, as mentioned above, semantic congruity also affected reaction times ($\beta = -78.46$, $SE = 16.81$, $t = -4.67$, $p < 0.001$); congruent sentences were responded to faster than incongruent sentences. This effect was expected and is in line with the results reported in Magne et al. 2007. Presence of IA and semantic congruency did not interact ($\beta = 10.66$, $SE = 18.04$, $t = 0.59$, $p = 0.55$, *ns*).

6.3.3.2 EEG

6.3.3.2.1 P2

As expected, neither ±IA nor ±S modulated the P2 amplitude ($p = 0.42$ and $p = 0.59$, *ns* respectively). This means that differences we find on the later metrical N400 and semantic N400 cannot be attributed to differences on the early P2, held to reflect more bottom-up processing of purely acoustic information (Hillyard & Picton, 1987).

6.3.3.2.2 Metrical N400

The data reveal a main effect of ±IA, i.e. ±IA words modulated the metrical

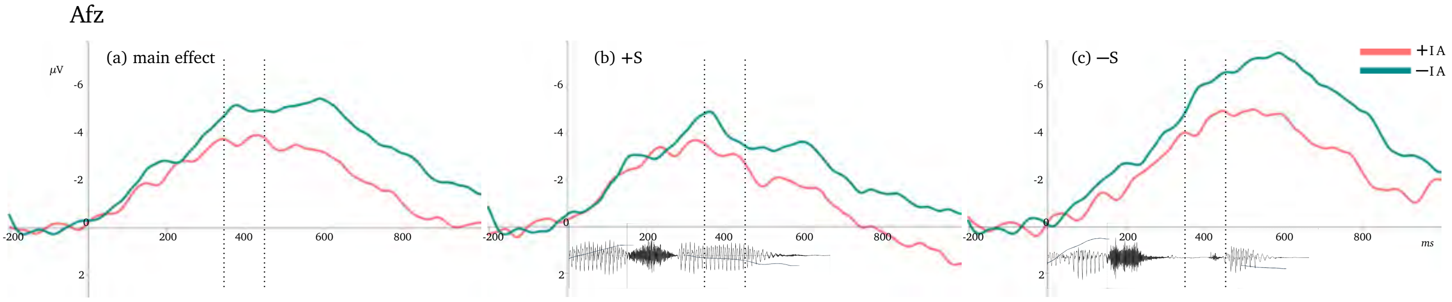


Figure 6.16: Grand average metrical N400 in the ±IA condition (−IA in green, +IA in pink), recorded at the Afz (anterio-frontal) electrode for: (a) main effect, (b) congruent sentences, (c) incongruent sentences. The tested time-window is indicated by dashed vertical lines. Furthermore to indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [løsemi] (*leucémie*, +S) and [dɑ̃site] (*densité*, −S) are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are lowpass filtered at 10 Hz.

N400 regardless of semantic congruency (critical t-score: ±3.9078, $df = 17$, $p < 0.05$). Words −IA elicited a larger N400 than did words +IA 375 ms post target word onset in the anterio-frontal region (Afz) (see figure 6.16).

Semantic congruency had no effect on the metrical N400 ($p = 0.14$, *ns*) nor did it interact with presence of IA ($p = 0.15$, *ns*).

Table 6.12: Overview linear mixed models. The model fitting the data best takes semantic congruency as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent significantly contributes to the prediction of reaction times when it is the only fixed effect and marginally contributes when entered together with semantic congruency.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	976.30***	977.21***	1016.63***	966.57***	1006.12***	1008.88***
	(25.07)	(26.89)	(28.18)	(27.24)	(28.52)	(28.91)
congruency			−78.46***		−78.27***	−83.64***
			(16.81)		(16.84)	(19.14)
±IA				21.00*	20.58*	15.20
				(9.37)	(9.31)	(13.03)
±IA:congruency						10.66
						(18.04)
AIC	17757.83	17567.17	17542.20	17557.93	17533.11	17527.13
BIC	17773.40	17603.50	17583.72	17599.45	17579.82	17579.03
Log Likelihood	−8875.92	−8776.58	−8763.10	−8770.96	−8757.55	−8753.57
Num. obs.	1326	1326	1326	1326	1326	1326
Num. groups: subj	20	20	20	20	20	20
Var: subj (Intercept)	11996.42	12415.76	12432.85	12339.64	12363.26	12377.19
Var: Residual	36671.85	25898.44	25946.90	25905.44	25948.81	25959.58
Num. groups: congr:stimuli		160	160	160	160	160
Num. groups: ±IA:stimuli		160	160	160	160	160
Num. groups: stimuli		80	80	80	80	80
Var: congr.stimuli (Intercept)		11031.88	7977.87	11057.78	8021.99	8016.48
Var: ±IA.stimuli (Intercept)		417.00	366.70	234.50	193.37	202.53
Var: stimuli (Intercept)		267.65	2153.78	591.00	336.86	1534.69

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

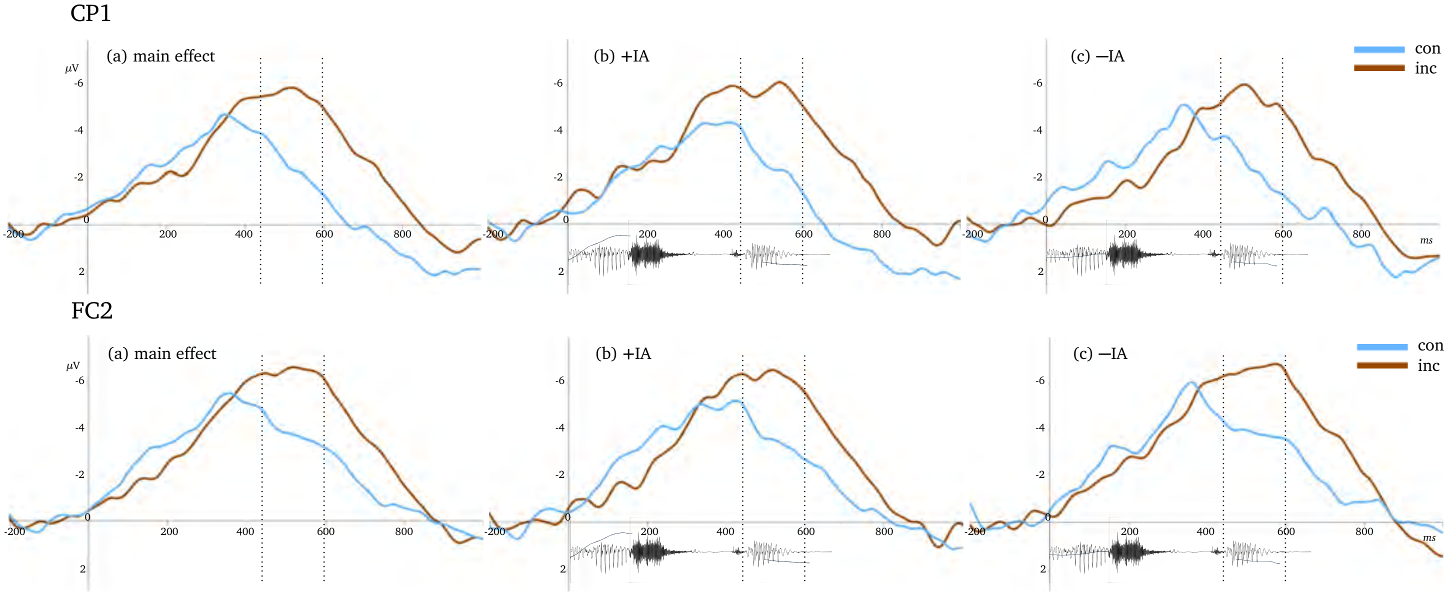


Figure 6.17: Grand average semantic N400 in the ±s condition (−s in brown, +s in blue), recorded at the CP1 (centro-parietal, top) and FC2 (fronto-central, bottom) electrodes for: (a) main effect, (b) +IA, (c) −IA. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [dãsite] +IA and −IA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are low-pass filtered at 10 Hz.

6.3.3.2.3 Semantic N400

In this later time-window, the ERP data show a main effect of semantic congruity (critical t-score: ±4.1627, $df = 17$, $p < 0.05$): semantically incongruent sentences elicited a larger N400 between $492 - 593$ ms after the onset of the target word than semantically congruent sentences in the left centro-parietal region (CP1) and the right fronto-central region (FC2) (see figure 6.17). This difference in N400 amplitude was also significant within the condition −IA (critical t-score: ±4.1038, $df = 17$, $p < 0.05$) and marginally significant within the condition +IA (critical t-score: ±4.1861, $df = 17$, $p = 0.086$).

The main effect of IA was not significant ($p = 0.28, ns$), however, we did observe an interaction between ±IA and ±s. The interaction effect between our two manipulations was significant between $523 - 593$ located at centro-parietal and frontal electrodes (critical t-score: ±4.0893, $df = 17$, $p < 0.05$, at Af4, Afz, CP1 and FC2), such that, in this later time-window, ±IA had continued to modulate N400 amplitude, but only in the sentences that were semantically incongruent (see figure 6.16).

Furthermore, visual inspection suggested a difference in N400 onset latency between semantically congruent and incongruent sentences, but only in the −IA condition, indicating that conflict resolution starts later for incongruent words without initial accent. Because this visual effect is important for the discussion of the additional semantic processing cost when words are presented −IA, we computed a regression analysis with, as dependent variable, peak amplitude latency, ±IA, semantic congruency and electrode cite (parietal, centro-parietal and central) as fixed effects, and participants as random effects. However, the analysis was not significant at $p = 0.11$. These results are further interpreted below.

6.4 Discussion

In the present study, we examined the phonological status of the French initial accent and its role in semantic processing. We were particularly interested in modulations of the N400 ERP component, a component typically observed subsequent to violations of lexico-semantic expectations (e.g. Kutas & Hillyard, 1980; Brown & Hagoort, 1993). Below, we present each of our findings in turn, starting with the main effect of ±IA on the fronto-central metrical N400 to then discuss the interaction between metrical expectancy and semantic congruence on the centro-parietal N400. Finally, we will examine our behavioral data which suggest that violated metrical anticipations slow down semantic conflict resolution during speech processing.

6.4.1 Metrical N400

During the metrical N400 time-window, presence of initial accent modulated N400 amplitude in the anterio-frontal brain area, irrespective of semantic congruency, i.e. words without initial accent elicited a larger N400 than did words with initial accent (figure 6.16). This, again, indicates that listeners expected words to be presented with initial accent, in line with the results reported in the previous sections. Furthermore, because our manipulation of IA did not modulate the acoustic P2, the metrical effect can be interpreted to reflect a more controlled process in the phonological processing of the initial accent, i.e. IA is phonologically natural.

Note also that, because in this time-window, we observed a main effect of our ±IA manipulation, this negativity may well be another instance of the N325 and indicate difficulties in stress extraction (*cf.* Böcker et al., 1999). This finding has two important consequences for interpreting the role of IA and more generally the domain of accentuation in French. First, replicating the results reported in the previous section is far from trivial for a language allegedly without accent wherein stress is not lexically distinctive and has been mostly ignored by the scientific community. Replication is at the core of science, and particularly the functional value of IA—the traditionally secondary and optional accent—has been gravely understudied. Moreover, while there has been more scientific interest for the contributions of stress in speech comprehension in stress based languages, metrical stress extraction during speech processing as reflected by a modulation of the N325 had only been shown in Böcker et al. (1999), and predominantly when listeners were performing a stress discrimination task requiring them to *explicitly attend* the metrical information. In the current work, we twice, and in two different paradigms, observed French listeners to have a metrical expectation for the initial accent as reflected in a modulation of the N325, which they extract and use in the task at hand, i.e. lexical retrieval and semantic access.

The second major conclusion we can draw in observing a main effect of IA in the current study, is that stress extraction is hindered when words are presented without their expected initial accent marking their onset, *even* when the word is *embedded* within a sentence (i.e. not presented in isolation). Indeed, as was explained above, the previous ERP studies had always manipulated IA on isolated words where the accent was in utterance initial position, which made it difficult to rule out advantages applying to the levels higher in the prosodic hierarchy. Here, however, we obtain the same effects despite IA not being utterance initial, underscoring the phonological status of IA as marker of the left boundary of the word (*cf.* Astésano et al., 2007, 2012; Garnier et al., 2016; Garnier, 2018).

6.4.2 Semantic N400

During the semantic N400 time-window, semantic congruity modulated the N400 in the centro-parietal regions, with semantically incongruent sentences eliciting a more ample N400 than did semantically congruent sentences (figure 6.17). This effect was however more pronounced when words were presented without IA than when they had been presented with IA , suggesting an interaction effect between semantic congruity and metrical expectation, such that pre-semantic processes (in this case the extraction of the initial accent) facilitated subsequent semantic evaluation. Indeed, the processes of word recognition and semantic retrieval unfold, due to the temporal nature of speech input, in a cascading manner (see chapter 2). Phonological analysis is required before semantic evaluation and this analysis is facilitated when the input meets phonological and metrical expectations.

Note that the findings therefore indicate that speech comprehension is impaired when the analysis of unexpected metrical stress templates has a downstream impact on semantic retrieval and integration (e.g. Praamstra & Stegeman, 1993; Dumay et al., 2001; DeLong et al., 2005; Robson et al., 2017). The results then contradict the hypothesis that the N400 can only be modulated by hindered post-lexical processes such as contextual integration (van den Brink et al., 2001; Brown & Hagoort, 1993; Boulenger et al., 2011), and, instead suggest phonological processes also affect N400 amplitudes. In this view, the N400 thus reflects the degree of lexical pre-activation with higher levels of pre-activation facilitating lexico-semantic processes and reducing N400 amplitude (Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; DeLong et al., 2005; Gilbert, 2014).

Recall from section 3.2.3, that such a view takes the N400 to reflect predictive, anticipatory processes that need not exclusively be of semantic nature, but can be phonological as well (Praamstra & Stegeman, 1993; Dumay et al., 2001; DeLong et al., 2005; Lau et al., 2008; Robson et al., 2017). That is, our results suggest that semantic as well as phonological predictions are generated prior to bottom-up information becoming available. Frontal regions are suggested to be involved in the generation of expected information that drive top-down modulations of sensory processing (Desimone & Duncan, 1995) and may replace missing speech information (Shahin et al., 2009; Boulenger et al., 2011). Such a 'phonological illusion' may account for the findings reported in Jankowski et al. (1999) where the initial accent was perceived, even when its phonetic correlates were suppressed, and may account for the ERP modulations observed in the current study. In fact, because the acoustic manipulations in Jankowski and colleagues were different than the manipulations here (i.e. they had mainly manipulated the onset duration, with f_0—the modulated phonetic parameter in the current study—neutralized), the combined results further point to the phonetically-independent identity of the initial accent.

Moreover, we observed an interaction effect between semantic congruity and the presence of the initial accent, such that ±IA continued to modulate N400 amplitudes, but only when sentences were semantically incongruent (see figure 6.16). This suggest that when a word did not make sense in the semantic context of the sentences, listeners re-evaluated the phonological make-up of the word. So, our results underline that listeners had a phonological preference for words to be marked with the initial accent in their underlying stress pattern, in line with the findings presented in the previous sections.

6.4.3 Delayed semantic resolution

Visual inspection of the ERP waveforms further suggested a delay in N400 latency (although this latency difference was not significant) when semantically incongruent words had been presented without initial accent, indicating that, when words are presented without initial accent and thus mismatch the listener's metrical anticipation, semantic conflict resolution starts later. Our behavioral results are in line with this interpretation. The results in response latencies showed a main effect of IA, such that when words were presented without initial accent, participants were slower to respond than when they had been presented with initial accent. This, indeed, suggests that semantic ambiguities were resolved after participants had attended to the metrical hindrance when words were presented without their expected stress template.

We also obtained a main effect of ±IA on error rates, such that listeners made more errors when words had been presented without initial accent than when they had been presented with IA. Furthermore, listeners appeared to make most errors on sentences that were semantically congruent, but metrically unexpected, indicating that presenting the words without the initial accent misdirected the participants on the word' identity. This is in line with the results reported in Magne et al. (2007), wherein metrical *congruity* was manipulated (i.e. the authors lengthened the medial syllable, a violation in French), while here we manipulated metrical *probability* (i.e. the presence of the initial accent). Whereas we predicted listeners to prefer words to be presented with initial accent, reducing its phonetic correlates did not create an illegal stress pattern. Still finding an effect of IA thus shows a *strong* expectation for the, allegedly, "secondary and *optional*" French accent.

Together with our ERP results on the semantic N400, the findings suggest a strong memory trace for the initial accent, such that lexical candidates matching the memory trace are easier to recognize, responded to faster and generate smaller N400s than when candidates are less easy to match (i.e. hold a less established memory trace). In other words, if listeners continuously predict upcoming speech input, they may have prepared for expected upcoming words by activating their expected phonological, metrical and semantic features from the mental lexicon (e.g. Lau et al., 2008). When all these features mismatched, reaction times were slowed down, and ERP amplitudes and latencies, which index prediction errors, increased.

IN SUM, we investigated the status of the French initial accent and its function in lexico-semantic processing. The initial accent was previously thought of as an optional and secondary accent in French, sub-serving the primary final accent in the marking of phrase boundaries. Previous ERP studies which also investigated the phonological status of IA (e.g. Astésano et al., 2013; Aguilera et al., 2014, and the studies presented in the previous sections), showed a phonological expectancy for IA and a disruption in pre-lexical stress processing when IA had been omitted. However, in the studies, words were presented in isolation, with IA in utterance initial position. Therefore, it had remained unclear whether the facilitatory effects of IA really applied to the lexical domain.

In the current study, the initial accent was not utterance initial but embedded in a sentence. We found the presence of IA to modulate the N400 not only in the fronto-central brain regions, but also in the centro-parietal regions. That is, when asking listeners to judge the semantic congruity of sentences that differed only in the explicit presence of the initial accent, lexico-semantic processing (as reflected by the N400) was still affected. Pre-lexical stress templates serve to access the mental lexicon. Our data demonstrate that presenting words without IA obstructs lexical access, which in turn, cascades up the process of speech comprehension to additionally hinder post-lexical processing.

In other words, French speech processing *naturally*, *automatically* and *crucially* engages metrical stress processing throughout comprehension.

Part IV

CONCLUSION AND OUTLOOK

The purpose of the current dissertation was to determine whether there is metrical stress in French and what role it plays during speech comprehension. Indeed, metrical stress is well-known to have an invaluable role during speech processing. Metrical stress serves as the anchor point of the intonation contour, and it is the regulating force behind rhythm. Additionally, metrical stress provides listeners with syllables that are perceptually stable, typically marks the boundaries of a word and, according to the Metrical Segmentation Strategy (MSS; Cutler & Norris, 1988; Cutler, 1990), the Attentional Bounce Hypothesis (ABH; Pitt & Samuel, 1990) and the Dynamic Attending Theory (DAT; Large & Jones, 1999), stress interplays with attention by attracting attention from bottom-up through its acoustic salience, as well as a priori harnessing attention top-down by means of its predictability.

The role of metrical stress in French speech processing had, however, been less straightforward. French is a language alleged without accent, which made it difficult to position the language within these theoretical frameworks. That is, French accentuation is commonly attributed a low phonological and post-lexical status because it is not lexically distinctive and frequently overlaps with intonation. The French primary accent (FA) is held to exclusively be located at the right phrasal boundary, such that it is perceptually overshadowed by higher level constituent marking. That is, phrase final accents acoustically merge with the intonation contour such that their phonetic parameters are spread and diluted over nearby syllables (i.e. “acoustic chiasm”, Fónagy, 1980), and phrase internal accents may be reduced in order to favor syntactic and semantic coherence within the phrase. Similarly, the French secondary and optional accent (IA) is held to serve only post-lexical functions in pragmatically expressing emotional or para-linguistic emphasis, or in the rhythmic structuration of the utterance, yielding to FA when not rhythmically necessary.

This means that, in French, the post-lexical re-structuration of the speech stream not only blurs local accents phonetically (i.e. at the surface level), but also functionally, which has given way to the notion that French listeners have no representation of stress in their mental lexicon, and, worse still, are unable to hear stress. As a result, French accentuation has attracted rather little interest in the linguistic field, with most descriptions of French prosody focusing on the tonal and intonational organization of the language, while leaving aside the characteristics of stress other than pitch movements as well as ignoring the possibility for lexical functions for stress during speech comprehension.

In the current dissertation, we sought to address this gap in the academic domain. We argued that the initial and final accents carry metrical weight and are represented in bipolar, cognitive stress templates underlying the lexical word (*cf.* Di Cristo, 1999). Such a metrical perspective allowed us to imagine a phonological role for stress in French speech processing. That is, in assuming IA and FA are metrically strong and mark both boundaries of the word, the accents are more readily integrated in theoretical frameworks on speech processing, according to which the accents are crucially involved in the analysis of speech.

In our investigation of the role of the metrical stress templates, we took a functional approach and examined whether presenting listeners with words without their two markers, hindered word-level processing. That is, we combined the method of Event-Related Potentials (ERP) with three paradigms (i.e. the oddball paradigm, the lexical decision paradigm, and the semantic anomaly paradigm) which allowed us to separate FA from its collaborative functions in the marking of the phrase, as well as demonstrate that IA is more than a rhythmic counterweight, or worse, merely a heavy emphatic stress, but, above all, structural in nature (*cf.* Astésano et al., 2007; Astésano & Bertrand, 2016; Astésano, 2017)

Instead, we showed the *metrical status* of the two French accents. That is, the results in each of our studies point to a cognitive representation and phonological anticipation for the French accents at the level of the word. Listeners consistently expected words to be marked with both accents at their lexical boundaries. The listener *naturally* and *automatically* extracts the metrical information, which s/he then uses during the pre-lexical, lexical and post-lexical processing stages of speech comprehension. The results therefore demonstrate the value of metrical stress processing in French and appeal for metrical stress to be given a more prominent place in the descriptions of French prosody.

Outlook

In the current work, the mechanism behind the contribution of metrical stress in the analysis of speech, is assumed to rely both on its bottom-up attentional grasp due to the acoustic salience and the top-down a priori attentional allocation due to its metric predictability. Such a view implies that syllables can be metrically strong even if they are not fully realized acoustically, analogous to the perception of a metrical beat underlying music. This mechanism was however not fully explored here since our metrical manipulation predominantly concentrated on the surface realization of the accents. Future studies are needed to more precisely determine to what extent French metrical stress predictively guides listeners during speech decoding.

Recall from chapter 3 that, indeed, studies in many cognitive domains (e.g. visual cognition, auditory cognition) show neural systems to chunk time; sampling, integrating and analyzing perceptual information in discontinuous time windows. According to the oscillation based functional model (e.g. Giraud & Poeppel, 2012), intrinsic oscillations at different frequency bands interact with neuronal activity generated by an incoming speech signal. This interaction is thought to heavily rely on the metrical and rhythmic structure underlying the utterances, and acts as a mechanism of sampling and packaging of the input spike trains, such that spike timing is phased-locked to generate a hierarchic organization of temporal windows. These hierarchically organized windows then embed the also hierarchical organization of linguistic units.

Most studies investigating this model focused only on two frequency bands: the gamma band (25 — 35 Hz) corresponding to phonemes and the theta band (4—8 Hz) corresponding to syllables. For instance, it has been suggested that gamma and theta rhythms work together, such that the phase of theta oscillations modulates the power and possibly also the phase of gamma oscillations during the analysis of auditory information (e.g. Schroeder & Lakatos, 2009). Speech-tracking studies investigating this type of nesting and neural alignment have shown that the phase of theta-band neural activity discriminates different sentences and may additionally be predictive of speech intelligibility (Luo & Poeppel, 2007; Gross et al., 2013; Peelle et al., 2013; Doelling et al., 2014; Hyafil et al., 2015; Park et al., 2015; Steinmetzger & Rosen, 2017, see however Zoefel & VanRullen 2015 for conflicting evidence). Such a tracking mechanism likely relies on predictive modulations and could be the mechanism behind the segmentation of speech.

Prosodic information is generally recognized to play an important role in speech segmentation and comprehension processes, and to provide language with much of its predictable properties. Prosody organizes syllables such that they may become part of larger structures of prominence networks, which could clearly benefit the neural tracking of speech if tracking were to rely on the predictive accents (i.e. metrical stress) as anchor points. Still, neural alignment to the higher levels of linguistic abstraction that correspond to time windows longer than the phoneme or syllable (stress patterns, words, prosodic phrases) has yet to be demonstrated.

Moreover, studies on neural alignment to speech are often ambiguous regarding to whether tracking relies on entrainment of internal oscillations and/or the predictive role of rhythm in speech processing. In other words, it has not yet been shown whether speech tracking indeed reflects a dynamic oscillatory mechanism, or whether it can also be explained in terms of a bottom-up/passive reflection of rhythmic modulations to the acoustic signal itself (e.g. Kösem et al., 2016; Haegens & Zion Golumbic, 2017). Distinguishing between the two explanations can only be done if one is to demonstrate top-down effects (i.e. goal directed, such as attention and linguistic processing) on neural speech tracking or decoupling speech acoustics from the linguistic or phonological information inside the signal.

For instance, future studies could investigate whether prior metrical regularity (either through a rhythmic prime or an artificially regular sentence context) or musical experience or competence, benefits behavioral performance in a linguistic task in French (see e.g. Cutler & Foss, 1977; Pitt & Samuel, 1990; Gow & Gordon, 1993; Quené & Port, 2005; Rothermich et al., 2010, 2012; Cason & Schön, 2012; Cason, 2013; Cason et al., 2015; Harding, 2016; Magne et al., 2016, for studies using these types of paradigms to show that speech processing indeed heavily relies on the perception of metrical structure). Such studies would provide more direct evidence that the perception of French metrical stress, requiring sensitivity to complex rhythmic structures, actively directs listeners in a forward looking manner through speech comprehension.

Further, speech tracking studies showing neural tracking to metric or pulse information not present in the signal may additionally provide support for a functional role of metrical stress in French comprehension processes. In a recent study, Meyer et al. (2017) demonstrated that tracking coincided with internal preferences to syntactic structures and phrasal boundaries, even in the presence of conflicting acoustic cues to phrasal boundaries in German (see also Kösem et al., 2016). Similarly, in a study presenting speech stimuli wherein the acoustic envelope only reflected syllable onsets but did not contain cues regarding lexical or phrasal boundaries, two delta band responses appeared to lock onto the word and phrase boundaries, but only when speech was intelligible (i.e. phase-locking disappeared to foreign speech) (Ding et al., 2016a). However, while these findings do relate neural alignment to higher levels of linguistic abstraction, it was related to online syntactic structure building, leaving open the question in what way whether they may have demonstrated the "phonological illusion" of prosodic boundaries or local prominences not present in the acoustic speech signal. That is, the results may also be explained in terms of the distinction between surface level acoustic manifestation and the perceptual effect of underlying phonological representations.

Another recent study, which links entrainment to top-down control conditioned on the temporal predictability in speech, manipulated rhythmic structure of English speech fragments by altering the distribution of pauses between syllables or words, and found that speech tracking along the prefrontal delta band was significantly reduced (Kayser et al., 2015). This suggests that at least some aspects of the speech tracking mechanism for speech, particularly in the lower frequencies attributed to prosody and attention, can be attributed to the inherent rhythmic structure of speech, and break down when this information deviates from expectation. It should be interesting to replicate these findings in French.

Additionally, studies could employ the cocktail party paradigm, describing an environment of several competing speakers, wherein one of the speakers has to be selected within the 'noise' of the other. This paradigm requires listeners to tap into top-down attentional processes, thereby allowing researchers to examine whether speech tracking can be modulated by top-down factors. The temporal precision of speech tracking in such a paradigm is thought to rely on extracting and utilizing predictive information from the speech signal (Zion Golumbic et al., 2013; Ding & Simon, 2014; Rimmele et al., 2015; Keitel et al., 2017). Therefore, this paradigm provides a way for studying prosodic parsing during speech comprehension. It could, for example, be interesting to see if musicians or individuals who are better at discriminating between complex rhythms, also exhibit better performance in selecting one speaker over the other.

Finally, this line of research could lead to a clearer understanding of the perceptual division of rhythmic structures between languages. For instance, while the oscillatory mechanism presumably applies cross-linguistically, there may be subtle differences in the respective dominance between oscillators. That is, while in traditionally classified stress based languages the foot- or stress-oscillator could be dominant and embed the syllable-oscillator while guiding the phrase-oscillator, the respective dominance may be different in languages traditional classified as syllable based. Perhaps in these languages, the syllable-oscillator takes the lead. Such a view could account for the perceptual distinction between syllable-timed and stress-timed languages and for the (subtle) differences obtained in segmentation studies.

UNTIL RECENTLY, the existence of metrical stress and its role in lexical processing was poorly defined in French. In the present work, we showed not only that accentuation *does* exist in French, but also that it likely applies to the *word* domain, is *expected* by listeners and is *actively* and *automatically* used in word-level processing. We hope that the present work will lead to a new appreciation for French metrical stress in the academic field, as it clearly lays way to compelling future work which will most certainly advance our understanding on the machinery behind the comprehension of speech.

BIBLIOGRAPHY

Abercombie, D. 1967. *Elements of general phonetics*. Aldine Pub. Company. Cited in 1.3 and 1.4.1.

Abercrombie, D. 1976. Stress and some other terms. *Work in Progress* 9.51–53. Cited in 1, 2, 1.1, 1.4, and 2.3.

Adda-Decker, M., Boula de Mareüil, P., Adda, G., & Lamel, L. 2005. Investigating syllabic structures and their variation in spontaneous French. *Speech Commun.* 46.119–139. Cited in 2.3.

Aguilera, M., El Yagoubi, R., Espesser, R., & Astésano, C. 2014. Event-Related Potential investigation of Initial Accent processing in French. *Speech Prosody 2014*, 383–387. Cited in 1.5.3, 1.6, 3.2.1, 4, 5.1.2.2, 6.1, 6.1.1.3, 6.1.2.2.2, 6.1.3, 6.1.3.2, 6.1.3.2.1, 6.1.3.2.2, 6.2.1.2.2, 6.3.2.2, 6.4.3, and C.1.1.

Allen, G. D. 1975. Speech rhythms: Its relation to performance universals and articulatory timing. *J. Phon.* Cited in 1.3.

Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A.-L. 2009. Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29.13445–13453. Cited in 3.1 and 3.3.

Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66.46–63. Cited in 1.4.3.

Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *J. Phon.* 40.351–373. Cited in 1.3 and 1.4.3.

Astésano, C. 1999. Levels of rhythmicity in French: a comparison between three speaking styles. *Actes de International Congress of Phonetic Sciences (San Francisco)*, 253–256. Cited in 1.3 and 6.3.

Astésano, C. 2001. *Rythme et accentuation en français: invariance et variabilité stylistique*. L'Harmattan.

Cited in 6, 1.4.3, 1.5.1, 1.5.2, 1.5.3, 2, 5.1.2.1, 5.1.3.1, 6.3, 6.3.2.2, and A.

Astésano, C. 2016. Prosodic Characteristics of Reference French. *Varieties of Spoken French*, ed. by Sylvain Detey, Jacques Durand, and Bernard Lyche Chantal, Laks.

Cited in I, 1.3, 1.5.1, 1.5.1, 18, 1.5.2, and A.

Astésano, C. 2017. *Le statut de l'Accent Initial dans la phonologie prosodique du français: enjeux descriptifs et psycholinguistiques*. Habilitat, UT2J.

Cited in I, 1.5.3, 1.5.3, 2, 2.6, 6.1, 6.3, IV, and A.

Astésano, C., Bard, E. G., & Turk, A. 2007. Structural influences on initial accent placement in French. *Lang. Speech* 50.423–446.

Cited in I, 1.5.1, 1.5.3, 4, 5.1.1, 5.1.2.1, 6.2.1.2.2, 6.2.3.5, 6.3, 6.3.2.2, 6.4.1, IV, and A.

Astésano, C. & Bertrand, R. 2016. Accentuation et niveaux de constituance en français : enjeux phonologiques et psycholinguistiques. *Langue française* NÂ° 191.11–30.

Cited in I, 1.5.3, 1.5.3, 6.1, 6.2.3.2, IV, and A.

Astésano, C., Bertrand, R., Espesser, R., & Nguyen, N. 2012. Perception des frontières et des proéminences en français. *Actes de la conférence conjointe JEP-TALN-RECITAL*, 353–360.

Cited in 1.5.3, 1.5.3, 1.6, 4, 6.1, 6.2.3.5, 6.3, and 6.4.1.

Astésano, C., El Yagoubi, R., & Aguilera, M. 2013. Processing of the Initial Accent by French listeners: investigation of pre-attentional and attentional processing in an oddball paradigm. AMLaP : Architecture and Mechanisms of Language Processing.

Cited in 1.5.3, 1.6, 3.2.1, 4, 6.1, 6.1.1.3, and 6.4.3.

Astésano, C., Magne, C., El Yagoubi, R., & Besson, M. 2004. Influence du rythme sur le traitement sémantique en français: approches comportementale et électrophysiologique. *JEP 2004, Fes, Maroc.*

Cited in 1.5.3, 3.2.3, and 6.3.

Astheimer, L. B. & Sanders, L. D. 2009. Listeners modulate temporally selective attention during natural speech processing. *Biol. Psychol.* 80.23–34.

Cited in 2.2.2.

Auer, P. 1993. Is a rhythm-based typology possible. *A study on the role of prosody in phonological typology. Hamburg: Germanisches Seminar*.

Cited in 1.3.

Aylett, M. & Turk, A. 2004. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech* 47.31–56.

Cited in 1.4, 1.4.3, and 2.5.

Baayen, R. H., Davidson, D. J., & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59.390–412.

Cited in 5.2.1.1 and 5.2.1.2.

Bagou, O., Fougeron, C., & Frauenfelder, U. H. 2002. Contribution of prosody to the segmentation and storage of words in the acquisition of a new mini-language. *Speech Prosody 2002, International Conference*.

Cited in 2.5, 2.6, 6.1.3.2.2, 6.2, and 6.3.

Banel, M.-H. & Bacri, N. 1994. On metrical patterns and lexical parsing in French. *Speech Commun.* 15.115–126.

Cited in 1.4.1, 2.5, 2.6, 6.1.3.2.2, 6.2, and 6.3.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* 68.

Cited in 5.2.1.1.

Bates, D., Maechler, M., & Bolker, B. 2012. lme4: linear mixed-effects models using S4 classes. R package version.

Cited in 5.2.1, 6.2.1.2.5, 6.2.2.2.5, and 6.3.2.5.1.

Beckman, M. E. 1996. The Parsing of Prosody. *Lang. Cogn. Process.* 11.17–68.

Cited in 1.2, 1.4, and 1.6.

Beckman, M. E. & Edwards, J. 1990. of prosodic constituency. *Between the grammar and physics of speech*, 152.

Cited in 1.1.

Beckman, M. E. & Edwards, J. 1994. Articulatory evidence for differentiating stress categories. *Papers in laboratory phonology III: Phonological structure and phonetic form*, 7–33.

Cited in 1.1.

Beckman, M. E. & Pierrehumbert, J. B. 1986. Intonational structure in Japanese and English. *Phonology* 3.255–309.

Cited in 1.1 and 1.1.

Beier, E. J. & Ferreira, F. 2018. The Temporal Prediction of Stress in Speech and Its Relation to Musical Beat Perception. *Front. Psychol.* 9.431.

Cited in 3.1.

Bentin, S., McCarthy, G., & Wood, C. C. 1985. Event-related potentials, lexical decision and semantic priming. *Electroencephalogr. Clin. Neurophysiol.* 60.343–355.

Cited in 3.2, 3.3, and A.3.

Blair, R. C. & Karniski, W. 1993. An alternative method for significance testing of waveform difference potentials. *Psychophysiology* 30.518–524.

Cited in 6.2.1.2.5 and 6.2.2.2.5.

Böcker, K. B., Bastiaansen, M. C., Vroomen, J., Brunia, C. H., & de Gelder,

B. 1999. An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology* 36.706–720. Cited in 3.2, 3.2.2, 3.3, 6.2, 6.2.1.2.5, 6.2.3, 6.2.3.2, 6.2.3.4, 6.2.3.5, 6.3.2.5.2, 6.4.1, A.3, C.3.1, C.3.1.5, C.3.3, and C.4.1.

Boersma, P. & Weenink, D. 2016. Praat software. University of Amsterdam. Cited in 5.1.2.2, 5.1.3.2, 6.1.1.2.2, 6.1.2.2.2, 6.2.1.2.2, and 6.3.2.2.

Bohn, K., Knaus, J., Wiese, R., & Domahs, U. 2013. The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia* 51.760–771. Cited in 2.6, 3.2, and 3.2.3.

Bolinger, D. L. 1958. A Theory of Pitch Accent in English. *Word World* 14.109–149. Cited in 1 and 2.

Bond, Z. S. & Garnes, S. 1980. Misperceptions of fluent speech. *Perception and production of fluent speech*. Cited in 1.1 and 1.4.2.

Boulenger, V., Hoen, M., Jacquier, C., & Meunier, F. 2011. Interplay between acoustic/phonetic and semantic processes during spoken sentence comprehension: an ERP study. *Brain Lang.* 116.51–63. Cited in 3.2.3 and 6.4.2.

Bourguignon, M., De Tiège, X., de Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., & Jousmäki, V. 2013. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Hum. Brain Mapp.* 34.314–326. Cited in 3.1.

Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. 2014. Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Lang Cogn Neurosci* 29.1132–1146. Cited in 1.4.2 and 2.2.2.

van den Brink, D., Brown, C. M., & Hagoort, P. 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *J. Cogn. Neurosci.* 13.967–985. Cited in 2.4, 3.2.3, and 6.4.2.

Briscoe, T. 1989. Lexical Access in Connected Speech Recognition. *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics*, ACL '89, 84–90. Stroudsburg, PA, USA: Association for Computational Linguistics. Cited in 2.3 and A.2.

Brown, C. & Hagoort, P. 1993. The processing nature of the n400: evidence from masked priming. *J. Cogn. Neurosci.* 5.34–44. Cited in 3.2, 3.2.3, 3.3, 6.2.3.4, 6.3.2.5.2, 6.4, 6.4.2, and A.3.

Buzsáki, G. 2009. Rhythms of The Brain. *ResearchGate*, xiv, 448 p. Oxford University Press. Cited in 2.5, 3.1, 3.3, and A.3.

Byrd, D. & Saltzman, E. 2003. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31.149–180. Cited in 1.4.3.

Cason, N. 2013. *L'effet du rythme musical sur la parole* dissertation. Cited in IV.

Cason, N., Astésano, C., & Schön, D. 2015. Bridging music and speech rhythm: rhythmic priming and audio-motor training affect speech perception. *Acta Psychol.* 155.43–50. Cited in IV.

Cason, N. & Schön, D. 2012. Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50.2652–2658. Cited in 1.4.2, 3.1, and IV.

Chait, M., Greenberg, S., Arai, T., Simon, J. Z., & Poeppel, D. 2015. Multi-time resolution analysis of speech: evidence from psychophysics. *Front. Neurosci.* 9.214. Cited in 2.6 and 3.1.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. 2004. Phonological phrase boundaries constrain lexical access I. Adult data. *J. Mem. Lang.* 51.523–547. Cited in 6.1.3.2.2 and 6.3.

Colin, C., Hoonhorst, I., Markessis, E., Radeau, M., de Tourtchaninoff, M., Foucher, A., Collet, G., & Deltenre, P. 2009. Mismatch negativity (MMN) evoked by sound duration contrasts: an unexpected major effect of deviance direction on amplitudes. *Clin. Neurophysiol.* 120.51–59. Cited in 6.1.1.2.2 and 6.1.2.2.2.

Connine, C. M. 2004. It's not what you hear but how often you hear it: on the neglected role of phonological variant frequency in auditory word recognition. *Psychon. Bull. Rev.* 11.1084–1089. Cited in 2.2.1.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. 2008. Processing variant forms in spoken word recognition: the role of variant frequency. *Percept. Psychophys.* 70.403–411. Cited in 2.2.1.

Connolly, J. F. & Phillips, N. A. 1994. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6.256–266. Cited in 3.2, 3.2.2, 3.3, 6.2, and A.3.

Content, A., Kearns, R. K., & Frauenfelder, U. H. 2001a. Boundaries versus Onsets in Syllabic Segmentation. *J. Mem. Lang.* 45.177–199. Cited in 2.2.2, 2.3, 2.6, and 6.2.

Content, A., Meunier, C., Kearns, R. K., & Frauenfelder, U. H. 2001b. Sequence detection in pseudowords in French: Where is the syllable effect? *Lang. Cogn. Process.* 16.609–636. Cited in 5.

Cope, T. E., Sohoglu, E., Sedley, W., Patterson, K., Jones, P. S., Wiggins, J., Dawson, C., Grube, M., Carlyon, R. P., Griffiths, T. D., Davis, M. H., & Rowe, J. B. 2017. Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nat. Commun.* 8.2154. Cited in 2.6 and 3.1.

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. 2013. Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *J. Neurosci.* 33.4002–4010. Cited in 3.1.

Cumming, R. E. 2010. *Speech rhythm: the language-specific integration of pitch and duration*. University of Cambridge dissertation. Cited in 1.3.

Cummins, F. & Port, R. 1998. Rhythmic constraints on stress timing in English. *J. Phon.* 26.145–171. Cited in 1.4.3, 2.5, 3.1, and 3.3.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. 2010. Words as anchors: known words facilitate statistical learning. *Exp. Psychol.* 57.134–141. Cited in 2.2.2.

Cunillera, T., Laine, M., & Rodríguez-Fornells, A. 2016. Headstart for speech segmentation: a neural signature for the anchor word effect. *Neuropsychologia* 82.189–199. Cited in 2.2.2.

Cutler, A. & Foss, D. J. 1977. On the role of sentence stress in sentence processing. *Lang. Speech* 20.1–10. Cited in 1, 1.1, 1.4.2, 2.5, and IV.

Cutler, A. & Norris, D. 1988. The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* Cited in I, 1.1, 1.4, 1.4.1, 1.4.3, 1.6, 2.2, 2.2.2, 2.2.2, 6.2, 6.2.3.2, IV, and A.1.

Cutler, A. 1976. Phoneme-monitoring reaction time as a function of preceding intonation contour. *Percept. Psychophys.* 20.55–60. Cited in 1.4.2, 2.5, and 3.1.

Cutler, A. 1990. *Exploiting prosodic probabilities in speech segmentation*. The MIT Press. Cited in I, 1.1, 1.4, 1.4.1, 1.4.3, 1.6, 2.2.2, 6.2, IV, and A.1.

Cutler, A. 2010. Abstraction-based Efficiency in the Lexicon. *Lab. Phonol.* 1.13. Cited in 1.1, 3, 1.4, 2.2, 2.2.1, 6.1, 6.2, and 6.3.

Cutler, A. & Butterfield, S. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *J. Mem. Lang.* 31.218–236.

Cited in 1.4.1, 1.4.3, 1.6, 6.1.3.2.2, and A.1.

Cutler, A. & Carter, D. M. 1987. The predominance of strong initial syllables in the English vocabulary. *Comput. Speech Lang.* 2.133–142.

Cited in 1.1, 1.4.1, 1.4.3, 1.6, 3, 2.2.2, 6.2, and A.1.

Cutler, A. & Clifton, Jr, C. 1984. The use of prosodic information in word recognition. *Attention and performance X: Control of language processes*, 183–196. Erlbaum.

Cited in 1.1 and 2.2.2.

Cutler, A., Mehler, J., Norris, D., & Segui, J. 1986. The syllable's differing role in the segmentation of French and English. *J. Mem. Lang.* 25.385–400.

Cited in 1.4.1, 2.3, 2.6, 6.2, and A.2.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. 2001. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Lang. Cogn. Process.* 16.507–534.

Cited in 2.2.2.

Dahan, D. & Tanenhaus, M. K. 2004. Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *J. Exp. Psychol. Learn. Mem. Cogn.* 30.498–513.

Cited in 2.4.

Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *J. Phon.*

Cited in 1.3 and 1.4.3.

Davis, M. H. 2003. Connectionist modelling of lexical segmentation and vocabulary acquisition. *Connectionist models of development:*.

Cited in 2.2.2.

Davis, M. H. & Scharenborg, O. 2016. Speech perception by humans and machines. *Speech Perception and Spoken*.

Cited in 2.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. 2002. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 28.218.

Cited in 2.

Deguchi, C., Chobert, J., Brunellière, A., Nguyen, N., Colombo, L., & Besson, M. 2010. Pre-attentive and attentive processing of French vowels. *Brain Res.* 1366.149–161.

Cited in 3.2.1.

Delattre, P. 1966. A combarison of syllable length conditioning among languages. *IRAL Int. Rev. Appl. Linguist. Lang. Teach.*

Cited in 1.3 and 1.5.1.

DeLong, K. A., Urbach, T. P., & Kutas, M. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8.1117–1121.
Cited in 3.2.3, 6.2.3.4, 6.3, 6.3.2.5.2, and 6.4.2.

Delorme, A. & Makeig, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134.9–21.
Cited in 6.1.1.2.4, 6.1.2.2.4, 6.2.1.2.4, 6.2.2.2.4, and 6.3.2.4.

Denham, S. L. & Winkler, I. 2017. Predictive coding in auditory perception: challenges and unresolved questions. *Eur. J. Neurosci.*
Cited in 3.2.1 and 6.1.

Desimone, R. & Duncan, J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18.193–222.
Cited in 6.4.2.

DeWitt, L. A. & Samuel, A. G. 1990. The role of knowledge-based expectations in music perception: evidence from musical restoration. *J. Exp. Psychol. Gen.* 119.123–144.
Cited in 3.

Di Cristo, A. & Hirst, D. 1993. Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix.*
Cited in 1.5.3.

Di Cristo, A. 1976. Indices prosodiques et structure constituante. *Cahiers de Linguistique, d'Orientalisme et de Slavistique* 7.27–40.
Cited in 1.2 and 2.2.1.

Di Cristo, A. 1999. Vers une modélisation de l'accentuation du français: première partie. *Journal of French language studies* 9.143–179.
Cited in I, 6, 1.5.1, 1.5.1, 1.5.1, 1.5.3, 1.5.3, 2, 2.6, 6.1, 6.1.3.2.2, 6.2, 6.3, IV, A, and A.1.

Di Cristo, A. 2000. Vers une modélisation de l'accentuation du français (seconde partie). *Journal of French language studies* 10.27–44.
Cited in I, 1.5.2, 1.3, 1.5.2, 1.5.3, 6.1, and 6.3.

Ding, N. & He, H. 2016. Rhythm of Silence. *Trends Cogn. Sci.* 20.82–84.
Cited in 3.1.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. 2016a. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19.158–164.
Cited in 2.6, 3.1, 3.3, and IV.

Ding, N., Patel, A., Chen, L., Butler, H., Luo, C., & Poeppel, D. 2016b. Temporal Modulations Reveal Distinct Rhythmic Properties of Speech and Music.
Cited in 1.3 and 2.

Ding, N. & Simon, J. Z. 2012a. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U. S. A.* 109.11854–11859. Cited in 3.1.

Ding, N. & Simon, J. Z. 2012b. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107.78–89. Cited in 3.1.

Ding, N. & Simon, J. Z. 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33.5728–5735. Cited in 3.1.

Ding, N. & Simon, J. Z. 2014. Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8.311. Cited in 3.1, 3.3, and IV.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85 Pt 2.761–768. Cited in 2.6, 3.1, and IV.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., & Schlesewsky, M. 2008. The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25.1–36. Cited in 3.2.

Donovan, A. & Darwin, C. J. 1979. The perceived rhythm of speech. *Proceedings of the ninth international congress of phonetic sciences*, vol. 2, 268–274. Cited in 1.3.

Dumay, N., Benraïss, A., Barriol, B., Colin, C., Radeau, M., & Besson, M. 2001. Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *J. Cogn. Neurosci.* 13.121–143. Cited in 3.2.3, 6.3, and 6.4.2.

Dumay, N., Frauenfelder, U. H., & Content, A. 2002. The role of the syllable in lexical segmentation in French: word-spotting data. *Brain Lang.* 81.144–161. Cited in 2.2.2, 2.3, 2.6, 6.2, 6.3, and A.2.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., Polich, J., Reinvang, I., & Van Petten, C. 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin. Neurophysiol.* 120.1883–1908. Cited in 3.2.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. 1997. A Destressing

"Deafness" in French? *J. Mem. Lang.* 36.406–421. Cited in 1.5, 17, and 6.1.

Eisner, F. & McQueen, J. M. 2018. Speech Perception. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, ed. by John T Wixted, vol. 49, 1–46. Hoboken, NJ, USA: John Wiley & Sons, Inc. Cited in 1.1, 2.2.1, and 2.4.

Engel, A. K., Fries, P., & Singer, W. 2001. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2.704–716. Cited in 3.3.

Ernestus, M. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142.27–41. Cited in 2.2.1.

Ernestus, M., Baayen, H., & Schreuder, R. 2002. The recognition of reduced word forms. *Brain Lang.* 81.162–173. Cited in 2.2.1.

Eulitz, C. & Lahiri, A. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16.577–583. Cited in 1.1, 3, 1.4, 2.2, 2.2.1, 3.2.1, 6.1, 6.2, and 6.3.

Falk, S. & Dalla Bella, S. 2016. It is better when expected: aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience* 31.699–708. Cited in 1.4.2 and 3.1.

Fant, G., Kruckenberg, A., & Nord, L. 1991. Durational correlates of stress in Swedish, French, and English. *J. Phon.* Cited in 1.3, 8, 2.5, 2, and 6.3.

Féry, C. 2016. *Intonation and Prosodic Structure by Caroline Féry*. Cambridge University Press. Cited in 1.2.

Fletcher, J. 2010. The Prosody of Speech: Timing and Rhythm. *The Handbook of Phonetic Sciences*, ed. by William J Hardcastle, John Laver, and Fiona E Gibbon, 521–602. Oxford, UK: Blackwell Publishing Ltd. Cited in 1.3 and 6.

Fónagy, I. 1980. L'accent en français: accent probabilitaire. *L'accent en francÌ§ais contemporain (Studia Phonetica)*, ed. by I Fónagy and P Leon, vol. 15, 123–233. Cited in 1.5.1, 1.5.1, 1.5.3, and IV.

Fougeron, C. & Keating, P. A. 1997. Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am.* 101.3728–3740. Cited in 8 and 2.5.

Fox, A. 2000. *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. Oxford University Press. Cited in 1, 2, 1.2, 1.3, 1.4, 1.5.2, and 3.1.

Fraisse, P. 1982. Rhythm and tempo. *The psychology of music* 1.149–180.

Cited in 1.3.

Friston, K. 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360.815–836.

Cited in 2.2.1, 2.5, 3.2, 6.1, 6.2.3.2, and A.3.

Frota, S. 2012. Prosodic structure, constituents and their representations. *The Oxford handbook of laboratory phonology*, 255–265.

Cited in 1.2, 1.4, and 3.1.

Garami, L., Ragó, A., Honbolygó, F., & Csépe, V. 2017. Lexical influence on stress processing in a fixed-stress language. *Int. J. Psychophysiol.* 117.10–16.

Cited in 3.2.1 and 6.1.3.2.2.

Garde, P. 1968. *L'accent*, vol. 5. Presses Univ. de France.

Cited in 1.5.1 and 1.5.1.

Garnier, L. 2018. *Quels liens entre accentuation et niveaux de constituance en français? Une analyse perceptive et acoustique*. UT2J, Toulouse dissertation.

Cited in 1.5.3, 1.5.3, 1.6, 4, 6.1, 6.2.3.5, 6.3, and 6.4.1.

Garnier, L., Baqué, L., Dagnac, A., & Astésano, C. 2016. Perceptual investigation of prosodic phrasing in French. *Speech Prosody 2016*.

Cited in 1.5.3, 1.6, 4, 6.1, 6.2.3.5, 6.3, and 6.4.1.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. 2009. The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120.453–463.

Cited in 3.2.1 and 6.1.

Ghitza, O. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2.130.

Cited in 2.5, 3.1, 3.3, and A.2.

Ghitza, O. 2013. The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4.138.

Cited in 2.5, 3.1, and A.2.

Ghitza, O. 2014. Behavioral evidence for the role of cortical ϑ oscillations in determining auditory channel capacity for speech. *Front. Psychol.* 5.652.

Cited in 3.1.

Ghitza, O. 2016. Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience* 32.545–561.

Cited in 3.1, 3.1, and 3.3.

Ghitza, O., Giraud, A.-L., & Poeppel, D. 2012. Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Front. Hum. Neurosci.* 6.340.

Cited in 2.5, 2.6, and 3.1.

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* 66.113–126.

Cited in 3.1 and 3.3.

Gilbert, R. 2014. *Temporal properties of rehearsal in auditory-verbal short-term memory*. University of York dissertation.

Cited in 6.4.2.

Giraud, A.-L. & Poeppel, D. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15.511–517.

Cited in 1.4.3, 2.5, 2.6, 3.1, 3.1, 3.3, IV, A.2, and A.3.

Gordon, M. 2014. Disentangling stress and pitch-accent: a typology of prominence at different prosodic levels. *Word stress: Theoretical and typological issues*, 83.

Cited in 1.2 and 1.2.

Goswami, U. 2017. A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Curr. Dir. Psychol. Sci.* 27.56–63.

Cited in 3.1.

Goswami, U., Cumming, R., Chait, M., Huss, M., Mead, N., Wilson, A. M., Barnes, L., & Fosker, T. 2016. Perception of Filtered Speech by Children with Developmental Dyslexia and Children with Specific Language Impairments. *Front. Psychol.* 7.791.

Cited in 3.1.

Gow, Jr, D. W. & Gordon, P. C. 1993. Coming to terms with stress: effects of stress location in sentence processing. *J. Psycholinguist. Res.* 22.545–578.

Cited in IV.

Grabe, E. & Low, E. L. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology* 7.

Cited in 1.3.

Greenberg, S. & Arai, T. 2004. What are the essential cues for understanding spoken language? *IEICE Trans. Inf. Syst.*, 1059–1070.

Cited in 3.1.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. *J. Phon.* 31.465–485.

Cited in 3.1.

Groppe, D. M., Urbach, T. P., & Kutas, M. 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48.1711–1725.

Cited in 5.2.2.3, 5.2.2.3, 6.1.1.2.5, 6.1.2.2.5, 6.2.1.2.5, 6.2.2.2.5, and 6.3.2.5.2.

Grosjean, F. 1980. Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* 28.267–283. Cited in 6.2.3.2.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11.e1001752. Cited in 1.4.3, 2.5, 3.1, IV, and A.2.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press. Cited in 4.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. 2018. In Spoken Word Recognition, the Future Predicts the Past. *J. Neurosci.* 38.7585–7599. Cited in 2.2.2.

Haegens, S. & Zion Golumbic, E. 2017. Rhythmic facilitation of sensory processing: A critical review. *Neurosci. Biobehav. Rev.* 86.150–165. Cited in 3.1 and IV.

Haken, H., Kelso, J. A., & Bunz, H. 1985. A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* 51.347–356. Cited in 1.4.3.

Harding, E. E. 2016. *Neurocognitive entrainment to meter influences syntactic comprehension in music and language: An individual-differences approach*. University of Potsdam dissertation. Cited in 1.4.2, 3.1, 3.2, and IV.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6.e4794. Cited in 5.2.1.1, 5.2.1.2, and 3.

Hayes, B. 1989. THE PROSODIC HIERARCHY IN METER. *Rhythm and Meter*, ed. by Paul Kiparsky and Gilbert Youmans, 201–260. Academic Press. Cited in 1.2, 1.1, and 10.

Henry, M. J. & Herrmann, B. 2012. A precluding role of low-frequency oscillations for auditory perception in a continuous processing mode. *J. Neurosci.* 32.17525–17527. Cited in 1.4.3.

Henry, M. J. & Herrmann, B. 2014. Low-Frequency Neural Oscillations Support Dynamic Attending in Temporal Context. *Timing & Time Perception* 2.62–86. Cited in 1.4.3.

Henry, M. J., Herrmann, B., & Obleser, J. 2014. Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U. S. A.* 111.14935–14940.

Cited in 1.4.3, 3.1, 3.1, and 3.3.

Henry, M. J. & Obleser, J. 2012. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U. S. A.* 109.20095–20100.

Cited in 3.1 and 3.3.

Herbst, S. K. & Landau, A. N. 2016. Rhythms for cognition: the case of temporal processing. *Current Opinion in Behavioral Sciences* 8.85–93.

Cited in 3.1.

Hickok, G. & Poeppel, D. 2015. Neural basis of speech perception. *Handb. Clin. Neurol.* 129.149–160.

Cited in 2.1 and 2.

Hickok, G. & Small, S. L. 2015. *Neurobiology of Language*. Elsevier.

Cited in 2.1 and 3.1.

Hillyard, S. A. & Picton, T. W. 1987. Electrophysiology of cognition. *Handbook of physiology* 5.519–584.

Cited in 6.2.3.2 and 6.3.3.2.1.

Honbolygó, F. & Csépe, V. 2013. Saliency or template? ERP evidence for long-term representation of word stress. *Int. J. Psychophysiol.* 87.165–172.

Cited in 3.2.1, 6.1.1.2.2, and 6.1.3.2.2.

Honbolygó, F., Csépe, V., & Ragó, A. 2004. Suprasegmental speech cues are automatically processed by the human brain: a mismatch negativity study. *Neurosci. Lett.* 363.84–88.

Cited in 3.2.1 and 6.1.3.2.2.

Honbolygó, F., Kolozsvári, O., & Csépe, V. 2017. Processing of word stress related acoustic information: A multi-feature MMN study. *Int. J. Psychophysiol.* 118.9–17.

Cited in 3.2.1, 6.1.1.2.2, 6.1.2.2.2, and 6.1.3.2.2.

van der Hulst, H. 2014. *Word Stress: Theoretical and Typological Issues*. Cambridge University Press.

Cited in 1, 1.2, 1.2, 1.4, and 3.1.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A.-L. 2015. Speech encoding by coupled cortical theta and gamma oscillations. *Elife* 2015.e06213.

Cited in 3.1 and IV.

Hyman, L. 1977. On the nature of linguistic stress. *Studies in stress and accent* 4.

Cited in 1.1.

Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., & Winkler, I.

2004. Pre-attentive auditory processing of lexicality. *Brain Lang.* 88.54–67.
Cited in 6.1.3.2.2.

Jacobsen, T. & Schröger, E. 2003. Measuring duration mismatch negativity. *Clin. Neurophysiol.* 114.1133–1143.
Cited in 6.1.1.2.2 and 6.1.2.2.2.

Jankowski, L., Astésano, C., & Di Cristo, A. 1999. The initial rhythmic accent in French: Acoustic data and perceptual investigation. *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 1, 257–260.
Cited in 1.5.3, 1.6, 3.1, 6.1, 6.3, and 6.4.2.

Jassem, W. & Gibbon, D. 1980. Re-defining English accent and stress. *J. Int. Phon. Assoc.* 10.2–16.
Cited in 1 and 2.

Jensen, O. & Mazaheri, A. 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4.186.
Cited in 3.3.

Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, 145–165.
Cited in 2.2.

Jones, M. R. 1976. Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychol. Rev.* 83.323–355.
Cited in 1.4.3, 2.5, 3.1, 3.1, and 3.3.

Jones, M. R. & Boltz, M. 1989. Dynamic attending and responses to time. *Psychol. Rev.* 96.459–491.
Cited in 3.1.

Jones, M. R. 2010. Attending to sound patterns and the role of entrainment. *Attention and time*, 317–330.
Cited in 1.4.3, 1.4.3, 1.6, and A.1.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. 2002. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol. Sci.* 13.313–319.
Cited in 1.4.3, 1.4.3, 1.6, 2.5, 3.1, 3.1, 3.3, and A.1.

Jun, S.-A. 2007. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press.
Cited in 1.1.

Jun, S.-A. & Fougeron, C. 2000. A Phonological Model of French Intonation. *Intonation: Analysis, Modelling and Technology*, ed. by Antonis Botinis, Text, Speech and Language Technology, 209–242. Dordrecht: Springer Netherlands.
Cited in 1.5.1, 1.5.3, 2.6, and 6.2.

Kahn, D. 1980. *Syllable-based generalizations in English phonology*. New York, Garland.

Cited in 2.3.

Kayser, S. J., Ince, R. A. A., Gross, J., & Kayser, C. 2015. Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. *J. Neurosci.* 35.14691–14701.

Cited in 3.1 and IV.

Keitel, A., Gross, J., & Kayser, C. 2017. Speech tracking in auditory and motor regions reflects distinct linguistic features.

Cited in 2.6 and IV.

Kim, D., Stephens, J. D. W., & Pitt, M. A. 2012. How does context play a part in splitting words apart? Production and perception of word boundaries in casual speech. *J. Mem. Lang.* 66.509–529.

Cited in 2.4.

Klatt, D. H. 1976. Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59.1208–1221.

Cited in 5.1.3.2, 6.1.1.2.2, and 6.1.2.2.2.

Kösem, A., Basirat, A., Azizi, L., & van Wassenhove, V. 2016. High-frequency neural activity predicts word parsing in ambiguous speech streams. *J. Neurophysiol.* 116.2497–2512.

Cited in 3.1 and IV.

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., & Connolly, J. F. 2004. Activation in the anterior left auditory cortex associated with phonological analysis of speech input: localization of the phonological mismatch negativity response with MEG. *Brain Res. Cogn. Brain Res.* 21.106–113.

Cited in 3.2.2 and 6.2.

Kuperberg, G. R. & Jaeger, T. F. 2016. What do we mean by prediction in language comprehension? *Language, cognition and neuroscience*.

Cited in 6.

Kutas, M. & Hillyard, S. A. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207.203–205.

Cited in 3.2, 3.2.3, 6.4, and 6.4.2.

Kutas, M. & Federmeier, K. D. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62.621–647.

Cited in 3.2, 3.2.3, 3.3, 6.2.3.4, 6.3.2.5.2, 6.4.2, and A.3.

Ladd, D. R. 1986. Intonational Phrasing: The Case for Recursive Prosodic Structure. *Phonology Yearbook* 3.311–340.

Cited in 4.

Ladd, D. R. 2008. *Intonational Phonology*. Cambridge University Press.

Cited in 5.1.2.1, 5.1.2.2, and 6.3.2.2.

Large, E. W. 2008. Resonating to musical rhythm: theory and experiment. *The psychology of time*, 189–232. Cited in 1.3, 1.4.3, and 3.1.

Large, E. W. & Jones, M. R. 1999. The dynamics of attending: How people track time-varying events. *Psychol. Rev.* 106.119. Cited in I, 1.1, 1.3, 1.4, 1.4.3, 1.4.3, 1.6, 2.5, 2.6, 3.1, 3.1, 3.3, 6.2.3.2, 6.3, IV, and A.1.

Large, E. W. & Snyder, J. S. 2009. Pulse and meter as neural resonance. *Ann. N. Y. Acad. Sci.* 1169.46–57. Cited in 1.3, 1.4.3, and 2.5.

Lau, E. F., Phillips, C., & Poeppel, D. 2008. A cortical network for semantics: (de)constructing the N400. *Nat. Rev. Neurosci.* 9.920–933. Cited in 3.2.3, 6.2.3.4, 6.3.2.5.2, 6.4.2, and 6.4.3.

Laver, J. & John, L. 1994. *Principles of Phonetics*. Cambridge University Press. Cited in 1 and 2.

Lehiste, I. 1973. Rhythmic units and syntactic units in production and perception. *J. Acoust. Soc. Am.* 54.1228–1234. Cited in 1.1, 1.3, 6, and 1.5.3.

Liberman, M. & Prince, A. 1977. On Stress and Linguistic Rhythm. *Linguist. Inq.* 8.249–336. Cited in 1.3, 10, and 1.5.1.

Luce, P. A. & Pisoni, D. B. 1998. Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19.1–36. Cited in 2.2.2.

Luck, S., Huang, S., & Lopez-Calderon, J. 2010. *Erplab toolbox*. Cited in 6.1.1.2.4, 6.1.2.2.4, 6.2.1.2.4, 6.2.2.2.4, and 6.3.2.4.

Luck, S. 2014. The Mass Univariate Approach and Permutation Statistics. *ERP analysis*. Cited in 5.2.2.3, 5.2.2.3, 6.1.1.2.5, 6.1.2.2.5, 6.2.1.2.5, 6.2.2.2.5, and 6.3.2.5.2.

Luck, S. J. 2005. *Event-related potentials: a methods handbook*. MIT Press, Cambridge, Mass. Cited in 3.1, 3.2, 5.2.2.2, and 5.2.2.2.

Luo, H. & Poeppel, D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54.1001–1010. Cited in IV.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. 2007. Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cereb. Cortex* 17.2659–2668. Cited in 1.1, 1.5.3, 3.2, 3.2.3, 5.1.1, 6.1, 6.1.3.2.2, 6.3, 6.3.2.2, 6.3.2.5.2, 6.3.3.1.1, 6.3.3.1.2, and 6.4.3.

Magne, C., Jordan, D. K., & Gordon, R. L. 2016. Speech rhythm sensitivity

and musical aptitude: ERPs and individual differences. *Brain Lang.* 153-154.13–19.
Cited in IV.

Mai, G., Minett, J. W., & Wang, W. S.-Y. 2016. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *Neuroimage* 133.516–528.
Cited in 2.6 and 3.1.

Makeig, S., Bell, A. J., Jung, T.-P., & Sejnowski, T. J. 1996. Independent Component Analysis of Electroencephalographic Data. *Advances in Neural Information Processing Systems 8*, ed. by D S Touretzky, M C Mozer, and M E Hasselmo, 145–151. MIT Press.
Cited in 5.2.2.2.

Manly, B. F. J. 2006. *Randomization, Bootstrap and Monte Carlo Methods in Biology, Third Edition*. CRC Press.
Cited in 1.

Marie, C., Magne, C., & Besson, M. 2011. Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23.294–305.
Cited in 3.2.3.

Maris, E. & Oostenveld, R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164.177–190.
Cited in 5.2.2.3.

Marslen-Wilson, W., Moss, H. E., & van Halen, S. 1996. Perceptual distance and competition in lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 22.1376–1392.
Cited in 2.2.2 and A.2.

Marslen-Wilson, W. & Warren, P. 1994. Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychol. Rev.* 101.653–675.
Cited in 2.2.2.

Marslen-Wilson, W. D. & Welsh, A. 1978. Processing interactions and lexical access during word recognition in continuous speech. *Cogn. Psychol.* 10.29–63.
Cited in 2, 2.2, 2.2.2, 2.2.2, 6, 2.6, 6.1.3.2.2, and A.2.

Martin, A. E. 2016. Language Processing as Cue Integration: Grounding the Psychology of Language in Perception and Neurophysiology. *Front. Psychol.* 7.120.
Cited in 3.1.

Martin, J. G. 1972. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychol. Rev.* 79.487–509.
Cited in 1.4.3 and 2.5.

Mathworks, I. 2014. MATLAB: R2014a.
Cited in 6.1.1.2.4, 6.1.1.2.5, 6.1.2.2.4, 6.1.2.2.5, 6.2.1.2.4, 6.2.2.2.4, and 6.3.2.4.

Mattys, S. L. 2004. Stress versus coarticulation: toward an integrated approach to explicit speech segmentation. *J. Exp. Psychol. Hum. Percept. Perform.* 30.397–408. Cited in 8.

Mattys, S. L. & Samuel, A. G. 1997. How Lexical Stress Affects Speech Segmentation and Interactivity: Evidence from the Migration Paradigm. *J. Mem. Lang.* 36.87–116. Cited in 2.2.2.

Mattys, S. L., White, L., & Melhorn, J. F. 2005. Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.* 134.477–500. Cited in (document), 1.1, 2.2.2, 2.4, 2.3, and 2.4.

McClelland, J. L. & Elman, J. L. 1986. The TRACE model of speech perception. *Cogn. Psychol.* 18.1–86. Cited in 2, 2.2, 2.2.2, 2.2.2, 3, 2.6, and A.2.

McClelland, J. L. & Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88.375–407. Cited in 2.2.2.

McQueen, J. M. 2007. Eight questions about spoken-word recognition. *The Oxford handbook of psycholinguistics*, 37–53. Cited in 2.2.2.

McQueen, J. M., Jesse, A., & Norris, D. 2009. No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *J. Mem. Lang.* 61.1–18. Cited in 2.4.

McQueen, J. M., Norris, D., & Cutler, A. 1999. Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *J. Exp. Psychol. Hum. Percept. Perform.* 25.1363. Cited in 2.2.2.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20.298–305. Cited in 1.4.1, 2.3, 2.6, and 6.2.

Mertens, P. 1993. Intonational grouping, boundaries, and syntactic structure in French. *Proceedings of an ESCA Workshop on Prosody*, 156–159. Lund University, Department of Linguistics. Cited in 1.5.1.

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. 2017. Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations. *Cereb. Cortex* 27.4293–4302. Cited in 3.1, 3.3, and IV.

Michelas, A., Esteve-Gibert, N., & Dufour, S. 2018. On French listeners' ability to use stress during spoken word processing. *J. Cogn. Psychol.* 30.198–206.

Cited in 1.5.3 and 6.1.

Michelas, A., Frauenfelder, U. H., Schön, D., & Dufour, S. 2016. How deaf are French speakers to stress? *J. Acoust. Soc. Am.* 139.1333–1342.

Cited in 1.5.3 and 6.1.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. 2016. Out-of-synchrony speech entrainment in developmental dyslexia. *Hum. Brain Mapp.* 37.2767–2783.

Cited in 3.1.

Morillon, B., Schroeder, C. E., & Wyart, V. 2014. Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* 5.5255.

Cited in 3.1 and 3.3.

Morton, J., Marcus, S., & Frankish, C. 1976. Perceptual centers (P-centers). *Psychol. Rev.* 83.405.

Cited in 1.3.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385.432–434.

Cited in A.3.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118.2544–2590.

Cited in 1.5.3, 3.2, 3.2.1, 3.3, 6.1, and 6.1.3.2.1.

Nespor, M. & Vogel, I. 1983. Prosodic Structure Above the Word. *Prosody: Models and Measurements*, ed. by Anne Cutler and D Robert Ladd, Springer Series in Language and Communication, 123–140. Berlin, Heidelberg: Springer Berlin Heidelberg.

Cited in 1.1 and 1.5.1.

New, B., Pallier, C., Ferrand, L., & Matos, R. 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE//A lexical database for contemporary french: LEXIQUE. *L'année psychologique* 101.447–462.

Cited in 5.1.1.

Newman, R. L. & Connolly, J. F. 2009. Electrophysiological markers of prelexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biol. Psychol.* 80.114–121.

Cited in 3.2.2 and 6.2.

Nguyen, N. 2012. Representations of speech sound patterns in the speaker's brain: Insights from perception studies. *Oxford handbook of laboratory phonology*, 359–368. Cited in 2.2.1.

Nguyen, N., Wauquier, S., & Tuller, B. 2009. The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories. *Approaches to phonological complexity*, 5–31. Cited in 2.2.1.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohut, Z., Kulakova, E., Meziere, D., Politzer-Ahles, S., Rousselet, G., Rueschemeyer, S.-A., Segaert, K., Tuomainen, J., & Von Grebmer Zu Wolfsthurn, S. 2018. Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. Cited in 6.

Nolan, F. & Asu, E. L. 2009. The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66.64–77. Cited in 1.4.3.

Norris, D., McQueen, J. M., & Cutler, A. 1995. Competition and segmentation in spoken-word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 21.1209–1228. Cited in 1.4.1, 2.2.2, and 2.6.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. 1997. The possible-word constraint in the segmentation of continuous speech. *Cogn. Psychol.* 34.191–243. Cited in 2.2.2.

Norris, D. 1994. Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52.189–234. Cited in 2, 2.2, 2.2.2, and A.2.

Norris, D. & McQueen, J. M. 2008. Shortlist B: a Bayesian model of continuous speech recognition. *Psychol. Rev.* 115.357–395. Cited in 2, 2.2, 2.2.2, 2.2.2, 2.6, and A.2.

Palmer, S. D., Hutson, J., White, L., & Mattys, S. L. 2018. Lexical knowledge boosts statistically-driven speech segmentation. *J. Exp. Psychol. Learn. Mem. Cogn.* Cited in 2.2.2.

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25.1649–1653. Cited in IV.

Partanen, E., Vainio, M., Kujala, T., & Huotilainen, M. 2011. Linguistic multi-

feature MMN paradigm for extensive recording of auditory discrimination profiles. *Psychophysiology* 48.1372–1380.
Cited in 6.1.3.2.1.

Peelle, J. E. & Davis, M. H. 2012. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* 3.320.
Cited in 2.5, 3.1, 3.3, and A.2.

Peelle, J. E., Gross, J., & Davis, M. H. 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23.1378–1387.
Cited in IV.

Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Typological studies in language* 45.137–158.
Cited in 3, 2.2, and 2.2.1.

Pike, K. L. 1945. *The Intonation of American English*. University of Michigan Press, Ann Arbor, Mich.
Cited in 1.1, 1.3, 1.4.1, and 2.3.

Pitt, M. A., Dilley, L., & Tat, M. 2011. Exploring the role of exposure frequency in recognizing pronunciation variants. *J. Phon.* 39.304–311.
Cited in 2.2.1.

Pitt, M. A. & Samuel, A. G. 1990. The use of rhythm in attending to speech. *J. Exp. Psychol. Hum. Percept. Perform.* 16.564–573.
Cited in I, 1.1, 1.4, 1.4.2, 1.4.3, 1.6, 2.5, 2.6, 3.1, 3.3, 6.2, 6.2.3.2, 6.3, IV, IV, and A.1.

Poeppel, D., Idsardi, W. J., & van Wassenhove, V. 2008. Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363.1071–1086.
Cited in 2.1, 2.5, 2.6, 3.1, 3.1, 3.3, and A.3.

Port, R. F. 2003. Meter and speech. *J. Phon.* 31.599–611.
Cited in 1.4.3, 16, 2.5, and 3.1.

Post, B. M. B. 2000. *Tonal and phrasal structures in French intonation*. Katholieke universiteit Nijmegen dissertation.
Cited in 1.4.

Power, A. J., Mead, N., Barnes, L., & Goswami, U. 2013. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 7.777.
Cited in 3.1.

Praamstra, P. & Stegeman, D. F. 1993. Phonological effects on the auditory N400 event-related brain potential. *Brain Res. Cogn. Brain Res.* 1.73–86.
Cited in 3.2.3, 6.3, and 6.4.2.

Pulvermüller, F. & Shtyrov, Y. 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* 79.49–71.
Cited in 6.1.1.2.2 and 6.1.3.2.2.

Quené, H. & Koster, M. L. 1998. Metrical segmentation in Dutch: vowel quality or stress? *Lang. Speech* 41 (Pt 2).185–202.

Cited in 2.4.

Quené, H. & Port, R. F. 2005. Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62.1–13.

Cited in 1.1, 1.4.2, 3.1, and IV.

Radeau, M. & Morais, J. 1990. The uniqueness point effect in the shadowing of spoken words. *Speech Commun.* 9.155–164.

Cited in 2.1 and 2.2.2.

Ramus, F., Nespor, M., & Mehler, J. 2000. Correlates of linguistic rhythm in the speech signal. *Cognition* 75.AD3–AD30.

Cited in 1.3.

te Rietmolen, N., El Yagoubi, R., Espesser, R., Magnen, C., & Astesano, C. 2016. Investigating the phonological status of the Initial Accent in French: an Event-Related Potentials study. *Speech Prosody 2016*.

Cited in C.4.1.

Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68.144–154.

Cited in IV.

Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. *Linguistic controversies* 73.79.

Cited in 1.3.

Robson, H., Pilkington, E., Evans, L., DeLuca, V., & Keidel, J. L. 2017. Phonological and semantic processing during comprehension in Wernicke's aphasia: An N400 and Phonological Mapping Negativity Study. *Neuropsychologia* 100.144–154.

Cited in 3.2.3, 6.3, and 6.4.2.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. 2012. Temporal expectation improves the quality of sensory information. *J. Neurosci.* 32.8424–8428.

Cited in 3.1.

Rohenkohl, G. & Nobre, A. C. 2011. α oscillations related to anticipatory attention follow temporal expectations. *J. Neurosci.* 31.14076–14084.

Cited in 1.4.3, 3.1, and 3.3.

Rolland, G. & Lœvenbruck, H. 2002. Characteristics of the accentual phrase in French: An acoustic, articulatory and perceptual study. *Speech Prosody 2002, International Conference*. isca-speech.org.

Cited in 1.5.1, 2.5, 2.6, 6.2, and 6.3.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. 2013. Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One* 8.e56000.

Cited in 1.4.2.

Rosen, S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 336.367–373.

Cited in 3.1, 3.1, and A.3.

Rossi, M. 1980. Le français, langue sans accent? *L'accent en francÌ§ais contemporain (Studia Phonetica)*, ed. by I Fonagy and P Leon, vol. 15, 13–51.

Cited in I, 1.5, 1.5.1, 6.1, 6.3, A, and A.1.

Rossi, M. 1972. Le seuil différentiel de durée. *Papers in Linguistics and Phonetics to the memory of Pierre Delattre* 54.435.

Cited in 5.1.3.2, 6.1.1.2.2, and 6.1.2.2.2.

Rothermich, K. & Kotz, S. A. 2013. Predictions in speech comprehension: fMRI evidence on the meter-semantic interface. *Neuroimage* 70.89–100.

Cited in 1.1, 3.2.3, and 6.3.

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. 2012. Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia* 50.232–244.

Cited in 1.1, 1.4.2, 2.6, 3.1, 3.2, 3.2.3, 6.2.3.4, 6.3, 6.3.2.5.2, and IV.

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. 2010. Event-related potential responses to metric violations: rules versus meaning. *Neuroreport* 21.580–584.

Cited in 1.1, 1.4.2, 2.6, 3.1, 3.2, 3.2.3, 6.2.3.1, 6.2.3.4, 6.3, 6.3.2.5.2, and IV.

Roux, G., Bertrand, R., Ghio, A., & Astésano, C. 2016. Naïve listeners' perception of prominence and boundary in French spontaneous speech. *Speech Prosody 2016*.

Cited in 1.5.3.

Salverda, A. P., Dahan, D., & McQueen, J. M. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90.51–89.

Cited in 2.

Samuel, A. G. 1981. Phonemic restoration: insights from a new methodology. *J. Exp. Psychol. Gen.* 110.474–494.

Cited in 3.

Samuel, A. G. 1991. A further examination of attentional effects in the phonemic restoration illusion. *Q. J. Exp. Psychol. A* 43.679–699.

Cited in 3.

Samuel, A. G. & Ressler, W. H. 1986. Attention within auditory word perception: insights from the phonemic restoration illusion. *J. Exp. Psychol. Hum. Percept. Perform.* 12.70–79.

Cited in 3.

Samuel, A. G. 1996. Does lexical information influence the perceptual restoration of phonemes? *J. Exp. Psychol. Gen.* 125.28.

Cited in 3.

Sanders, L. D., Newport, E. L., & Neville, H. J. 2002. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* 5.700–703. Cited in 2.2.2.

Scharinger, M., Bendixen, A., Trujillo-Barreto, N. J., & Obleser, J. 2012. A sparse neural code for some speech sounds but not for others. *PLoS One* 7.e40953. Cited in 2.2.1 and 6.1.

Scharinger, M., Monahan, P. J., & Idsardi, W. J. 2016. Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *Neuroimage* 128.293–301. Cited in 2.2.1, 3.2, 3.2.1, 6.1, 6.2.3.2, and A.3.

Scharinger, M., Steinberg, J., & Tavano, A. 2017. Integrating speech in time depends on temporal expectancies and attention. *Cortex* 93.28–40. Cited in 3.2.1.

Schmidt-Kassow, M., Heinemann, L. V., Abel, C., & Kaiser, J. 2013. Auditory–motor synchronization facilitates attention allocation. *Neuroimage* 82.101–106. Cited in 3.1 and 3.3.

Schmidt-Kassow, M. & Kotz, S. A. 2008. Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Res.* 1226.144–155. Cited in 1.4.2, 3.1, and 3.2.

Schmidt-Kassow, M., Schubotz, R. I., & Kotz, S. A. 2009. Attention and entrainment: P3b varies as a function of temporal predictability. *Neuroreport* 20.31–36. Cited in 1.4.2, 3.1, 3.1, and 3.3.

Schroeder, C. E. & Lakatos, P. 2009. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32.9–18. Cited in 1.4.3, 3.1, 3.1, and IV.

Selkirk, E. 1996. The prosodic structure of function words. *Signal to syntax: Bootstrapping from speech to*. Cited in 4.

Selkirk, E. O. 1986. *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA, USA: MIT Press. Cited in 1.1.

Shahin, A. J., Bishop, C. W., & Miller, L. M. 2009. Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage* 44.1133–1143. Cited in 6.4.2.

Shattuck-Hufnagel, S. & Turk, A. E. 1996. A prosody tutorial for investigators of auditory sentence processing. *J. Psycholinguist. Res.* 25.193–247. Cited in 4, 1.2, 1.2, and 1.1.

Shatzman, K. B. & McQueen, J. M. 2006. Prosodic knowledge affects the recognition of newly acquired words. *Psychol. Sci.* 17.372–377.

Cited in 2.2.1 and 2.

Shoemaker, E. M. 2009. Acoustic cues to speech segmentation in spoken French: Native and non-native strategies.

Cited in 2.3.

Shukla, M., Nespor, M., & Mehler, J. 2007. An interaction between prosody and statistics in the segmentation of fluent speech. *Cogn. Psychol.* 54.1–32.

Cited in 2.5.

Singmann, H. & Kellen, D. 2017. An Introduction to Mixed Models for Experimental Psychology. *New Methods in Neuroscience and Cognitive Psychology*. Psychology Press Hove.

Cited in 5.2.1.1 and 5.2.1.2.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. 2012. Predictive top-down integration of prior knowledge during speech perception. *J. Neurosci.* 32.8443–8453.

Cited in 2.6 and 3.1.

Spinelli, E., Grimault, N., Meunier, F., & Welby, P. 2010. An intonational cue to word segmentation in phonemically identical sequences. *Atten. Percept. Psychophys.* 72.775–787.

Cited in 2.5, 2.6, 6.2, and 6.3.

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. *Lang. Cogn. Process.* 22.828–859.

Cited in 2.6 and 6.2.

Steinhauer, K. & Connolly, J. F. 2008. Event-Related Potentials in the Study of Language. *Concise Encyclopedia of Brain and Language*. Elsevier.

Cited in 3.2, 3.2.2, 3.3, 6.3.2.5.2, and A.3.

Steinmetzger, K. & Rosen, S. 2017. Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech. *Neuropsychologia* 95.173–181.

Cited in IV.

Sulpizio, S. & McQueen, J. M. 2012. Italians use abstract knowledge about lexical stress during spoken-word recognition. *J. Mem. Lang.* 66.177–193.

Cited in 2.2.1.

Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. 2014. The five myths of MMN: redefining how to use MMN in basic and clinical research. *Brain Topogr.* 27.553–564.

Cited in 6.1 and 6.1.3.2.1.

Sussman, E. 2007. A New View on the MMN and Attention Debate: The Role of Context in Processing Auditory Events. *J. Psychophysiol.* 21.164–175. Cited in 6.1 and 6.1.3.2.1.

Sussman, E. S. & Shafer, V. L. 2014. New perspectives on the mismatch negativity (MMN) component: an evolving tool in cognitive neuroscience. *Brain Topogr.* 27.425–427. Cited in 6.1 and 6.1.3.2.1.

Tajima, K. & Port, R. F. 2003. Speech rhythm in English and Japanese. *Phonetic interpretation: Papers in laboratory phonology VI*, 317–334. Cited in 1.4.3 and 3.1.

Tavano, A. & Scharinger, M. 2015. Prediction in speech and language processing. *Cortex* 68.1–7. Cited in 3.2 and 6.2.3.2.

Team, R. C. 2014. The R project for statistical computing. Cited in 5.2.1, 6.2.1.2.5, 6.2.2.2.5, and 6.3.2.5.1.

Ten Oever, S., van Atteveldt, N., & Sack, A. T. 2015. Increased Stimulus Expectancy Triggers Low-frequency Phase Reset during Restricted Vigilance. *J. Cogn. Neurosci.* 27.1811–1822. Cited in 3.1 and 3.3.

Ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., Mehta, A. D., Mégevand, P., Groppe, D. M., & Zion-Golumbic, E. 2017. Low-Frequency Cortical Oscillations Entrain to Subthreshold Rhythmic Auditory Stimuli. *J. Neurosci.* 37.4903–4912. Cited in 1.4.3, 3.1, and 3.3.

Tilsen, S. & Johnson, K. 2008. Low-frequency Fourier analysis of speech rhythm. *J. Acoust. Soc. Am.* 124.EL34–9. Cited in 1.4.3.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. 2010. Continuous perception and graded categorization: electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychol. Sci.* 21.1532–1540. Cited in 2.2.2.

Turk, A. & Shattuck-Hufnagel, S. 2013. What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Lab. Phonol.* Cited in 1.3, 1.4, 1.4.3, 2.5, and 3.1.

Turk, A. 2010. Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis. *Lab. Phonol.* 1.11. Cited in 1.4.3 and 2.5.

Turk, A. & Shattuck-Hufnagel, S. 2014. Timing in talking: what is it used for, and how is it controlled? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369.20130395. Cited in 1.4, 1.4.3, and 2.5.

Vaissière, J. 1991. Rhythm, accentuation and final lengthening in French. *Music, Language, Speech and Brain: Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5–8 September 1990*, ed. by Johan Sundberg, Lennart Nord, and Rolf Carlson, Wenner-Gren Center International Symposium Series, 108–120. London: Macmillan Education UK.

Cited in 6, 1.5.1, and 6.3.

Vaissiere, J. 1997. Langues, prosodies et syntaxe. *Revue Traitement Automatique des Langues, numéro spécial Prosodie et syntaxe*, 53–82.

Cited in 1.5.1.

Välimaa-Blum, R. 2009. The phoneme in cognitive phonology: episodic memories of both meaningful and meaningless units? *CogniTextes*.

Cited in 2.2.1.

VanRullen, R., Busch, N. A., Drewes, J., & Dubois, J. 2011. Ongoing EEG Phase as a Trial-by-Trial Predictor of Perceptual and Attentional Variability. *Front. Psychol.* 2.60.

Cited in 3.1.

VanRullen, R. 2016. Perceptual Cycles. *Trends Cogn. Sci.* 20.723–735.

Cited in 3.1 and 3.3.

VanRullen, R. & Koch, C. 2003. Is perception discrete or continuous? *Trends Cogn. Sci.* 7.207–213.

Cited in 3.1 and 3.3.

VanRullen, R., Zoefel, B., & Ilhan, B. 2014. On the cyclic nature of perception in vision versus audition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369.20130214.

Cited in 3.1.

Vroomen, J., van Zon, M., & de Gelder, B. 1996. Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Mem. Cognit.* 24.744–755.

Cited in 1.4.1, 1.4.3, 1.6, 2.3, 6.1.3.2.2, and A.1.

Vroomen, J. & de Gelder, B. 1995. Metrical segmentation and lexical inhibition in spoken word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 21.98–108.

Cited in 1.1, 1.4.1, 1.4.3, 1.6, 2.2.2, 2.4, 6.2, and A.1.

Wauquier-Gravelines, S. 1999. Segmentation lexicale de la parole continue : la linéarité en question. *Recherches linguistiques de Vincennes* nÂ° 28.8–8.

Cited in 2.3, 2.6, 6.3, and A.2.

Welby, P. & Lœvenbruck, H. 2006. Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics*.

Cited in 1.5.3.

Welby, P. 2007. The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Commun.* 49.28–48.

Cited in 2.5, 2.6, 6.2, and 6.3.

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 dissertation.

Cited in 1.5.1 and 1.5.3.

Wenk, B. J. & Wioland, F. 1982. Is French really syllable-timed? *J. Phon.*

Cited in 1.3.

White, L., Mattys, S. L., Stefansdottir, L., & Jones, V. 2015. Beating the bounds: Localized timing cues to word segmentation. *J. Acoust. Soc. Am.* 138.1214–1220.

Cited in 2.5.

White, L., Mattys, S. L., & Wiget, L. 2012. Segmentation cues in conversational speech: robust semantics and fragile phonotactics. *Front. Psychol.* 3.375.

Cited in 2.4.

Wilson, M. 1990. Activation, Competition, and Frequency in Lexical Access. *Cognitive Models of Speech Processing*, ed. by G T M Altmann, 148–172. The MIT Press.

Cited in 2, 2.1, 2.2, 2.2.2, 2.2.2, 6, 2.6, and 6.1.3.2.2.

Winkler, I., Debener, S., Müller, K.-R., & Tangermann, M. 2015. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015.4101–4105.

Cited in 5.2.2.2.

Winkler, I., Denham, S. L., & Nelken, I. 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn. Sci.* 13.532–540.

Cited in 3.2.1 and 6.1.

Winter, B. 2013. Linear models and linear mixed effects models in R with linguistic applications.

Cited in 5.2.1.1 and 5.2.1.2.

Yan, S., Kuperberg, G. R., & Jaeger, T. F. 2017. Prediction (Or Not) During Language Processing. A Commentary On Nieuwland et al.(2017) And Delong et al.(2005). *bioRxiv*.

Cited in 6, 6.2.3.4, and 6.3.2.5.2.

Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P. 2016. Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain Lang.* 162.72–80.

Cited in 3.2.1.

Ylinen, S., Strelnikov, K., Huotilainen, M., & Näätänen, R. 2009. Effects of prosodic familiarity on the automatic processing of words in the human brain. *Int. J. Psychophysiol.* 73.362–368.

Cited in 3.2.1 and 6.1.3.2.2.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77.980–991.

Cited in 2.6, 3.1, and IV.

Zoefel, B. & Heil, P. 2013. Detection of Near-Threshold Sounds is Independent of EEG Phase in Common Frequency Bands. *Front. Psychol.* 4.262.

Cited in 3.1.

Zoefel, B., Reddy Pasham, N., Brüers, S., & VanRullen, R. 2015. The ability of the auditory system to cope with temporal subsampling depends on the hierarchical level of processing. *Neuroreport* 26.773–778.

Cited in 3.1.

Zoefel, B. & VanRullen, R. 2015. The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Front. Hum. Neurosci.* 9.651.

Cited in IV.

Zoefel, B. & VanRullen, R. 2017. Oscillatory Mechanisms of Stimulus Processing and Selection in the Visual and Auditory Systems: State-of-the-Art, Speculations and Suggestions. *Front. Neurosci.* 11.296.

Cited in 3.1.

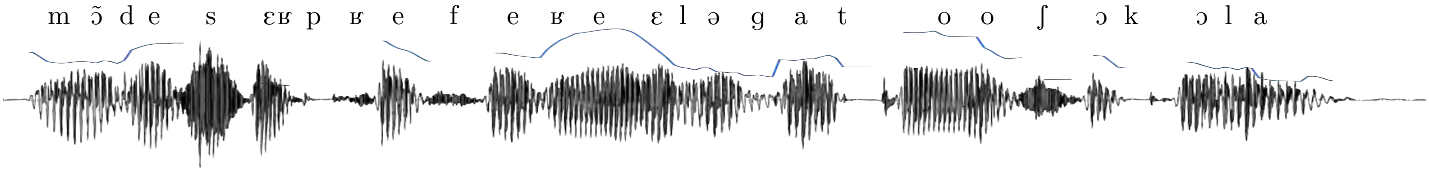
Zora, H., Riad, T., Schwarz, I.-C., & Heldner, M. 2016. Lexical Specification of Prosodic Information in Swedish: Evidence from Mismatch Negativity. *Front. Neurosci.* 10.533.

Cited in 6.1.3.2.2.

Part V

APPENDIX

A French summary



Parmis les sons que les individues peuvent rencontrer dans leurs environnement, la parole est celui qui comporte le plus d'information. A partir d'un signal vocal, l'auditeur peut extraire des informations extra-linguistiques ou para-linguistiques sur, par exemple, l'identité, la santé ou l'état émotionnel du locuteur, et il peut extraire les informations segmentales (par exemple, les phonèmes tels que voyelles et consonnes) et supra-segmentales (c'est-à-dire des informations plus grandes que les segmentales individuels) nécessaires pour reconnaître les mots. Une représentation acoustique de la parole (la phrase "Mon dessert préféré est le gâteau au chocolat") est représentée ci-dessus. Le signal acoustique indique que l'amplitude du son et la fréquence fondamentale (f_0) varient dans le temps et reflètent une partie de la richesse contenue dans le flux vocal.

L'information dans le flux de parole n'est pas présentée d'une manière robotique, staccato ou continue, mais organisée dans un 'arrangement musical' qui structure et regroupe le flux de parole, de sorte qu'il peut être perçu dans des énonés cohérents. Dans cet arrangement musical de la parole—mieux connu sous le nom de prosodie—intonation, accentuation et rythme, s'auto-organisent pour former un cadre prosodique hiérarchique, avec, en son cœur, un 'accent métrique'.

Les syllabes métriquement fortes peuvent être considérées comme les piliers qui guident les auditeurs dans leur analyse d'un énoncé. L'accent métrique sert de point d'ancrage au contour de l'intonation et de force régulatrice du rythme. Les modulations des caractéristiques acoustiques de l'accentuation métriques servent à indiquer la profondeur de la hiérarchie prosodique (c'est-à-dire, les niveaux incrémentiels de réalisation phonétique peuvent distinguer, par exemple, le niveau du groupe ou le marquage de l'intonation). De plus, l'accent métrique sous-tend la représentation abstraite du mot - la plus petite unité informationnelle - de sorte qu'il contribue facilement au traitement lexical. En effet, l'accent métrique fournit aux auditeurs des syllabes qui sont perceptiblement stables ; l'auditeur peut se référer à ces syllabes dans des environnements d'écoute difficiles ou bruyants dans lesquels le signal vocal peut être corrompu. De plus, les syllabes accentuées marquent souvent les frontières d'un mot, qui, comme nous le verrons plus loin dans ce travail, ne sont généralement pas évidentes à partir du signal acoustique. Enfin, l'accent métrique interagit avec l'attention, simultanément par sa proéminence acoustique au niveau de la surface et par sa prédictibilité temporelle sous-jacente. En d'autres termes, l'accent métrique capte l'attention du 'bottom-up' grâce à sa saillance acoustique qui attire l'attention, et il guide l'attention du 'top-down' en fournissant un cadre métrique que l'auditeur peut utiliser pour faire rebondir l'attention d'une syllabe accentuée à la suivante. Compte tenu de toutes ces caractéristiques, il est clair que l'accent métrique constitue la base du traitement de la parole.

Le français, la langue étudiée dans la présente thèse, peut être considéré comme une langue particulière par rapport aux autres langues, puisqu'elle décrite comme sans accent (Rossi, 1980). Les raisons de ce point de vue curieux, mais, jusqu'à présent, traditionnellement accepté, seront discutées—et contestées—tout au long de cette thèse. À ce stade, il est important de noter que la vision traditionnelle a eu deux conséquences majeures pour le traitement de la parole en français. Premièrement, la langue française était difficile à positionner dans les modèles théoriques avancés, où l'accent métrique joue un rôle central dans notre compréhension de la manière dont les auditeurs analysent la parole. Deuxièmement, comme le franÃ§ais est décrit comme une langue sans accent, les phénomènes accentuels français, et en particulier leur rôle dans le traitement au niveau des mots, ont suscité peu d'intérêt scientifique. Cela signifie que, même s'il y avait des accents lexicales ou métriques en français, et même si ils contribuaient au traitement lexical, cette fonction n'est pas facilement reconnaissable, soit parce qu'elle n'est pas du tout étudiée, soit parce que les fonctions sont attribuées à des domaines autres que le mot.

Nous nous interrogeons ici sur la conception du français comme "langue sans accent". Nous nous basons sur deux modèles métriques de l'accentuation française, qui proposent de coder l'accent dans des modèles cognitifs sous-jacents à la représentation abstraite du mot (Di Cristo, 1999; Astésano, 2001; Astésano et al., 2007; Astésano, 2016; Astésano & Bertrand, 2016; Astésano, 2017), et nous permettent de concevoir un rôle pour les accentuations françaises dans le traitement au niveau des mots. Dans notre étude interdisciplinaire du traitement des accents métriques en français, nous adoptons une approche fonctionnelle, mais enracinée dans la métrique. Nous utilisons la méthode des potentiels évoqués (ERP), qui nous fournit une mesure très sensible et précise au niveau temporel qui nous permet de déterminer s'il y a un stress métrique en français et dans quelle mesure ce stress métrique aide l'auditeur à comprendre la parole.

A.1 Chapitre 1 : Définition de l'accent métrique

Dans le chapitre 1, nous tenterons de définir les accents métriques. En effet, l'accent métrique, ou stress ou accentuation en général, n'est pas facile à définir. L'accentuation a été évoquée par une série de termes différents qui, selon l'auteur, peuvent désambiguïser le status phonologique de l'accent, son domaine prosodique et sa fonction linguistique. Dans la section 1.1 : *Accent lexical*, nous présenterons l'accent comme une entité abstraite et cognitive qui est spécifiée dans le lexique mental et qui fait partie de l'entrée lexicale. Nous dirons que l'accent au niveau des mots est indispensable dans le traitement lexical, parce qu'il interagit avec l'attention pendant la compréhension de la parole. Le stress attire l'attention du 'bottom-up' par sa proéminence acoustique et il capte l'attention du 'top-down' grâce à sa prévisibilité.

Nous verrons cependant dans la section 1.2 : *Proéminence de niveau supérieur* et section 1.3 : *Rythme et Metre* que de nombreux facteurs extérieurs au mot peuvent moduler la réalisation phonétique en surface de l'accent lexicale. Cela peut rendre difficile la reconnaissance d'un stress lexical approprié. Cependant, il sera démontré que l'accent est identifié de façon plus fiable lorsqu'on distingue sa forme phonologique, puisqu'il sous-tend la représentation cognitive des mots, des propriétés phonétiques avec lesquelles il se manifeste dans la parole. En effet, nous démontrerons que, bien que l'accent puisse être réalisé avec une diversité de paramètres phonétiques (par exemple f_0, durée, intensité), ceux-ci sont differents de sa representation abstraite et cognitive, et nous ferons valoir que les influences extérieures qui co-déterminent la manifestation phonétique en surface de l'accentuation ne modulent ni sa representation lexicale ni son rôle lexical dans la compréhension du discours. C'est alors à travers le rôle de l'accentuation dans le processus de compréhension de la parole, que nous pouvons observer son identité.

Dans la section 1.4 : *Fonctions de l'accent dans les modèles de traitement de la parole*, nous présenterons trois cadres théoriques qui aident à expliquer le rôle fonctionnel de l'accent dans la compréhension de la parole : la stratégie de segmentation métrique (Metrical Stress Segmentation, MSS; Cutler & Norris, 1988; Cutler, 1990), l'hypothèse de rebondissement attentionnel (Attentional Bounce Hypothesis, ABH; Pitt & Samuel, 1990) et la théorie dynamique des soins (Dynamic Attending Theory, DAT; Large & Jones, 1999; Jones et al., 2002; Jones, 2010).

Selon MSS, l'accent joue un rôle indispensable dans le traitement de la parole, parce qu'il marque les frontières des mots dans un flux de parole continu et informe l'auditeur du moment où il doit commencer à accéder au lexique (par exemple Cutler & Carter, 1987; Vroomen & de Gelder, 1995; Cutler & Norris, 1988; Cutler, 1990; Cutler & Butterfield, 1992; Vroomen et al., 1996). ABH se concentre plus spécifiquement sur la prévisibilité des accents métriques et propose des ressources attentionnelles fluctuantes tout en assurant des pics d'attention sur les accents qui peuvent ensuite être encodés de façon optimale (Pitt & Samuel, 1990). DAT, qui visait à l'origine à modéliser la perception de la musique, propose un mécanisme similaire pour optimiser le traitement en s'appuyant sur la structure rythmique du son, mais qui permet en outre de suivre plusieurs plans temporels (par exemple Large & Jones, 1999; Jones et al., 2002; Jones, 2010). Ainsi, l'accent interagit avec l'attention tout au long de la compréhension, c'est-à-dire qu'il capte l'attention par sa proéminence acoustique, et guide l'attention par sa régularité temporelle.

Enfin, dans la section 1.5 : *Le stress dans la prosodie française*, nous discutons du statut de l'accentuation dans la langue étudiée dans la présente thèse : Français. L'accentuation française a traditionnellement un statut phonologique faible et post-lexical. Le stress n'est pas marqué dans les dictionnaires français et se superpose à l'intonation, tant phonétique que fonctionnelle, ce qui fait que la langue est décrite comme "une langue sans accent" (Rossi, 1980). Par conséquent, l'accent français a suscité assez peu d'intérêt dans le domaine linguistique, et son rôle dans la compréhension de la parole est, à l'heure actuelle, mal compris. En effet, dans sa description traditionnelle de langue sans accent, le français est difficile à positionner dans les cadres théoriques présentés dans la section 1.4.

Cependant, nous montrerons qu'une des raisons qui explique la vision traditionnelle de l'accentuation française est que la confusion entre l'accent en tant qu'entité acoustique, caractérisé par les propriétés phonétiques avec lesquelles il fait surface dans la parole, et l'accent dans sa forme phonologique, qui sous-tend la représentation cognitive des mots. En effet, la plupart des descriptions de la prosodie française sont axées sur l'organisation tonale de la langue. Nous verrons que l'accent français a un poids métrique et est représenté dans les modèles de l'accent cognitif qui sous-tendent le mot lexical (*cf.* Di Cristo, 1999). En acceptant ce point de vue, l'accent français est plus facilement intégré dans les cadres théoriques présentés et peut être envisagé comme ayant un rôle fonctionnel dans la compréhension de la parole. Nous argumenterons que, pour observer au mieux les pattrons de l'accent métriques français, il est nécessaire d'adopter une approche fonctionnelle et d'étudier les interactions.

A.2 Chapitre 2 : Traitement de la parole

Afin d'apprécier les rôles fonctionnels de l'accentuation française dans le processus de compréhension de la parole, il sera nécessaire de mieux comprendre les défis qu'un système vocal doit relever lorsqu'il est confronté à un signal acoustique de parole. Dans le chapitre 2, nous examinerons de plus près ce que le traitement de la parole implique. C'est-à-dire à la section 2.1 : *Trois étapes dans la perception de la parole*, les trois processus impliqués dans la perception de la parole seront présentés et nous expliquerons pourquoi il est nécessaire d'extraire et d'analyser les patrons de l'accent.

En outre, il est souligné que le traitement de la parole est difficile à modéliser de manière computationnelle en raison de deux défis : le problème de variabilité de la parole (selon lequel, généralement, l'information segmentaire et supra-segmentaire ne sont présentée sous leur forme canonique) et le problème de segmentation, qui renvoie au manque d'indices marquant de manière cohérente les frontières entre les mots ou même entre phonèmes. Ces deux défis dans du traitement de la parole, sont précisés à la section 2.2 : *Problèmes computationels* grâce à la description de trois modèles de computationels bien connus sur la reconnaissance des mots et la segmentation de la parole (c'est-à-dire le modèle Cohort, Marslen-Wilson & Welsh 1978; Marslen-Wilson et al. 1996; TRACE, McClelland & Elman 1986; et Shortlist, Norris 1994; Norris & McQueen 2008). Nous argumenterons à travers l'évolution de ces modèles que les processus d'accès lexical et de segmentation de la parole impliquent forçement le décodage de l'information métrique pour identifier pré-lexicalement les débuts des mots (Briscoe, 1989).

Dans la section 2.3 : *Le rôle de la syllabe dans le traitement de la parole en français*, nous décrirons le rôle de la syllabe (c'est-à-dire la prétendue unité métrique française) dans le traitement de la parole en français. Nous verrons que même si les auditeurs français peuvent effectivement utiliser la syllabe dans le traitement de texte, la parole fournit souvent de multiples signaux à l'auditeur (présentés à la section 2.4 : *La hiérarchie des poids des indices de segmentation*), auxquels il peut se référer au besoin (par exemple Cutler et al., 1986; Wauquier-Gravelines, 1999; Dumay et al., 2002). De plus, si l'accentuation française a un poids métrique et s'applique au domaine proche du mot, les auditeurs français pourraient analyser la parole en se basant à la fois sur l'information syllabique et l'information sur l'information accententuelle.

En effet, dans la section 2.5 : *Segmentation temporelle parallèle*, nous décrirons les développements récents dans le domaine des neurosciences, qui indiquent aux auditeurs de traiter la parole sur plusieurs fenêtres temporelles qui indiquent que les auditeurs traitent le signal de parole en parallèle. Ce mécanisme repose sur des oscillations neuronales qui s'alignent sur des événements rythmiques et donc prévisibles dans le temps afin de moduler l'excitabilité des réseaux neuronaux de telle sorte que les points importants dans le temps soient traités de manière optimale (par exemple Giraud & Poeppel, 2012; Ghitza, 2011, 2013; Peelle & Davis, 2012; Gross et al., 2013). Cette stratégie de traitement de la parole repose vraisemblablement sur l'interaction entre la prosodie et l'attention, et leur effet combiné sur l'excitabilité neurale, et fournit une nouvelle perspective sur le rôle de l'accentuation/- stress métrique dans la compréhension de la parole en français.

A.3 Chapitre 3 : Alignement neural sur le rythme de la parole

Le chapitre 3 décrit la relation entre l'accentuation et l'attention, particulièrement en ce qui concerne l'excitabilité neurale. Il est expliqué de quelle manière ABH et DAT envisagent l'accentuation pour guider l'excitabilité neurale de manière à faciliter la compréhension de la parole. En d'autres termes, en se basant sur les deux théories de l'attention, comment peut-on expliquer que la manipulation de l'accent module l'excitabilité neurale réelle et donc le traitement de la parole ? Il sera expliqué dans ce chapitre en quoi DAT et ABH fournissent des théories biologiquement plausibles de la façon dont l'information prosodique facilite le traitement de la parole.

En effet, dans la section 3.1 : *Excitabilité neurale et échantillonnage attentionnel*, nous verrons que les états d'excitabilité des ensembles neuronaux fluctuent rythmiquement dans ce que l'on appelle une oscillation neurale ou cérébrale. Cela signifie que la phase d'une oscillation influence la probabilité qu'un neurone se déclenche (Buzsáki, 2009). Des développements récents ont découvert que lorsque l'information auditive présente certaines régularités temporelles, elle est traitée par un mécanisme oscillatoire dynamique, qui suit l'entrée sur de multiples fréquences, basé sur des prédictions.

En particulier dans l'analyse de la parole, un tel mécanisme peut servir à plusieurs fins. La parole est intrinsèquement rythmique. Les rythmes de la parole sont de nature hiérachique, les différentes couches de la hiérarchie prosodique transmettant l'information sur de multiples échelles de temps (par exemple Rosen, 1992). Un mécanisme dynamique, tel que proposé dans DAT, assure une sensibilité neuronale élevée aux parties les plus importantes du signal de parole. De plus, il peut expliquer comment la parole est simultanément traitée et segmentée sur plusieurs échelles temporelles (par exemple Poeppel et al., 2008; Giraud & Poeppel, 2012).

La plupart des travaux sur l'apprentissage neuronal (c'est-à-dire l'un des mécanismes à l'origine de la manipulation de l'excitabilité neurale) de la parole se sont toutefois concentrés sur les limites temporelles du traitement des niveaux inférieurs de la hiérarchie prosodique, les études utilisant la méthode des potentiels liés aux événements (la méthode utilisée dans le courant) traitant plus directement de la manière dont la compréhension des structures métriques facilite la compréhension des processus vocaux.

Section 3.2 : *Potentiels evoqués et codage prédictif* présente l'utilisation de la méthode des potentiels evoqués (ERP) dans les recherches sur le traitement de la parole. Les ERP s sont des signaux cérébraux à verrouillage temporel moyen, provoqués par an événement externe. Les ERP reflètent généralement l'inadéquation entre les événements sensoriels et les événements attendues et, à ce titre, ils sont étroitement liés au cadre théorique du "predictive coding". Les décalages entre l'information sensorielle et la prédiction descendante entraînent ce qu'on appelle des erreurs de prédiction qui nécessitent un traitement supplémentaire. C'est alors cet effort supplémentaire qui est reflété ou mesuré dans la modulation des composantes ERP. Cette interprétation des ERP s en ce qui concerne le traitement du langage souligne que la perception de la parole repose constamment sur la prédiction en ligne, générant des hypothèses préalables basées soit sur le contexte, soit sur la mémoire à long terme.

Nous discuterons de quatre composantes pertinentes dans le cadre des travaux présentés : MMN, PMN, N325 et N400 (Näätänen et al., 1997; Bentin et al., 1985; Connolly & Phillips, 1994; Böcker et al., 1999; Steinhauer & Connolly, 2008; Brown & Hagoort, 1993; Kutas & Federmeier, 2011). Ces quatre composantes ont toutes en commun le fait que leur modulation reflète une inadéquation entre une attente fondée sur des représentations de la mémoire à long terme ou des représentations phonologiques/linguistiques établies et la violation dans le contexte expérimental. À ce titre, elles reflètent chacune le résultat du codage prévisible (Friston, 2005) et permettent de tirer des conclusions sur le décours temporel (c'est-à-dire l'étape du traitement) et les représentations phonologiques prévues (auxquelles les entrées ne correspondent pas) pendant la compréhension vocale (voir également Scharinger et al., 2016).

A.4 Chapitre 4 : Questions de recherche et hypothèses

Ce travail a pour objectif de tester si s'il y a un accent métrique en français. Plus précisément, nous adoptons une approche fonctionnelle pour examiner si les accents initiaux et finaux français (c.-à-d. IA et FA, respectivement) sont codés phonologiquement dans des modèles de l'accent bipolaires sous-jacents à la représentation du mot lexical, et si la présentation des mots sans leurs modèles de l'accent entrave le traitement au niveau des mots.

Dans le chapitre 4, nous exposons les questions qui ont motivé notre travail et présentons nos hypothèses initiales.

A.5 Chapitre 5 : Stratégie expérimentale

Le chapitre 5 présente les méthodes utilisées dans la thèse. Afin d'étudier les questions qui viennent d'être présentées, nous avons manipulé la présence de l'accent initial et/ou de l'accent final sur les mots trisyllabiques dans les paradigmes ERP. Plus spécifiquement, la section 5.1 : *Création de stimuli* décrit comment les stimuli ont été crés, c'est-à-dire comment nous avons manipulé la présence de l'accent initial et final en termes de modulation de f_0 et de durée.

Et la section 5.2 : *Procédures statistiques* nous allons motiver et détailler notre choix d'analyse de données (c'est-à-dire modèles mixtes pour les données comportementales et test de permutation $t_{\max}$ non paramétrique pour les données ERP).

A.6 Chapitre 6 : Études

Dans ce chapitre, nous présenterons les cinq études réalisées dans le cadre de cette thèse. Les études sont présentées avec une introduction dans laquelle nous expliquons précisément quelle question a motivé le recueil de données et comment nous avons concu nos prédictions.

Le cas échéant, certaines études sont regroupées afin de faciliter la lecture. Par exemple, la section 6.1 présente deux études sur la représentation phonologique de l'accentuation française, soit FA dans la première et FA et IA mixte dans la seconde. De même, dans la section 6.2, nous présentons deux études de décisions lexicales qui ont été construites pour observer l'interaction des deux accents français (l'accent initial dans le premier et l'accent final dans le second) avec le processus de l'accès lexical.

Enfin, dans la section 6.3, nous présentons une étude dans laquelle l'attente métrique et la congruence sémantique sont manipulées orthogonalement afin de déterminer si les difficultés du traitement métrique ont également une incidence sur les étapes ultérieures, par exemple la récupération sémantique ou l'intégration sémantique, dans la compréhension de la parole.

Afin de souligner davantage la cohérence naturelle entre les études, chaque étude motivant la suivante, les résultats de chaque étude (groupée) seront interprétés en détail dans une discussion au cours de laquelle nous nous ferons part de nos conclusions.

Dans l'ensemble, nous montrerons comment les résultats ont renforcé et motivé notre étude sur le traitement des contraintes métriques en français.

A.6.1 Représentation phonologique des accents métriques en français

A.6.1.1 MisMatch Negativity: accent final

Question de recherche : Les auditeurs français sont-ils sensibles à l'accent final primaire français (FA) et l'accent fait-il partie du modèle de l'accent phonologique attendu en français ?

Procédure : 19 participants ont regardé passivement un film muet dans un paradigme de “oddball”. Dans une condition, les déviants étaient présentés sans (−FA) et les standards avec accent final (+FA), tandis que dans une autre condition, ces positions étaient changées (consultez la figure A.1 pour un exemple des stimuli dans cette étude et table A.1 pour les caractéristiques des stimuli).

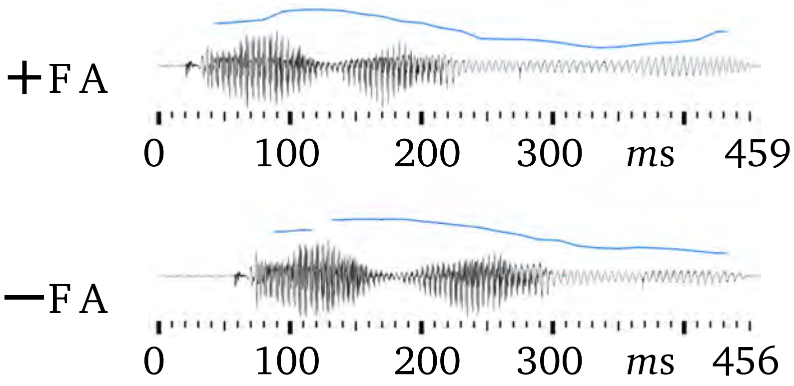


Figure A.1: Example of durational manipulation for +FA (top) and −FA (bottom) of the stimulus paʁadi (*paradis*). The two waveforms and associated pitch tracks show how syllable duration was shortened substantially for the final syllable, and moderately for the initial two syllables.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd syllable
	ms	f_0	*ms*	f_0	*ms*	f_0	*ms*	f_0	onset
CASINO									
+FA	503.3	117.8	112.9	125.8	157.1	122.2	233.3	111.0	269.0
−FA	500.8	119.0	146.9	125.1	178.1	122.3	178.8	110.7	325.0
PARADIS									
+FA	459.0	106.5	121.5	121.7	111.6	118.9	225.9	93.4	233.1
−FA	456.5	110.9	168.8	123.3	139.3	119.1	148.3	95.1	308.1

Table A.1: Overview of durational and f_0 values, plus the timing of the third syllable (holding ±FA) onset for both 'casino' and 'paradis' with and without final accent.

Résultats : Nous avons obtenu des amplitudes d'onde MMN asymétriques, telles que le déviant -FA induit un MMN plus ample que le déviant +FA (qui n'induit pas un MMN, figure A.3). De plus, la différence d'amplitude d'onde entre des stimuli identiques dans des positions différentes à l'intérieur des paradigmes de l'oddball, a indiqué que −FA étaient défavorisés, qu'il s'agisse des déviants ou des standards (figure A.2).

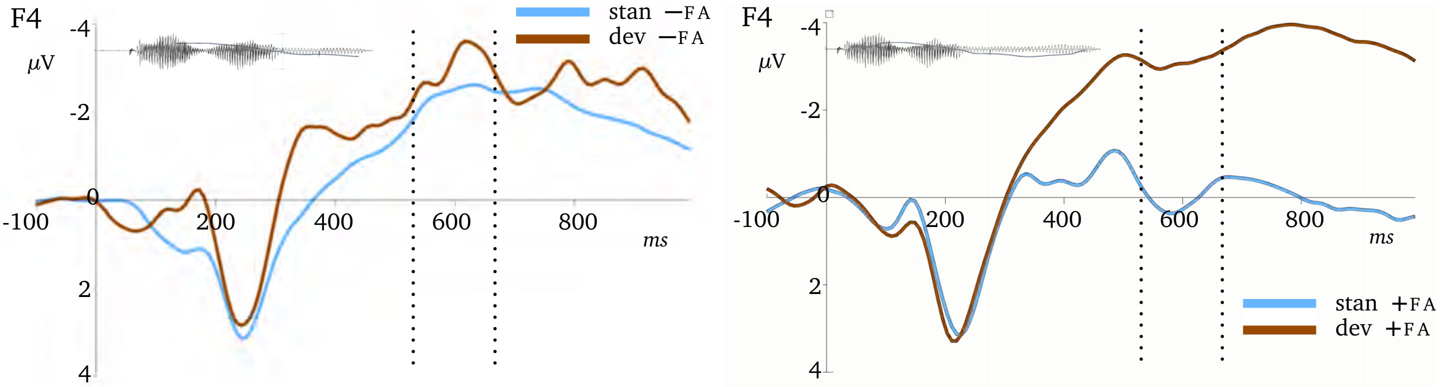


Figure A.3: ERP components for −FA (left) and +FA (right) stimuli, recorded at the F4 (right frontal) electrode, with the oscillograms of [paʁadi] plotted in the background to indicate the relation between the offset of the ±FA manipulation and the resulting stimulus-locked ERPs. The tested time-window is indicated by dashed vertical lines. For ease of presentation, ERP waveforms are low-pass filtered at 10 Hz and negativity is plotted as an upward deflection.

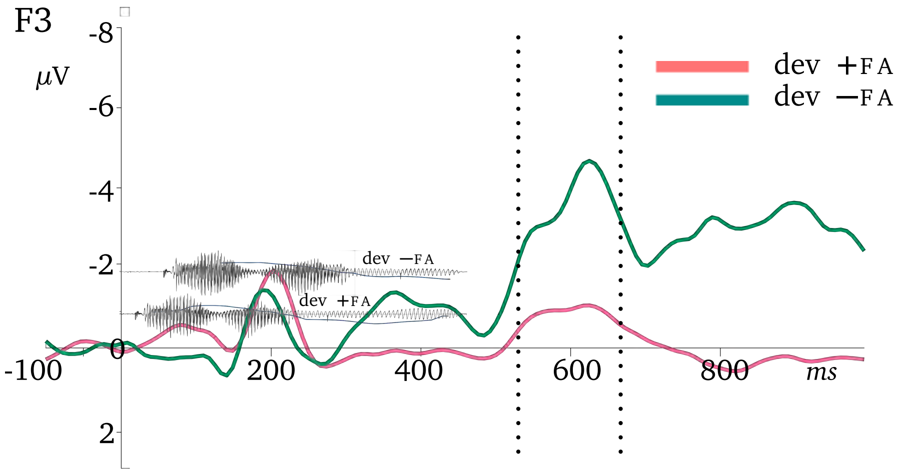


Figure A.2: MMN components for +FA (in pink) and −FA (in green) deviants, recorded at the F3 (left frontal) electrode, with the oscillogram of the deviant stimuli [paʁadi] plotted in the background. Waveforms and oscillograms are temporally aligned to indicate the relation between the offset of the ±FA manipulation and the resulting stimulus-locked event-related potentials. The tested time-window is indicated by dashed vertical lines. For ease of presentation, ERP waveforms are low-pass filtered at 10 Hz and negativity is plotted as an upward deflection.

Conclusion : Les auditeurs français ne sont pas sourds à l'accent final. Les résultats indiquent plutôt que l'accent final est encodé phonologiquement et à attacher au patron de l'accent attendu qui sous-tend la représentation du mot lexical.

MisMatch Negativity: Accent final et initial

Question de recherche : L'accent final français et l'accent initial secondaire sont-ils codés de la même façon et les auditeurs ont-ils des attentes similaires entre les deux accents ?

Procédure : 20 participants ont regardé passivement un film muet dans un paradigme de oddball. Les standards étaient toujours présentés avec l'accent initial et l'accent final, tandis que les déviants étaient présentés sans accent final (—FA) ou sans accent initial —IA (consultez la figure A.4 pour un exemple des stimuli dans cette étude et table A.2 pour les caractéristiques des stimuli).

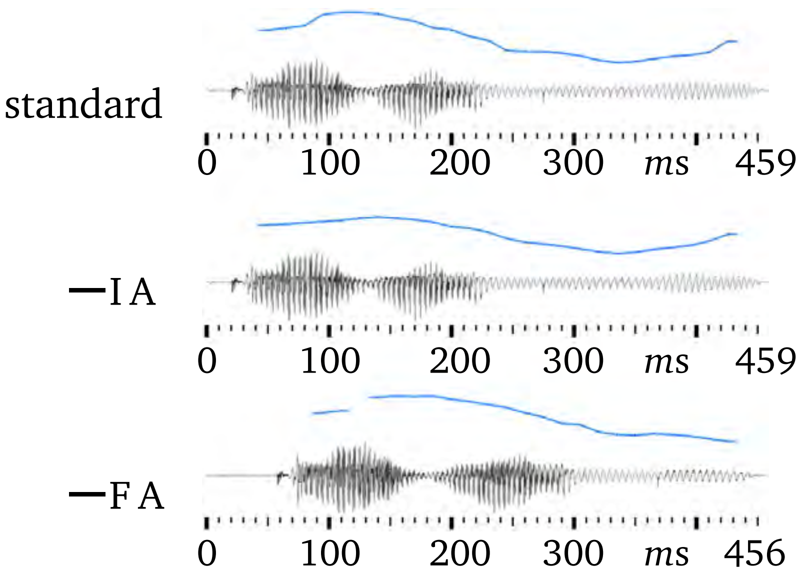


Figure A.4: Example of the stimulus *paradis* [paʁadi]. At the top, the standard with both IA and FA, the deviant —IA in the middle, and the deviant —FA at the bottom.

Table A.2: Overview of stimulus properties (durational and f_0 values, plus the timing of the third syllable onset) for both 'casino' and 'paradis' standards, deviants —IA and deviants —FA.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd syllable
	ms	f_0	*ms*	f_0	*ms*	f_0	*ms*	f_0	onset
CASINO									
standard	503.3	117.8	112.9	125.8	157.1	122.2	233.3	111.0	269.0
dev—IA	503.3	116.0	112.9	120.5	157.1	121.6	233.3	109.9	269.0
dev—FA	500.8	119.0	146.9	125.1	178.1	122.3	178.8	110.7	325.0
PARADIS									
standard	459.0	106.5	121.5	121.7	111.6	118.9	225.9	93.4	233.1
dev—IA	459.0	104.1	121.5	114.4	111.6	114.4	225.9	93.4	233.1
dev—FA	456.5	110.9	168.8	123.3	139.3	119.1	148.3	95.1	308.1

Résultats : Nous avons obtenu des MMNs à la fois aux déviants —FA et aux déviants IA, bien que les déviants —FA aient suscité un MMN plus ample (consultez figure A.5).

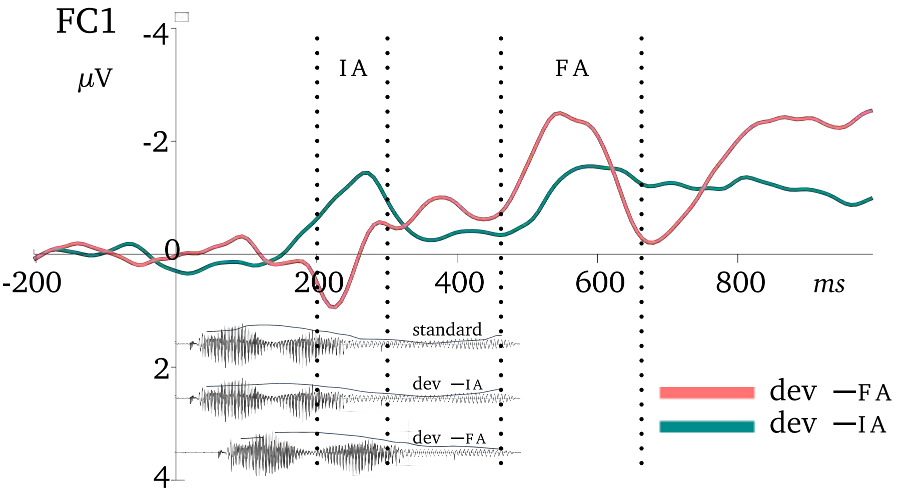


Figure A.5: MMN components for —IA (in green) and —FA (in pink) deviants, recorded at the FC1 (left frontal) electrode, with the oscillogram of all stimuli type of [paʁadi] plotted in the background. Waveforms and oscillograms are temporally aligned to indicate the relation between the offset of the ±IA and ±FA manipulation and the resulting stimulus-locked event-related potentials. Tested time-windows are indicated by dashed vertical lines. For ease of presentation, ERP waveforms are low-pass filtered at 10 Hz and negativity is plotted as an upward deflection.

Conclusion : L'accent initial et l'accent final ont été facilement perçus par les auditeurs français, ce qui va à l'encontre de la notion de surdité de l'accent en français. Les résultats suggèrent toutefois que les rôles respectifs de l'accent final et de l'accent initial sont différents.

A.6.2 L’accent métrique dans la reconnaissance des mots

Dans la section 6.2, nous présentons deux études de décision lexicale qui ont été construites pour déterminer l’interaction des deux accents français (l’accent initial et l’accent final, présentés dans la section 1.5) avec le processus d’accès lexical.

Décision lexicale : accent initial

Question de recherche : L’accent initial est-il attendu comme sous-jacent à la représentation des mots et est-il traité principalement de manière pré-lexicale ou intervient-il également dans les étapes ultérieures de l’accès lexical ?

Procédure : 23 participants ont repondu à une tâche de décision lexicale, dans laquelle la lexicalité (noms français réels par rapport à des pseudos-mots) et le patron de l’accent métrique (+IA vs −IA) ont été manipulés orthogonalement sur des stimuli trisyllabiques (consultez la figure A.6 pour un exemple des stimuli dans cette étude et table A.3 pour les caractéristiques des stimuli).

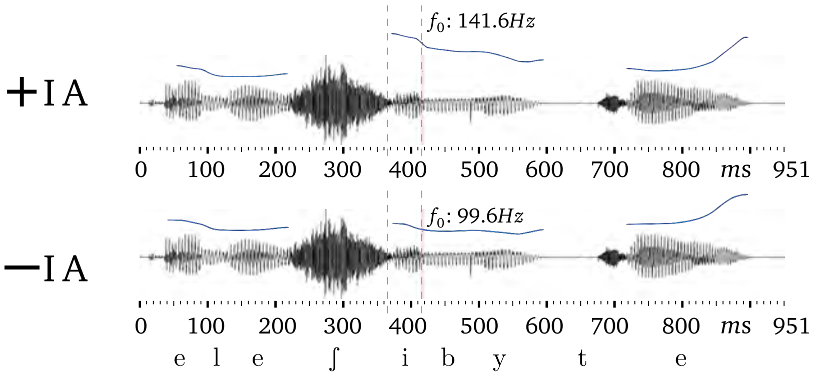


Figure A.6: Example of f_0 resynthesis +IA (top) and −IA (bottom) of the stimulus ʃibyte (chibuté), with quadratic interpolation from the f_0 value of the preceding determinant to the f_0 value at the beginning of the last stressed syllable for −IA targets.

Table A.3: Overview of mean stimulus properties for both lexical words and pseudowords ±IA (total duration, total first syllable and syllable-vowel durations, and first syllable-vowel and determinant-vowel f_0 values).

	Total duration		1st syllable ms		1st vowel ms		1st vowel f_0		Det vowel f_0	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
WORD										
+IA	647.7	62.5	176.6	36.1	79.0	22.4	263.1	19.2	199.4	11.7
−IA	647.7	62.5	176.2	35.8	79.0	22.4	217.6	13.4	196.7	11.8
PSEUDOWORD										
+IA	658.6	44.8	169.6	25.3	75.2	18.2	272.7	28.9	197.5	29.5
−IA	658.6	44.8	169.6	25.3	75.2	18.2	217.4	10.5	196.3	8.5

Résultats : Résultats comportementaux, effect de lexicalité (consultez figure A.7 et table A.4 et table A.5) :

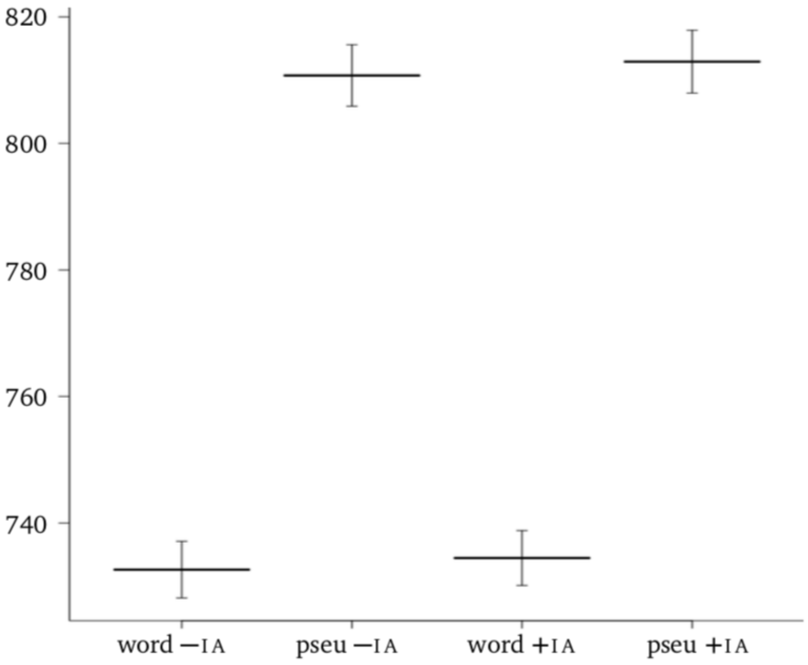


Figure A.7: Error-bar plot of mean reaction times for all four conditions (word −IA, pseudoword −IA, word +IA, pseudoword +IA) showing a clear significant effect of lexicality and no effect of presence of initial accent, nor an interaction between both manipulations.

Response latencies	*m*	*sd*	maximum	minimum	% errors
Word					
−IA	732.63	165.43	1774	325	1.2%
+IA	734.47	160.26	1679	306	0.5%
Pseudoword					
−IA	810.73	179.13	1793	358	0.9%
+IA	812.92	182.79	1754	436	1.5%

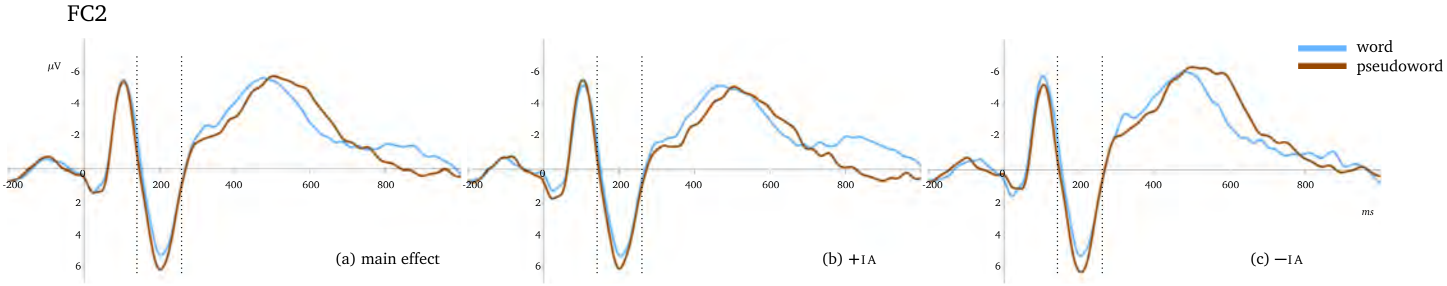
Table A.4: Reaction times per condition. Data analysis revealed a significant effect of lexicality, but no effect of ±IA and no interaction.

Nous avons obtenu un effet de ±IA sur l'amplitude des composantes de N325, de sorte que les stimuli −IA ont suscité un N325 plus grand que les stimuli +IA (figure A.9). Cet effet était plus faible dans la condition pseudo-mot, que dans la condition mot lexical bien que l'inspection visuelle suggère des processus similaires. De plus, nous avons obtenu un effet de lexicalité sur la composante P2, qui est censée refléter un chevauchement temporel entre la N325 et la P2, ce qui

Table A.5: Overview linear mixed models. The model fitting the data best takes lexicality as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent did not significantly contribute to the prediction of reaction times.

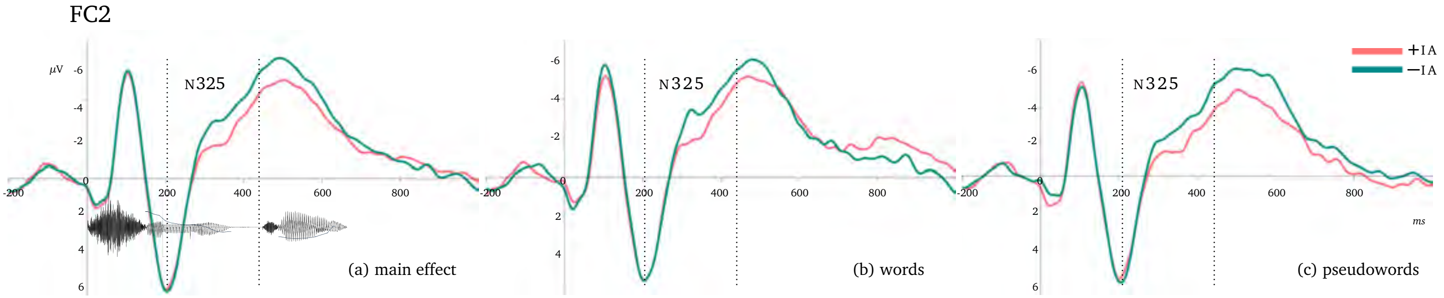
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	772.37^{***}	772.72^{***}	655.34^{***}	775.46^{***}	658.15^{***}	657.02^{***}
	(17.61)	(17.99)	(20.67)	(22.17)	(23.19)	(39.18)
lexicality			78.25^{***}		78.25^{***}	79.00^{***}
			(6.99)		(7.00)	(22.19)
±IA				−1.83	−1.87	−1.12
				(8.64)	(7.00)	(22.18)
±IA:lexicality						−0.50
						(14.03)
AIC	70588.80	70131.37	70026.59	70127.17	70022.79	70017.67
BIC	70608.61	70177.58	70079.41	70179.99	70082.22	70083.70
Log Likelihood	−35291.40	−35058.68	−35005.30	−35055.59	−35002.40	−34998.84
Num. obs.	5447	5447	5447	5447	5447	5447
Num. groups: subject	23	23	23	23	23	23
Var: subject (Intercept)	7027.58	7014.83	7017.44	7014.79	7017.39	7017.33
Var: Residual	24438.60	20923.99	20926.03	20923.97	20926.01	20925.98
Num. groups: stimuli		240	240	240	240	240
Num. groups: lex:stimuli		240	240	240	240	240
Num. groups: ±IA:stimuli		240	240	240	240	240
Var: stimuli (Intercept)		292.62	40.37	1016.38	173.44	210.20
Var: lex.stimuli (Intercept)		0.08	1126.62	146.17	1365.94	345.59
Var: ±IA.stimuli (Intercept)		0.05	787.50	919.59	345.11	1405.91

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$



(a) main effect (b) +IA (c) −IA

Figure A.8: Grand average P2 in the lexical condition (pseudoword-condition in brown, word-condition in blue), recorded at the FC2 (fronto-central) electrode for: (a) main effect, (b) +IA, (c) −IA. The tested time-window is indicated by dashed vertical lines. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are low-pass filtered at 10 Hz.



(a) main effect (b) words (c) pseudowords

Figure A.9: Grand average N325 in the ±IA condition (−IA-condition in green, +IA-condition in pink), recorded at the FC2 (fronto-central, bottom) electrode for: (a) main effect, (b) words, (c) pseudowords. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillogram and f_0 deflection of [ʃibyte] (*chibuté*) +IA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are lowpass filtered at 10 Hz.

suggère que notre manipulation métrique a affecté le traitement au début de la reconnaissance des mots figure A.8.

Conclusion : Les données révèlent à la fois l'automaticité de l'extraction pré-lexicale des contraintes et une préférence pour les modèles de contraintes avec accent initial pendant les premières étapes de l'accès lexical.

Décision lexicale : accent final

Dans la section 6.3 : L'accent initial français dans le traitement lexico-sémantique, nous présentons une étude dans laquelle l'espérance métrique et la congruence sémantique sont manipulées orthogonalement afin de déterminer si les difficultés du traitement métrique affectent les étapes ultérieures de la compréhension du langage telles que la récupération sémantique ou l'intégration sémantique.

Question de recherche : L'accent final français est-il attendu comme sous-jacent à la représentation des mots, semblable à l'accent initial, ou est-il traité principalement au cours des étapes ultérieures de l'accès lexical ?

Procédure : 20 participants ont rempli une tâche de décision lexicale, dans laquelle la lexicalité (noms français réels par rapport à des pseudos-mots) et le modèle de l'accent métrique (+FA vs −FA) ont été manipulés orthogonalement sur des stimuli trisyllabiques (consultez la figure A.10 pour un exemple des stimuli dans cette étude et table A.6 pour les caractéristiques des stimuli).

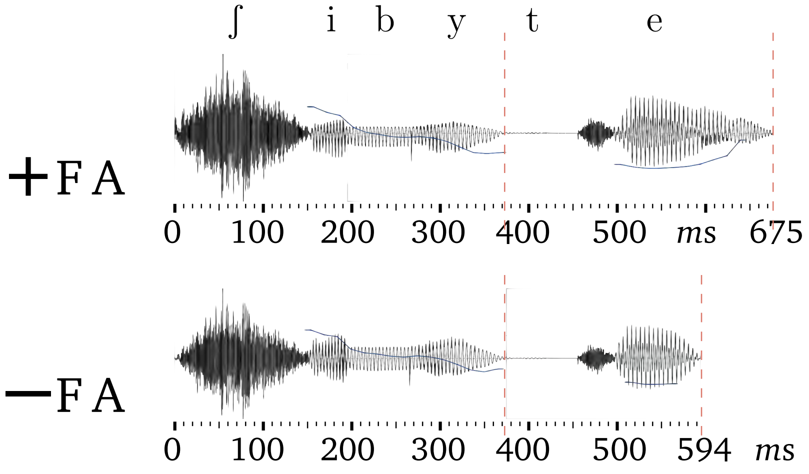


Figure A.10: Example of duration manipulation +FA (top) and −FA (bottom) of the stimulus ʃibyte (chibuté). The duration of the final syllable is equal to the duration of the unaccented second syllable for the −FA targets.

Table A.6: Overview of mean stimulus properties for both lexical words and pseudowords ±FA (total duration, syllable durations and timing of third syllable onset).

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd onset	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
WORD										
+FA	647.9	62.5	176.6	36.1	168.3	30.1	303.6	54.6	345.2	40.5
−FA	565.9	69.0	176.2	35.8	167.7	30.5	221.7	58.7	346.5	56.4
PSEUDOWORD										
+FA	658.9	44.8	169.6	25.3	175.3	31.3	314.0	47.2	345.0	31.9
−FA	564.0	49.0	169.6	25.3	175.3	31.3	219.0	47.2	345.0	31.9

Résultats : Résultats comportementaux, effect de lexicalité (consultez figure A.11 et table A.8 et table A.7) :

Table A.7: Overview linear mixed models. The model fitting the data best takes lexicality as fixed factor and subjects and stimuli variability as random factors. Presence of final accent did not significantly contribute to the prediction of reaction times.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	1061.94***	1059.83***	1112.79***	1054.66***	1107.89***	1101.76***
	(24.46)	(25.47)	(26.07)	(25.85)	(26.44)	(26.77)
lexicality			−108.71***		−108.71***	−96.05***
			(13.82)		(13.84)	(16.34)
±FA				10.40	9.83	22.14
				(8.80)	(8.75)	(12.19)
±FA:lexicality						−25.42
						(17.51)
AIC	38591.28	38415.85	38358.03	38410.27	38352.59	38344.92
BIC	38609.08	38457.39	38405.50	38457.74	38406.00	38404.26
Log Likelihood	−19292.64	−19200.93	−19171.01	−19197.13	−19167.29	−19162.46
Num. obs.	2792	2792	2792	2792	2792	2792
Num. groups: subj	18	18	18	18	18	18
Var: subj (Intercept)	10389.67	10491.87	10566.46	10499.48	10573.56	10575.08
Var: Residual	57741.78	50200.34	50185.95	50185.00	50171.06	50157.33
Num. groups: stimuli		160	160	160	160	160
Num. groups: lex:stimuli		160	160	160	160	160
Var: stimuli (Intercept)		1.17	2103.46	7371.37	1394.45	1227.17
Var: lex.stimuli (Intercept)		7552.73	2208.74	209.28	0.00	3466.80

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

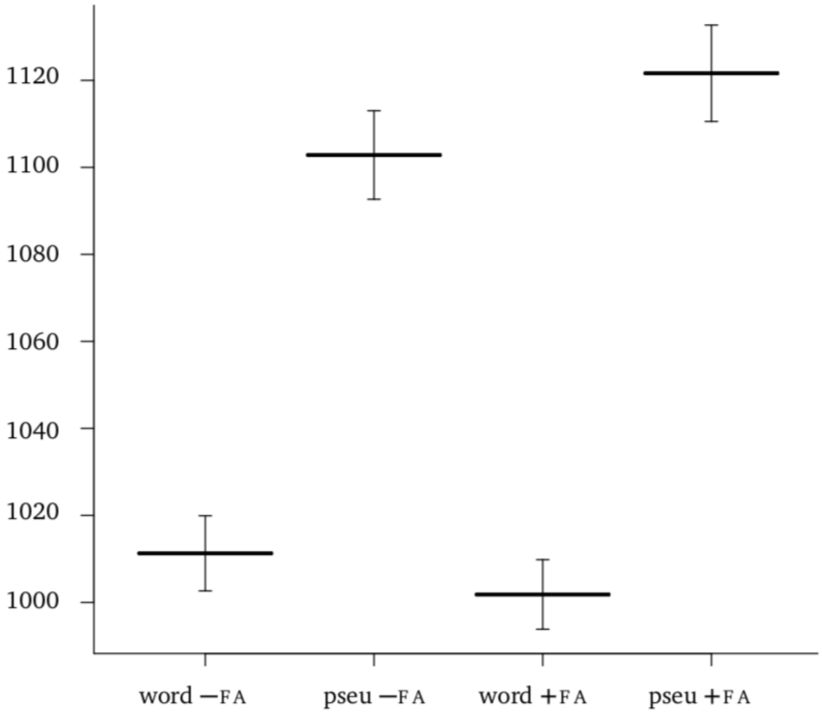


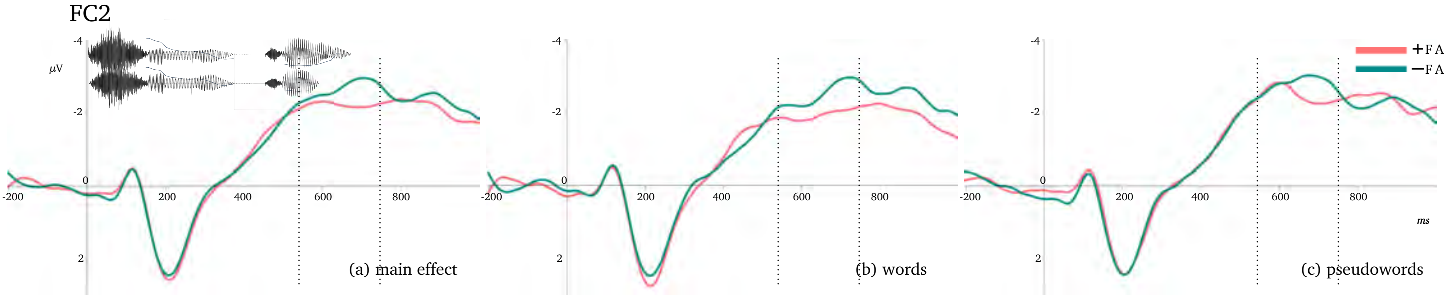
Figure A.11: Error-bar plot of mean reaction times for all four conditions (word −FA, pseudoword −FA, word +FA, pseudoword +FA) showing a clear significant effect of lexicality and no effect of presence of final accent, nor an interaction between both manipulations.

Response latencies	*m*	*sd*	maximum	minimum	% errors
WORD					
−FA	1011.29	224.25	2290.72	612.30	3.4%
+FA	1001.81	206.79	2444.67	651.06	2%
PSEUDOWORD					
−FA	1102.84	274.50	2461.48	623.52	1%
+FA	1121.36	297.45	2474.52	658.13	1%

Table A.8: Reaction times per condition. Data analysis revealed a significant effect both of lexicality, but no effect of ±FA and no interaction.

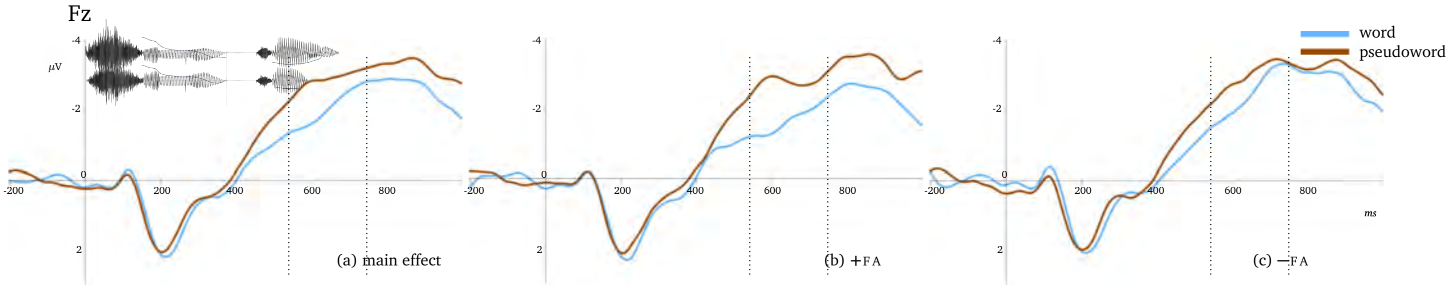
Nous avons obtenu un effet principal de ±FA sur l'amplitude des composantes N325, de sorte que les stimuli −FA ont suscité un N325 plus grand que les stimuli +FA (figure A.12). La lexicalité modulait

également l'amplitude de N325, mais seulement lorsque des stimuli avaient été présentés avec FA avec des pseudos-mots déclenchant un N325 plus grand que les mots (figure A.13).



(a) main effect (b) words (c) pseudowords

Figure A.12: Grand average N325 in the ±FA condition (−FA in green, +FA in pink), recorded at the FC2 (fronto-central) electrode for: (a) main effect, (b) words (c) pseudowords. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [ʃibyte] (*chibuté*) +FA and −FA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are low-pass filtered at 10 Hz.



(a) main effect (b) +FA (c) −FA

Figure A.13: Grand average N325 in the lexicality condition (pseudowords in brown, words in blue), recorded at the Fz electrode for: (a) main effect, (b) +FA, (c) −FA. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [ʃibyte] (*chibuté*) +FA and −FA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are lowpass filtered at 10 Hz.

Conclusion : Les résultats de l'ERP révèlent une interaction entre la présence de l'accent final français et la lexicalité. Bien que les pseudo-mot aient entravé le traitement, qu'ils aient ou non été présentés avec un accent final, seuls les mots −FA ont obstrué l'accès lexical. Les auditeurs avaient donc une attente lexicale pour l'accent final, c'est-à-dire que FA est naturel pour l'auditeur et facilite l'accès lexical.

A.6.3 Le stress français dans le traitement lexico-sémantique

Question de recherche : L'accent pré-lexical français initial interagit-il avec les étapes ultérieures du traitement de la compréhension de la parole, comme l'accès au sens et l'intégration contextuelle ?

Procédure : 18 auditeurs ont participé à l'étude N400. Ils ont jugé la congruence sémantique des mots cibles qui étaient sémantiquement congruents ou non congruents, et présentés avec ou sans accent initial (consultez la figure A.14 pour un exemple des stimuli dans cette étude et table A.9 pour les caractéristiques des stimuli).

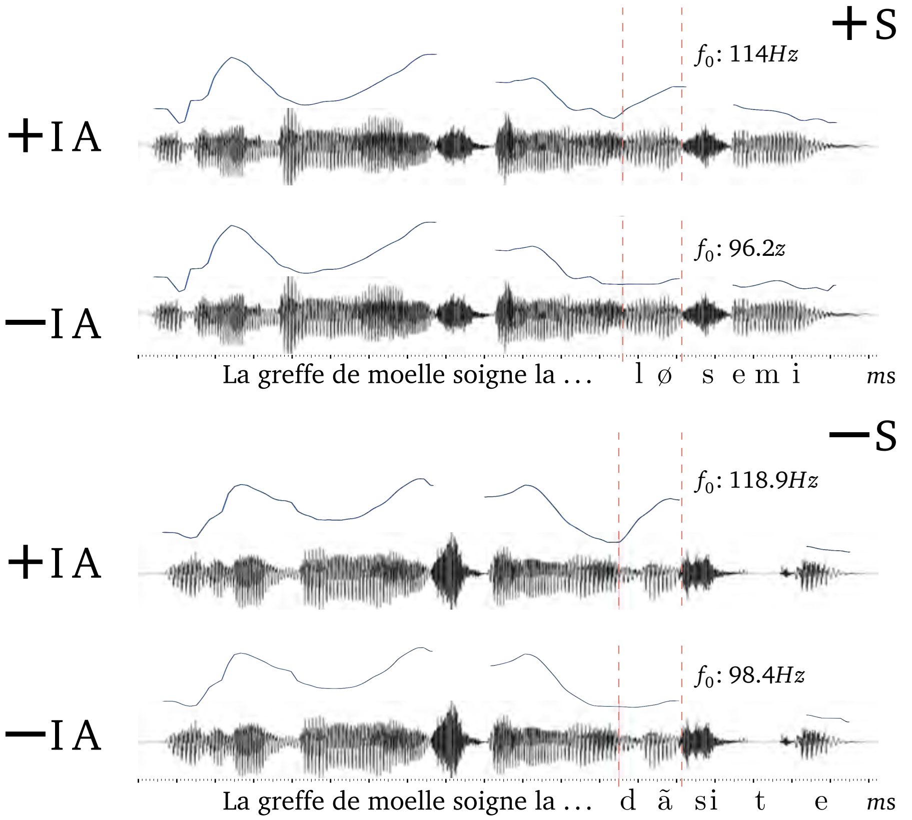


Figure A.14: Example of f_0 resynthesis with (+IA) and without initial accent (−IA) on semantically incongruent (+S, top two) and semantically congruent (−S, bottom two) sentences with quadratic interpolation from the f_0 value of the preceding determinant to the f_0 value at the beginning of the last stressed syllable for +IA targets (visible in blue). The time window of ±IA is indicated by vertical red dashed lines.

	Sentence *ms*		Target word *ms*		1st syllable *ms*		1st vowel *ms*		1st vowel f_0	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
SEMANTICALLY CONGRUENT										
−IA	2097.07	402.81	552.88	96.98	157.23	28.76	72.16	25.9	116.56	11.73
+IA	2092.07	402.81	552.88	96.98	157.23	28.76	72.16	25.9	126.38	12.2
SEMANTICALLY INCONGRUENT										
−IA	2122.59	411.72	583.45	61.72	160.57	32.16	77.86	27.23	123.02	42.18
+IA	2122.59	411.72	583.45	61.72	160.57	32.16	77.86	27.23	140.28	44.26

Table A.9: Overview of mean stimulus properties within the semantically congruent and incongruent conditions ±IA (total sentence and target-word duration, first syllable and syllable-vowel durations, and first syllable-vowel f_0 values).

Résultats : Résultats comportementaux, effect de ±IA et de la congruence semantique (consultez figure A.15 et table A.12 et table A.13) :

Table A.11: ANOVA: initial accent and semantic congruency as significant predictors for reaction times.

ANOVA	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
N400—intercept	2	540	550	-268	536			
N400—random structure	6	412	443	-200	400	135.502	4	0.00000
N400— ±IA	7	401	437	-193	387	13.329	1	0.00026
N400— ±S	7	405	441	-195	391	7.3802	0	0.06510
N400—interaction	9	402	449	-192	384	0.057	1	0.81140

Table A.10: Overview linear mixed models. The model fitting the data best takes semantic congruency as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent significantly contributes to the prediction of reaction times when it is the only fixed effect and marginally contributes when entered together with semantic congruency.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-3.00^{***}	-8.40^{***}	-9.50^{***}	-10.07^{***}	-10.52^{***}	-10.72^{***}
	(0.17)	(0.98)	(1.23)	(1.24)	(1.43)	(1.70)
congruency			1.73		1.69	1.93
			(0.94)		(0.98)	(1.41)
±IA				1.58^{*}	1.49^{*}	1.73
				(0.63)	(0.60)	(1.19)
±IA:congruency						-0.30
						(1.26)
AIC	539.70	412.20	405.14	400.87	399.81	401.76
BIC	550.10	443.39	441.53	437.26	441.41	448.55
Log Likelihood	-267.85	-200.10	-195.57	-193.44	-191.91	-191.88
Num. obs.	1338	1338	1338	1338	1338	1338
Num. groups: subj	20	20	20	20	20	20
Var: subj (Intercept)	0.16	1.10	1.31	1.40	1.31	1.32
Num. groups: congr:stimuli		160	160	160	160	160
Num. groups: ±IA:stimuli		160	160	160	160	160
Var: congr.stimuli (Intercept)		34.81	29.15	42.47	33.43	33.61
Var: ia.stimuli (Intercept)		0.00	3.72	2.20	2.11	2.11

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

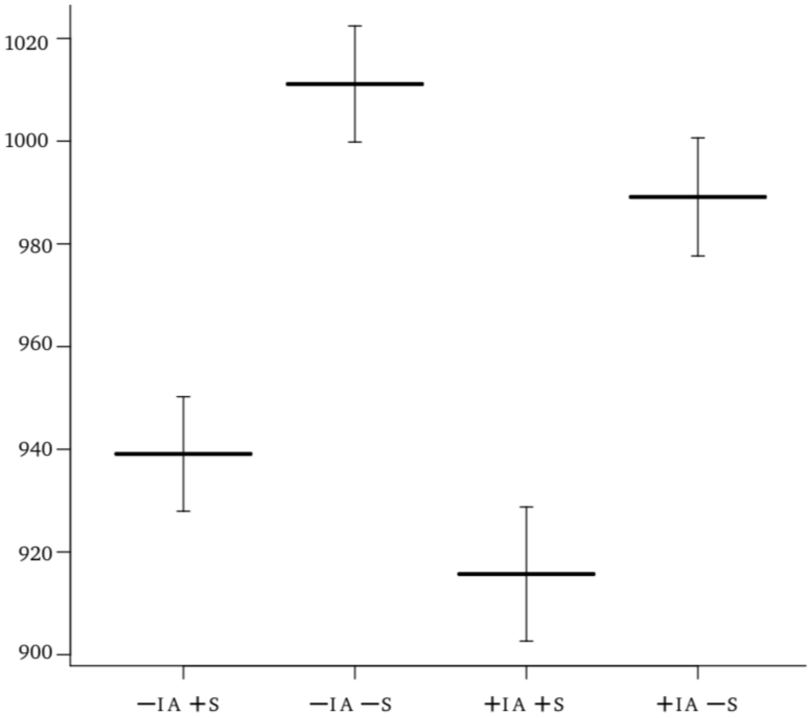


Figure A.15: Error-bar plot of mean reaction times for all four conditions (−IA +S, −IA −S, +IA +S, +IA −S) revealing a significant effect of both ±IA and of ±S, with no interaction between the two experimental manipulations.

Response latencies	*m*	*sd*	maximum	minimum	% errors
SEMANTICALLY CONGRUENT					
−IA	939.09	202.87	1494.07	359.01	11%
+IA	915.70	239.43	1561.71	25.01	7%
SEMANTICALLY INCONGRUENT					
−IA	1011.12	209.89	1511.31	572.25	4.5%
+IA	989.14	206.93	1564.43	573.43	3%

Table A.12: Reaction times per condition. Data analysis revealed a significant effect both of ±IA and of ±S, with no interaction between the two conditions.

Les résultats révèlent un effet d'interaction significatif : le N400 situé au fronto-centrale était plus grand pour les mots présentés sans IA (figure A.16).

De plus, nous avons trouvé un effet de congruence sémantique, sur la région centro-pariétale (la région traditionnelle pour N400), qui était plus grande pour le mot −IA que pour les mots +IA (figure A.17).

Enfin, nous avons observé une interaction telle que ±IA a continué à moduler l'amplitude N400, mais seulement dans les phrases qui

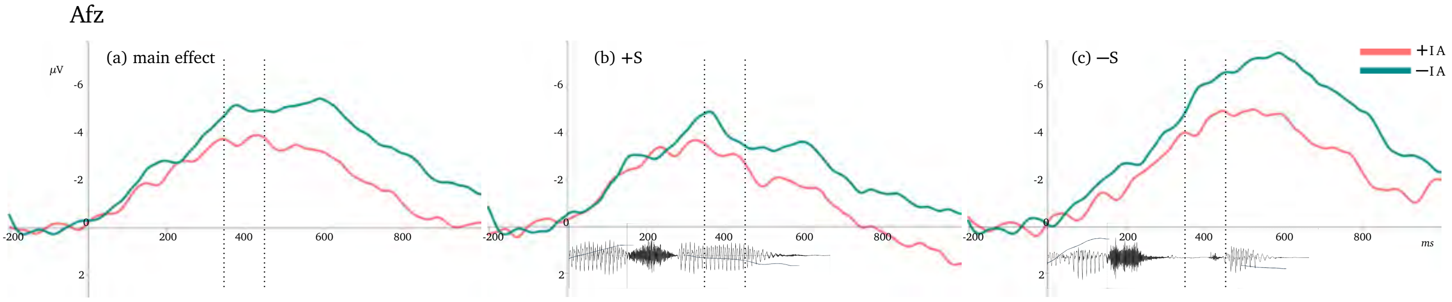


Figure A.16: Grand average metrical N400 in the ±IA condition (−IA in green, +IA in pink), recorded at the Afz (anterio-frontal) electrode for: (a) main effect, (b) congruent sentences, (c) incongruent sentences. The tested time-window is indicated by dashed vertical lines. Furthermore to indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [løsemi] (*leucémie*, +S) and [dɑ̃site] (*densité*, −S) are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are lowpass filtered at 10 Hz.

étaient sémantiquement inconciliables (figure A.16). De plus, comme les participants se sont penchés sur le contenu sémantique des phrases, le résultat souligne l'automaticité du traitement de l'accent.

Table A.13: Overview linear mixed models. The model fitting the data best takes semantic congruency as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent significantly contributes to the prediction of reaction times when it is the only fixed effect and marginally contributes when entered together with semantic congruency.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	976.30***	977.21***	1016.63***	966.57***	1006.12***	1008.88***
	(25.07)	(26.89)	(28.18)	(27.24)	(28.52)	(28.91)
congruency			−78.46***		−78.27***	−83.64***
			(16.81)		(16.84)	(19.14)
±IA				21.00*	20.58*	15.20
				(9.37)	(9.31)	(13.03)
±IA:congruency						10.66
						(18.04)
AIC	17757.83	17567.17	17542.20	17557.93	17533.11	17527.13
BIC	17773.40	17603.50	17583.72	17599.45	17579.82	17579.03
Log Likelihood	−8875.92	−8776.58	−8763.10	−8770.96	−8757.55	−8753.57
Num. obs.	1326	1326	1326	1326	1326	1326
Num. groups: subj	20	20	20	20	20	20
Var: subj (Intercept)	11996.42	12415.76	12432.85	12339.64	12363.26	12377.19
Var: Residual	36671.85	25898.44	25946.90	25905.44	25948.81	25959.58
Num. groups: congr:stimuli		160	160	160	160	160
Num. groups: ±IA:stimuli		160	160	160	160	160
Num. groups: stimuli		80	80	80	80	80
Var: congr.stimuli (Intercept)		11031.88	7977.87	11057.78	8021.99	8016.48
Var: ±IA.stimuli (Intercept)		417.00	366.70	234.50	193.37	202.53
Var: stimuli (Intercept)		267.65	2153.78	591.00	336.86	1534.69

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

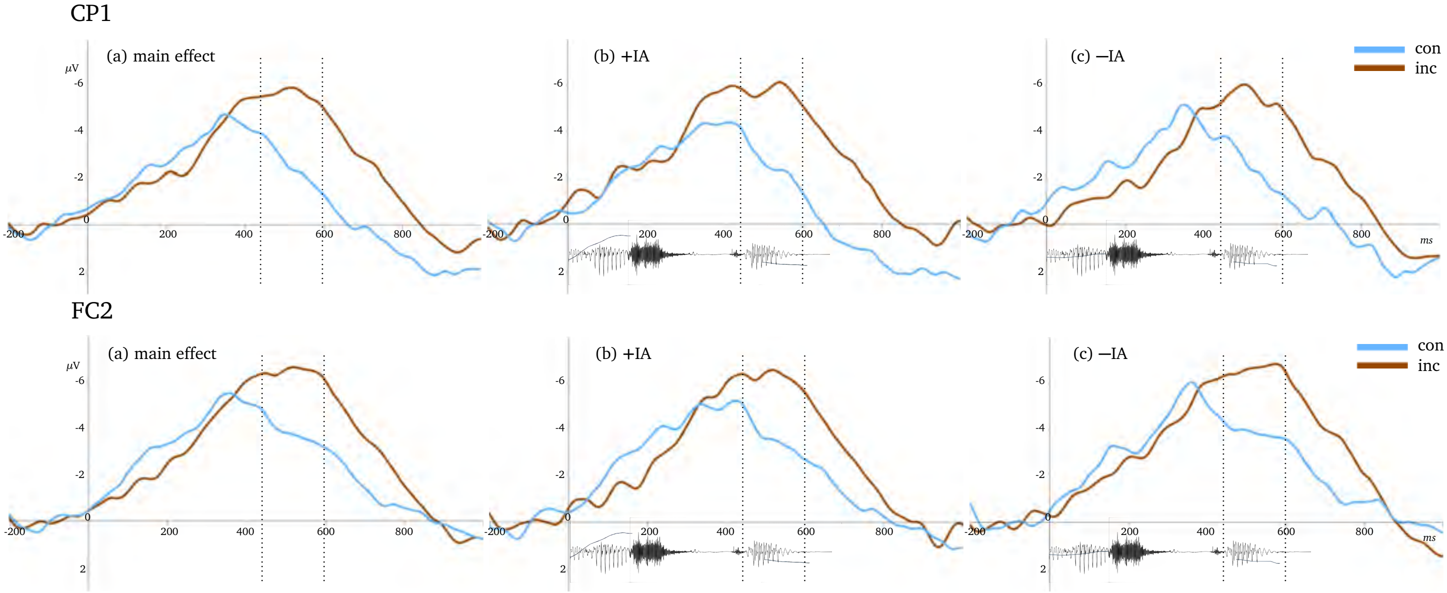


Figure A.17: Grand average semantic N400 in the ±S condition (−S in brown, +S in blue), recorded at the CP1 (centro-parietal, top) and FC2 (fronto-central, bottom) electrodes for: (a) main effect, (b) +IA, (c) −IA. The tested time-window is indicated by dashed vertical lines. To indicate the timing of the amplitude modulation with respect to the speech signal, the oscillograms and f_0 deflections of [dãsite] +IA and −IA are plotted in the background. Negativity is plotted upwards and, for ease of presentation only, ERP waveforms are low-pass filtered at 10 Hz.

Conclusion : Nos données montrent que l'accent initial est codé au niveau lexical et anticipé par les auditeurs. La présentation de mots sans l'accent initial entrave l'accès lexical et cascade le processus de compréhension de la parole pour se répercute sur le traitement sémantique.

EN RÉSUMENT, nous démontrons que l'accentuation joue un rôle crucial dans la compréhension de la parole en français.

B Stimuli lists

B.1 Corpus—Lexical decision

Lexical word		Pseudoword	
barbeles	leopards	becebaire	maromo
batiments	libellules	benalat	mazaquier
bicyclette	matinees	bepotere	metobo
bijoutiers	ministere	bugolier	micoria
boulimie	mobilier	cajitiste	mitimier
bricolage	narrateur	chalugueur	nacamien
candidat	nettoyage	charibol	namalion
caramel	nicotine	chibute	nemuphile
casino	noctambules	chofinet	nuzote
chapiteaux	nouveaute	cotemien	pecima
charabia	paradis	couvido	quebutal
cheminee	parasol	darofeur	ropamion
chocolat	perroquet	ditelia	rubinel
citoyens	physiciens	dontini	samuree
compagnon	protections	faruteur	silumiette
couturier	robinet	fepino	sinodi
debarras	saisonniers	finutie	solunade
defile	salarie	focagnon	suletier
delegues	seminaires	fudiri	talasol
depute	sympathie	galofet	tivonet
diffusion	tabourets	gotimia	toforol
figurants	tragedie	gucidie	toutaro
funambules	tyrannie	juluveur	vavitile
garderie	vacanciers	jurenia	vitonesse
gaspacho	vegetaux	kubeto	vonolion

geranium	veloute	lagimot	votacaire
gueridon	vibration	lametier	vupacier
judokas	videos	lizabo	vutala
lampadaires	voyageurs	losopat	zoduquet
langoustines	zigotos	lumarie	zopasson

B.2 Corpus—Semantic Congruency

	Semantically congruent +s		**Semantically incongruent −s**
001a	Ces friandises contiennent des calories	001b	Ces friandises contiennent des péronés
004a	Certains skieurs pratiquent la randonnée	004b	Certains skieurs pratiquent la pénurie
011a	Le garçonnet dort en pyjama	011b	Le garçonnet dort en délégué
012a	L’écrivain a écrit une parodie	012b	L’écrivain a écrit une boulimie
014a	Cette rivière est infestée de piranhas	014b	Cette rivière est infestée de désaveux
015a	Ce film se joue au cinéma	015b	Ce film se joue au thermostat
020a	Le dompteur maîtrise ses animaux	020b	Le dompteur maîtrise ses lavabos
021a	Le reporteur branche sa caméra	021b	Le reporteur branche sa fourberie
022a	Le plombier répare le robinet	022b	Le plombier répare le tournedos
023a	Cet accident est une tragédie	023b	Cet accident est une bourgeoisie
024a	Cet oiseau est un canari	024b	Cet oiseau est un pétunia
028a	Le soudeur allume son chalumeau	028b	Le soudeur allume son célibat
029a	Le pêcheur prépare son moulinet	029b	Le pêcheur prépare son camélia
031a	Cette danseuse se produit dans ce cabaret	031b	Cette danseuse se produit dans ce détritus
032a	Le cuisinier cherche un saladier	032b	Le cuisinier cherche un singulier
033a	Des vacanciers louent ce bungalow	033b	Des vacanciers louent ce cachalot
034a	Ces agriculteurs élèvent des bovidés	034b	Ces agriculteurs élèvent des pédaliers
035a	Certaines plantes aident á soigner ces maladies	035b	Certaines plantes aident á soigner ces parvenus
036a	Ce compositeur joue avec les harmonies	036b	Ce compositeur joue avec les libéraux
037a	Les soldats combattent la tyrannie	037b	Les soldats combattent la palmeraie
038a	Le dictateur rassemble ses généraux	038b	Le dictateur rassemble ses farineux
040a	Elle a une peur phobique des araignées	040b	Elle a une peur phobique des cerisiers
041a	L’annuaire donne la liste des abonnés	041b	L’annuaire donne la liste des capitaux
045a	Son complice lui fournit un alibi	045b	Son complice lui fournit un nirvana
046a	Sur le banc public, j’ai vu des amoureux	046b	Sur le banc public, j’ai vu des mausolées

047a	Elle dit á ses enfants de ne pas parler aux inconnus	047b	Elle dit á ses enfants de ne pas parler aux éboulis
048a	Le vieux sage était un érudit	048b	Le vieux sage était un chandelier
049a	Elle va au zoo pour voir des animaux	049b	Elle va au zoo pour voir des escabots
052a	Sa défense est assurée par un avocat	052b	Sa défense est assurée par un appétit
053a	Le cheval sent son écurie	053b	Le cheval sent son cheminot
054a	Ce n'est pas un phoque mais une otarie	054b	Ce n'est pas un phoque mais une embardée
056a	Cette civilisation a atteint son apogée	056b	Cette civilisation a atteint son wagonnet
057a	Le peintre a posé le tableau sur son chevalet	057b	Le peintre a posé le tableau sur son renégat
059a	Ces pauvres gens sont des malheureux	059b	Ces pauvres gens sont des manitous
060a	Jean a rangé les vieilles affaires dans le débarras	060b	Jean a rangé les vieilles affaires dans le renouveau
061a	Les prêtres ont fait voeu de chasteté	061b	Les prêtres ont fait voeu de dindonneau
063a	En orient, les hommes fument le narguilé	063b	En orient, les hommes fument le rectorat
064a	Pour faire réparer ma montre, j'irai chez un horloger	064b	Pour faire réparer ma montre, j'irai chez un parolier
066a	Pour aller en Egypte, il faut un vaccin contre la malaria	066b	Pour aller en Egypte, il faut un vaccin contre les bordereaux
068a	Dans sa nouvelle maison, il a changé le mobilier	068b	Dans sa nouvelle maison, il a changé le pancréas
070a	Aux Antilles, il y a des champs de bananiers	070b	Aux Antilles, il y a des champs de baleiniers
071a	Pour les vendanges, il fait appel á des saisonniers	071b	Pour les vendanges, il fait appel á des timoniers
072a	La serveuse a tâché son tablier	072b	La serveuse a tâché son doctorat
074a	Les jongleurs se rassemblent sous le chapiteau	074b	Les jongleurs se rassemblent sous le colibri
076a	Elle a acheté cette belle bague chez ce bijoutier	076b	Elle a acheté cette belle bague chez ce jardinier
078a	Les grands filets de pêche sont traînés par les chalutiers	078b	Les grands filets de pêche sont traînés par les teinturiers
080a	Le Vatican rassemble beaucoup de religieux	080b	Le Vatican rassemble beaucoup de dynamos
081a	Paris attire beaucoup de japonais	081b	Paris attire beaucoup de bâtonnets
082a	Le samouraï enfile son kimono	082b	Le samouraï enfile son cagibi
083a	Les greffes de moelle soignent la leucémie	083b	Les greffes de moelle soignent la densité

085a	Le coq chante très tôt dans la matinée	085b	Le coq chante très tôt dans la félonie
086a	Dans la prison, les gardiens contiennent les mutinés	086b	Dans la prison, les gardiens contiennent lesomégas
087a	Dans l'avion, l'hôtesse renseigne les passagers	087b	Dans l'avion, l'hôtesse renseigne les pédalos
088a	Le jardinier plante des tomates dans son potager	088b	Le jardinier plante des tomates dans son postulat
089a	Pour se soigner, certaines personnes vont voir un rebouteux	089b	Pour se soigner, certaines personnes vont voir un pintadeau
091a	Pour l'épreuve de français, les élèves ont écrit un résumé	091b	Pour l'épreuve de français, les élèves ont écrit un rescapé
092a	Le chef d'orchestre dirige une symphonie	092b	Le chef d'orchestre dirige une tabagie
093a	La manifestation a regroupé beaucoup de syndicats	093b	La manifestation a regroupé beaucoup de bassinets
095a	L'été, les champs sont remplis de coquelicots	095b	L'été, les champs sont remplis de consulats
096a	Mon vin préféré est le beaujolais	096b	Mon vin préféré est le karaté
097a	Sur la plage, elle met son bikini	097b	Sur la plage, elle met son quolibet
098a	Il fait des études de biologie	098b	Il fait des études de fantaisie
100a	Après la bataille, lesindiensfument le calumet	100b	Après la bataille, lesindiensfument le rococo
101a	Il passe ses journées sur le canapé	101b	Il passe ses journées sur le pugilat
108a	Pendant le carnaval, les enfants lancent des confettis	108b	Pendant le carnaval, les enfants lancent des cavités
110a	La Martinique est un coin de paradis	110b	La Martinique est un coin de pigeonneau
113a	Ce vent est aussi chaud que le Sirocco	113b	Ce vent est aussi chaud que le roudoudou
114a	Une mer calme est idéale pour les ricochets	114b	Une mer calme est idéale pour les libellés
115a	Ce chercheur est une sommité	115b	Ce chercheur est une mélopée
117a	Un des sept péchés capitaux est la Vanité	117b	Un des sept péchés capitaux est la rapsodie
118a	Il a été victime de la vendetta	118b	Il a été victime de la litanie
120a	Pour naviguer, il faut savoir utiliser les alizés	120b	Pour naviguer, il faut savoir utiliser les insoumis
121a	La salade de fruits est meilleure avec des ananas	121b	La salade de fruits est meilleure avec des inédits

123a	A la pétanque, il faut placer la boule au plus prêt du cochonnet	123b	A la pétanque, il faut placer la boule au plus prêt du démêlé
126a	Cet explorateur a longtemps vécu avec des eskimos	126b	Cet explorateur a longtemps vécu avec des minuties
127a	Pour repeindre son plafond, il a utilisé un escabeau	127b	Pour repeindre son plafond, il a utilisé un infini
128a	L'année dernière, la Reine d'Angleterre a fêté son jubilé	128b	L'année dernière, la Reine d'Angleterre a fêté son pékinois
129a	Mes enfants n'aiment pas trop les salsifis	129b	Mes enfants n'aiment pas trop les endémies
130a	La forêt méditerranéenne est majoritairement composée de résineux	130b	La forêt méditerranéenne est majoritairement composée de sympathie
131a	Pour pêcher la truite, le meilleur appât est l'asticot	131b	Pour pêcher la truite, le meilleur appât est l'artichaut

C Analyses—Behavioral and EEG

C.1 MMN—FA

C.1.1 Study overview

RQ: Is there a long-term expectation for word to be marked with FA (similar as study MMN-IA, Aguilera et al. 2014)?

Hypotheses: The MMN will be bigger when deviants are presented −FA than when deviants are presented +FA.

C.1.1.1 Procedure

Nr participants: 19 listeners (2 excluded, 8 dev −FA and 11 dev +FA)

Nr stimuli per condition: 986 standards, 106 deviants (2 stimuli, 4 lists)

Task: No

ISI: 600 ms

C.1.1.2 Stimuli

Manipulation: Mostly duration, equal length last syllable and 2nd (unstressed) syllable (see section 5.1.3 for more information).

Some descriptives:

Table C.1: Durations of 'casino' and 'paradis': total duration, 1st syllable, 2nd syllable, 3rd syllable and the onset of the 3rd (FA) syllable.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd syllable
	ms	f_0	*ms*	f_0	*ms*	f_0	*ms*	f_0	onset
CASINO									
+FA	503.3	117.8	112.9	125.8	157.1	122.2	233.3	111.0	269.0
−FA	500.8	119.0	146.9	125.1	178.1	122.3	178.8	110.7	325.0
PARADIS									
+FA	459.0	106.5	121.5	121.7	111.6	118.9	225.9	93.4	233.1
−FA	456.5	110.9	168.8	123.3	139.3	119.1	148.3	95.1	308.1

Nr electrodes:	64
Reference:	Mastoids
Filter and down-sampling:	0.01 − 30 bandpass, 128 Hz
Epoch length:	−100 − 1000

C.1.1.3 Analysis

Descriptive statistic:

deviant +FA, stimulus ‘paradis’

Peak latencies for MMN - ‘paradis’

chlabel	AF7	AF3	F1	F5	FC1	C3	AF8	AF4	Afz	Fz	F2	F6	FC2	FCz	Cz	C4
subjects	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
mean	584.8213	595.9821	582.5893	556.9196	575.8929	582.5893	584.8216	575.8929	569.1964	588.1694	581.4731	592.6339	584.8216	599.3303	609.3750	609.3750
sd	49.08600	28.42506	53.53276	37.93012	56.28517	42.51815	38.12086	49.35184	47.39919	47.67444	44.61980	44.06228	42.89271	32.74890	27.80498	27.43674
max	640.625	640.625	648.438	617.188	648.438	632.812	640.625	648.438	648.438	640.625	648.438	640.625	648.438	648.438	648.438	648.438
min	500.000	554.688	507.812	507.812	507.812	523.438	523.438	507.812	507.812	507.812	507.812	515.625	515.625	539.062	562.500	562.500

Peak latencies for deviant +FA - ‘paradis’

mean	584.8213	595.9821	582.5893	556.9196	575.8929	582.5893	584.8216	575.8929	569.1964	588.1694	581.4731	592.6339	584.8216	599.3303	609.3750	609.3750
sd	49.08600	28.42506	53.53276	37.93012	56.28517	42.51815	38.12086	49.35184	47.39919	47.67444	44.61980	44.06228	42.89271	32.74890	27.80498	27.43674
max	640.625	640.625	648.438	617.188	648.438	632.812	640.625	648.438	648.438	640.625	648.438	640.625	648.438	648.438	648.438	648.438
min	500.000	554.688	507.812	507.812	507.812	523.438	523.438	507.812	507.812	507.812	507.812	515.625	515.625	539.062	562.500	562.500

Peak latencies for standard −FA - ‘paradis’

mean	495.5360	495.5357	472.0983	472.0983	483.2590	487.7234	495.5359	501.1161	521.2054	472.0983	472.0983	512.2769	486.6074	476.5627	482.1430	524.5539
sd	49.58677	43.46456	22.48812	22.48812	29.72457	55.76628	42.27808	61.75148	65.78497	22.48812	22.48812	58.61234	31.15683	28.52717	35.14581	71.94674
max	562.500	570.312	507.812	507.812	531.250	609.375	570.312	625.000	640.625	507.812	507.812	625.000	531.250	531.250	531.250	632.812
min	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	460.938

deviant −FA, stimulus ‘paradis’

Peak latencies for MMN - ‘paradis’

chlabel	AF7	AF3	F1	F5	FC1	C3	AF8	AF4	Afz	Fz	F2	F6	FC2	FCz	Cz	C4
subjects	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
mean	498.4376	510.9376	493.7500	487.5002	493.7500	492.1874	492.1876	493.7500	493.7500	492.1876	471.8752	493.7500	470.3126	471.8752	495.3126	543.7502
sd	47.32823	70.65889	45.01519	33.42085	47.96880	48.15928	44.87940	45.01519	45.01519	44.87940	22.50765	47.96880	23.69660	22.50765	46.41673	46.08681
max	570.312	625.000	562.500	523.438	570.312	570.312	562.500	562.500	562.500	562.500	500.000	570.312	500.000	500.000	570.312	609.375
min	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	492.188

Peak latencies for deviant −FA - ‘paradis’

mean	476.5628	501.5628	479.6876	476.5626	464.0626	506.2502	464.0628	464.0626	464.0626	462.5002	478.1252	498.4374	462.5002	460.9376	475.0002	575.0002
sd	30.75803	71.04669	34.76336	29.23163	20.37249	63.80469	16.20051	20.37249	20.37249	16.93730	31.92639	76.22672	16.93730	13.53165	23.69625	70.22560
max	523.438	617.188	531.250	515.625	500.000	609.375	492.188	500.000	500.000	492.188	523.438	632.812	492.188	484.375	507.812	648.438
min	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	460.938

Peak latencies for standard +FA - ‘paradis’

mean	525.0002	532.8124	535.9376	542.1874	529.6874	523.4376	537.5002	539.0626	517.1876	535.9376	539.0628	564.0628	529.6874	529.6874	562.5002	565.6254
sd	53.38873	69.30690	73.41267	62.88954	80.32077	72.45047	75.42223	73.49603	60.86752	73.41267	73.90985	88.62997	80.32077	80.32077	86.99650	86.22124
max	617.188	609.375	617.188	609.375	625.000	617.188	617.188	617.188	617.188	617.188	617.188	648.438	625.000	625.000	648.438	648.438
min	484.375	460.938	460.938	476.562	460.938	460.938	460.938	468.750	460.938	460.938	460.938	453.125	460.938	460.938	460.938	460.938

deviant +FA, stimulus ‘casino’

Peak latencies for MMN - ‘casino’

chlabel	AF7	AF3	F1	F5	FC1	C3	AF8	AF4	Afz	Fz	F2	F6	FC2	FCz	Cz	C4
subjects	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
mean	599.6095	578.1250	580.0780	585.9375	599.6092	591.7970	578.1250	607.4217	597.6560	599.6090	605.4685	591.7967	603.5155	578.1250	599.6092	605.4690
sd	40.53264	40.84508	40.02713	45.55437	38.99720	92.96010	40.84508	48.31793	41.09322	42.49273	48.37055	56.47216	48.73717	36.64409	41.52369	47.94787
max	625.000	617.188	625.000	625.000	632.812	648.438	617.188	648.438	632.812	632.812	648.438	640.625	648.438	625.000	632.812	648.438
min	539.062	539.062	539.062	539.062	546.875	453.125	539.062	539.062	539.062	539.062	539.062	539.062	539.062	539.062	546.875	546.875

Peak latencies for deviant +FA - ‘casino’

mean	576.1717	537.1092	539.0622	580.0783	541.0158	599.6095	523.4375	531.2500	541.0155	539.0622	539.0622	582.0315	583.9845	541.0155	597.6560	589.8440
sd	32.13315	58.93994	61.51560	38.99757	55.74662	50.78141	53.36954	57.40978	48.73717	46.43906	49.82054	48.37055	48.73717	51.96933	31.57397	42.55265
max	617.188	578.125	593.750	617.188	585.938	648.438	578.125	578.125	585.938	578.125	578.125	648.438	648.438	585.938	640.625	648.438
min	539.062	453.125	453.125	539.062	460.938	531.250	453.125	453.125	476.562	476.562	468.750	539.062	531.250	468.750	570.312	546.875

Peak latencies for standard −FA - ‘casino’

mean	542.9690	541.0155	541.0155	539.0622	541.0155	511.7185	542.9688	539.0625	541.0155	541.0155	539.0622	541.0155	541.0155	539.0622	496.0935	544.9220
sd	50.02464	53.13085	53.13085	51.42817	53.13085	51.62607	51.62559	50.63104	53.13085	53.13085	51.42817	48.73674	53.13085	51.42817	18.59754	48.73671
max	585.938	593.750	593.750	593.750	593.750	585.938	593.750	585.938	593.750	593.750	593.750	593.750	593.750	593.750	515.625	593.750
min	492.188	484.375	484.375	484.375	484.375	476.562	484.375	484.375	484.375	484.375	484.375	492.188	484.375	484.375	476.562	492.188

deviant −FA, stimulus ‘casino’

Peak latencies for MMN - ‘casino’

chlabel	AF7	AF3	F1	F5	FC1	C3	AF8	AF4	Afz	Fz	F2	F6	FC2	FCz	Cz	C4
subjects	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
mean	500.0000	494.7917	505.2083	513.0207	492.1873	492.1873	502.6040	497.3957	492.1873	489.5830	473.9580	484.3747	489.5830	486.9790	489.5833	471.3543
sd	23.43800	19.66123	18.04220	56.51718	20.67012	20.67012	22.55303	25.11357	27.06387	22.55303	4.51026	20.66974	22.55303	25.11378	29.57804	25.11357
max	523.438	515.625	515.625	578.125	515.625	515.625	515.625	515.625	523.438	515.625	476.562	507.812	515.625	515.625	523.438	500.000
min	476.562	476.562	484.375	476.562	476.562	476.562	476.562	468.750	476.562	476.562	468.750	468.750	476.562	468.750	468.750	453.125

Peak latencies for deviant −FA - ‘casino’

mean	536.4583	528.6460	520.8337	528.6457	526.0417	526.0417	523.4373	536.4583	533.8540	520.8330	526.0417	520.8333	520.8333	520.8333	520.8333	575.5207
sd	36.92054	39.32152	35.22877	43.02765	43.02762	43.02762	51.23022	29.57725	32.52587	47.09149	50.83154	54.87320	54.87320	50.83107	50.83107	99.23150
max	578.125	570.312	554.688	570.312	570.312	570.312	578.125	570.312	570.312	570.312	578.125	578.125	578.125	570.312	570.312	632.812
min	507.812	492.188	484.375	484.375	484.375	484.375	476.562	515.625	507.812	476.562	476.562	468.750	468.750	468.750	468.750	460.938

Peak latencies for standard +FA - ‘casino’

mean	572.9167	572.9167	585.9377	539.0627	591.1460	580.7293	572.9167	572.9167	572.9167	572.9167	575.5207	559.8960	575.5210	578.1253	575.5210	562.5000
sd	11.934118	11.934118	28.168647	74.526657	31.573802	9.021675	11.934118	11.934118	11.934118	11.934118	9.021675	27.436836	11.934009	13.531936	11.934009	28.168508
max	585.938	585.938	617.188	585.938	625.000	585.938	585.938	585.938	585.938	585.938	585.938	585.938	585.938	585.938	585.938	585.938
min	562.500	562.500	562.500	453.125	562.500	570.312	562.500	562.500	562.500	562.500	570.312	531.250	562.500	562.500	562.500	531.250

Statistical analysis: $t_{\max}$ mass univariate permutation test, 2500 permutations in matlab

Electrodes: 11 electrodes (Fz, Cz, FC1, FC2, C3, C4, F1, F2, Fh, Fh, FCz)

Time-windows: $551-651$ for FA

C.1.1.4 Results

Within: *Effect of* −FA *as deviant; marginal effect* (critical t-score: ± 4.2958, test-wise α: 0.0528, $p = 0.0652$).

difference between both MMN, *significant* (critical t-score: ± 3.1505, test-wise α: 0.010324, $p = 0.0436$).

Effect of +FA *as deviant; not significant* (critical t-score: ± 4.3095, test-wise α: 0.001538, $p = 0.8396$).

Between: *within* −FA, *not significant* (critical t-score: ± 4.3954, test-wise α: 0.001348, $p = 0.8444$).

within +FA, *significant* (critical t-score: ± 3.7416, test-wise α: 0.003837, p between 0.048 and 0.0152).

C.1.2 All non-parametric results

C.1.2.1 $t_{\max}$ permutation tests

C.1.2.1.1 Within mmns

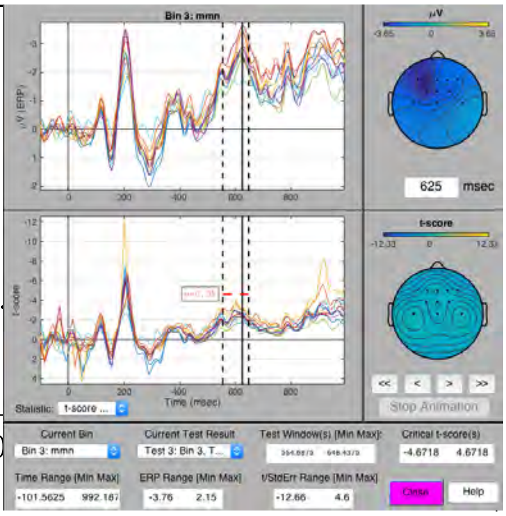
```
GND=tmaxGND(GND,3 , 'time_wind',[551  651], 'include_chans',
        {'Fz','Cz','FC1','FC2','C3','C4','F1','F2','F3','F4','FCz'});
```

standard +FA; *deviant* −FA

8 out of 8 participants have data in relevant bin. Attempting to use time boundaries of 551 to 651 ms for hypothesis test. Testing null hypothesis that the grand average ERPs in Bin 3 (mmn) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	11
Nr time points	13
Nr comparisons	143
Nr participants	8
t-score degrees of freedom	7
Estimated actual family-wise alpha level	0.052800
Critical t-score(s)	-4.2958 and 4.
Test-wise alpha level	0.003585.
Bonferroni test-wise alpha	0.000350.

ERPs are NOT significantly different from zero ($\alpha = 0.05$)
All $p-values >= 0.065200$

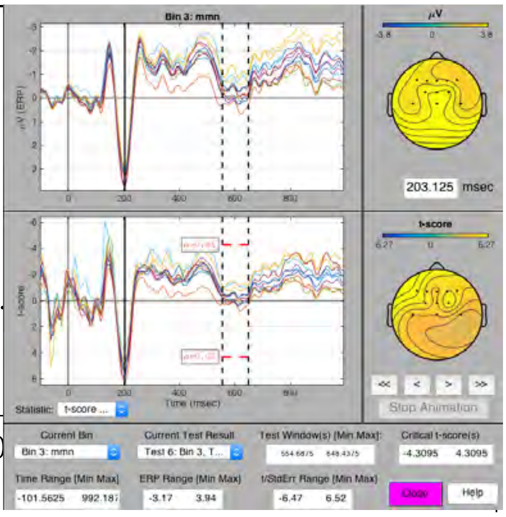


Current Bin: Bin 3: mmn
Current Test Result: Test 3: Bin 3, T...
Critical t-score(s): -4.6718 4.6718
Time Range [Min Max]: -101.5625 992.187
ERP Range [Min Max]: -3.76 2.15
t/StdErr Range [Min Max]: -12.66 4.6

standard −FA; *deviant* +FA

11 out of 11 participants have data in relevant bin. Attempting to use time boundaries of 551 to 651 ms for hypothesis test. Testing null hypothesis that the grand average ERPs in Bin 3 (mmn) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	11
Nr time points	13
Nr comparisons	143
Nr participants	11
t-score degrees of freedom	10
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.3095 and 4.
Test-wise alpha level of 0.001538.	
Bonferroni test-wise alpha	0.000350.
ERPs are NOT significantly different from zero ($\alpha = 0.05$) All $p - values >= 0.839600$	



Statistic: t-score ...

Current Bin: Bin 3: mmn

Current Test Result: Test 6: Bin 3, T...

Test Window(s) [Min Max]: 554.6875 648.4375

Critical t-score(s): -4.3095 4.3095

Time Range [Min Max]: -101.5625 992.187

ERP Range [Min Max]: -3.17 3.94

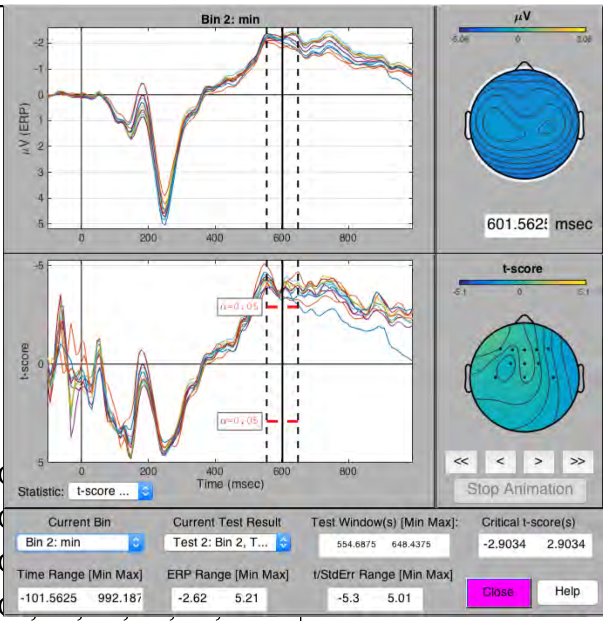
t/StdErr Range [Min Max]: -6.47 6.52

C.1.2.1.2 Between mmns

difference between both MMNs

Testing null hypothesis that the grand average ERPs in Bin 3 (mmn) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels
Nr time points
Nr comparisons
Nr participants
t-score degrees of freedom
Estimated actual family-wise alpha level
Critical t-score(s)
That corresponds to a test-wise alpha level
Bonferroni test-wise alpha
Significant differences from zero:
554 ms, electrode(s): Fz, Cz, FC1, FC2, C3,
562 ms, electrode(s): Fz, Cz, FC1, FC2, C3,
570 ms, electrode(s): Fz, Cz, FC1, FC2, C3,
578 ms, electrode(s): Fz, Cz, FC1, FC2, C3,
585 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
593 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
601 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
609 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
617 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
625 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
632 ms, electrode(s): Fz, Cz, FC1, FC2, C3, C4, F1, F2, F3, F4, FCz.
640 ms, electrode(s): Fz, Cz, FC2, C3, C4, F1, F2, F3, F4, FCz.
648 ms, electrode(s): Fz, Cz, FC2, C3, C4, F1, F2, F3, F4, FCz.
$p - values$ between $0.043600 - 0.043600$



601.562 msec

Statistic: t-score ...

Current Bin: Bin 2: min

Current Test Result: Test 2: Bin 2, T...

Test Window(s) [Min Max]: 554.6875 648.4375

Critical t-score(s): -2.9034 2.9034

Time Range [Min Max]: -101.5625 992.187

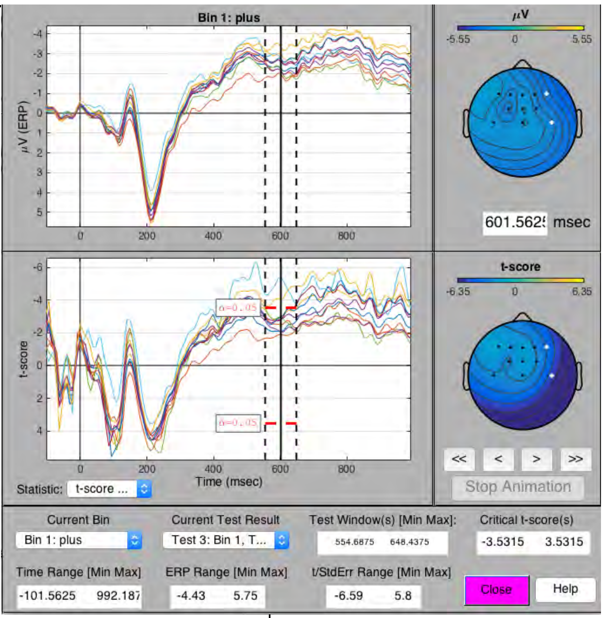
ERP Range [Min Max]: -2.62 5.21

t/StdErr Range [Min Max]: -5.3 5.01

deviant +FA − *standard* +FA

Testing null hypothesis that the grand average ERPs in Bin 1 (devplus-stanplus) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels
Nr time points
Nr comparisons
Nr participants
t-score degrees of freedom
Estimated actual family-wise alpha level
Critical t-score(s)
That corresponds to a test-wise alpha level
Bonferroni test-wise alpha
Significant differences from zero:
554 ms, electrode(s): C4.
562 ms, electrode(s): C4.
570 ms, electrode(s): C4.
578 ms, electrode(s): C4, F4.
585 ms, electrode(s): C4, F4.
593 ms, electrode(s): C4, F4.
601 ms, electrode(s): C4, F4.
609 ms, electrode(s): C4, F4.
617 ms, electrode(s): C4, F4.
625 ms, electrode(s): C4, F4.
632 ms, electrode(s): C4, F4.
640 ms, electrode(s): C4, F4.
648 ms, electrode(s): F1, F4.
$p - values$ between $0.048000 - 0.015200$



601.562! msec

Statistic: t-score ...

Current Bin: Bin 1: plus
Current Test Result: Test 3: Bin 1, T...
Test Window(s) [Min Max]: 554.6875 648.4375
Critical t-score(s): -3.5315 3.5315
Time Range [Min Max]: -101.5625 992.187
ERP Range [Min Max]: -4.43 5.75
t/StdErr Range [Min Max]: -6.59 5.8

deviant —FA – *standard* —FA

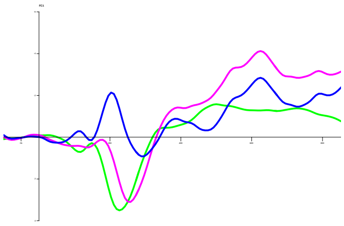
Testing null hypothesis that the grand average ERPs in Bin 2 (devmin-stanmin) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	11
Nr time points	13
Nr comparisons	143
Nr participants	11
t-score degrees of freedom	10
Estimated actual family-wise alpha level	0.050800
Critical t-score(s)	-4.394 and 4.394
Test-wise alpha level	0.001348.
Bonferroni test-wise alpha	0.000350.
ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed	
All $p - values >= 0.844400$	

C.1.3 All erp plots

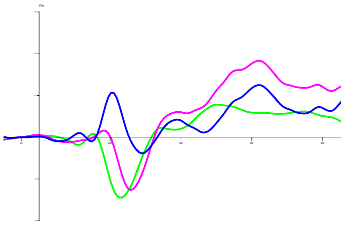
C.1.3.1 Deviant without final accent

FC1



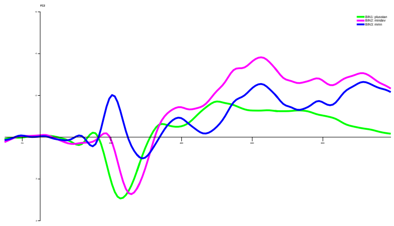
green:plus fa magenta:min fa blue:MMN

FCz



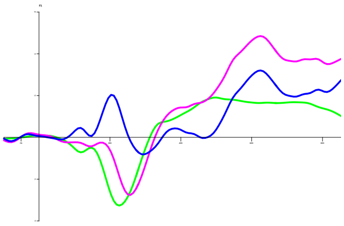
green:plus fa magenta:min fa blue:MMN

FC2



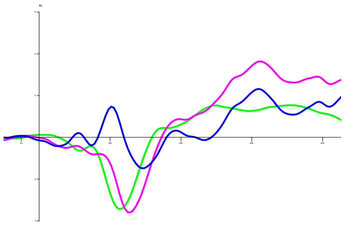
green:plus fa magenta:min fa blue:MMN

F1



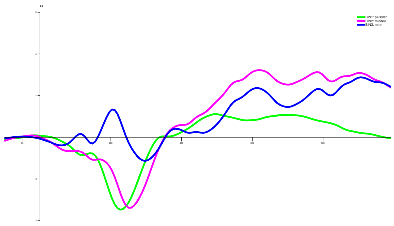
green:plus fa magenta:min fa blue:MMN

Fz



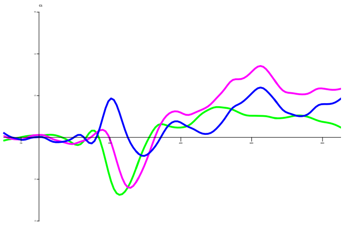
green:plus fa magenta:min fa blue:MMN

F2



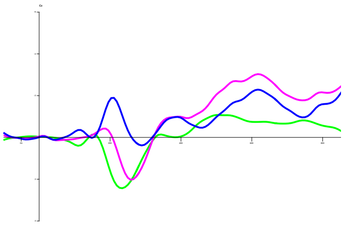
green:plus fa magenta:min fa blue:MMN

C1



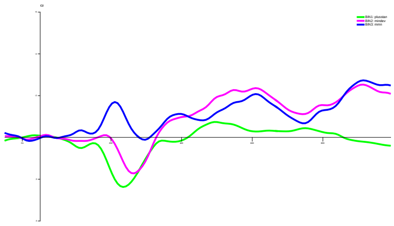
green:plus fa magenta:min fa blue:MMN

Cz



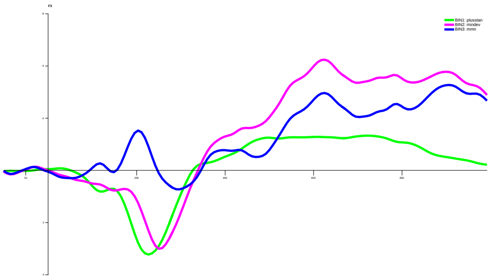
green:plus fa magenta:min fa blue:MMN

C2



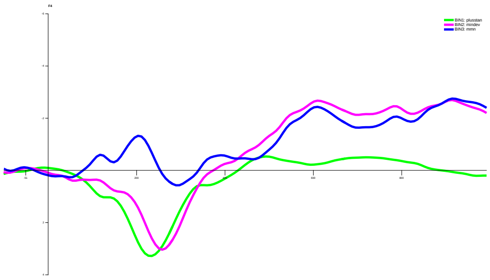
green:plus fa magenta:min fa blue:MMN

F3



green:plus fa magenta:min fa blue:MMN

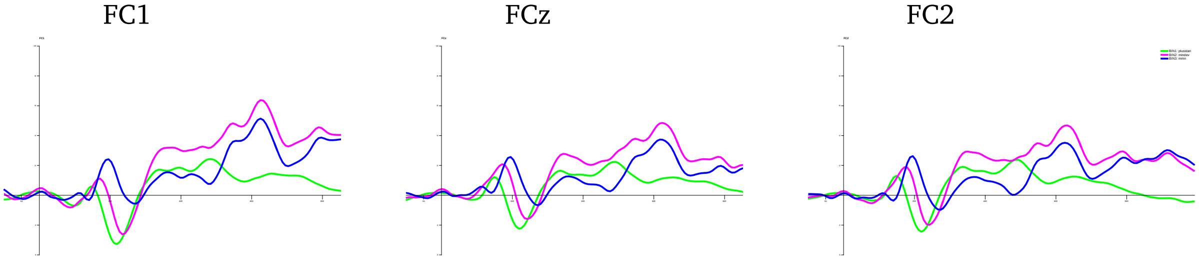
F4



green:plus fa magenta:min fa blue:MMN

Figure C.1: MMN-FA — raw ERP's for standard +fa and deviant -fa.

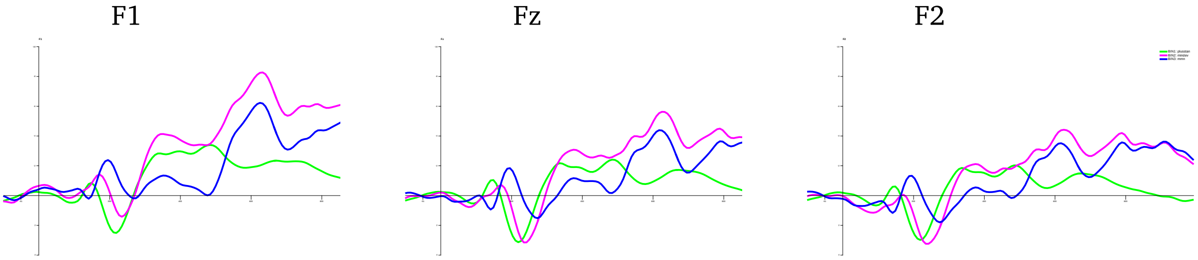
C.1.3.2 Deviant without final accent — casino (3 subjects)



green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

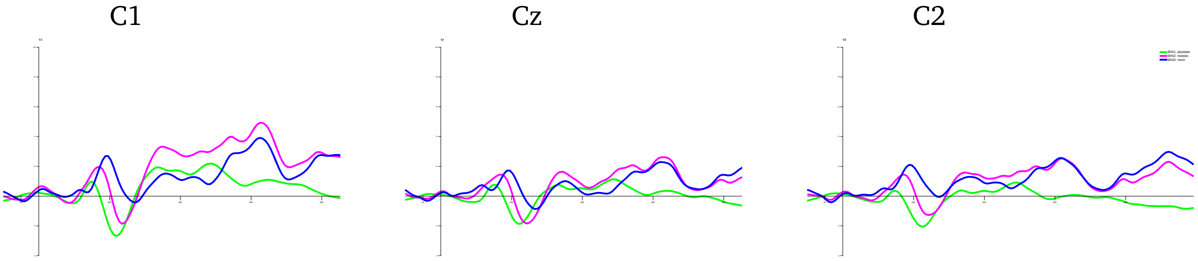
green:plus fa magenta:min fa blue:MMN



green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

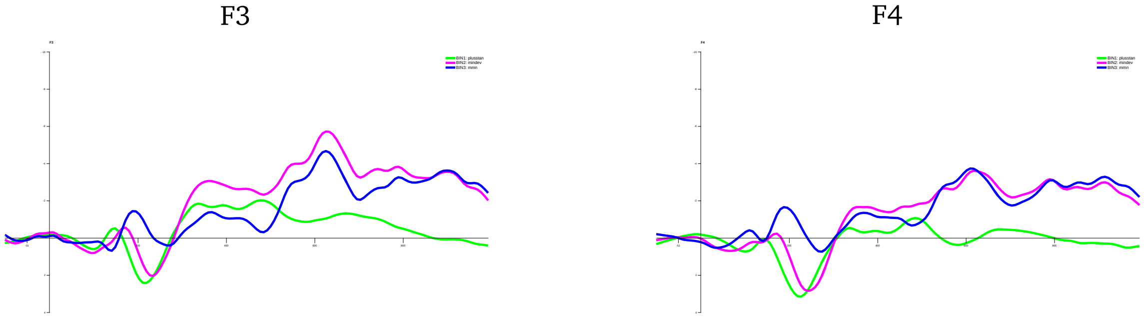
green:plus fa magenta:min fa blue:MMN



green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN



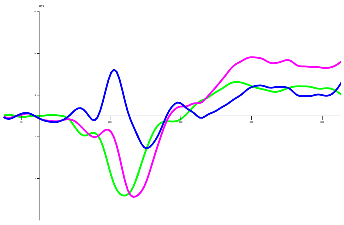
green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

Figure C.2: MMN-FA — raw ERP's for standard +fa and deviant -fa for ONLY casino.

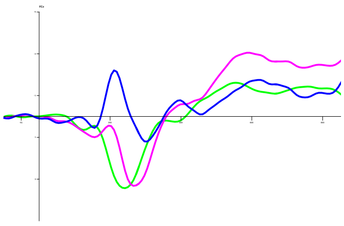
C.1.3.3 Deviant without final accent — paradis (5 subjects)

FC1



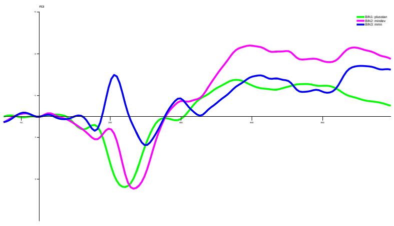
green:plus fa magenta:min fa
blue:MMN

FCz



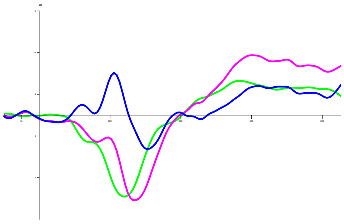
green:plus fa magenta:min fa
blue:MMN

FC2



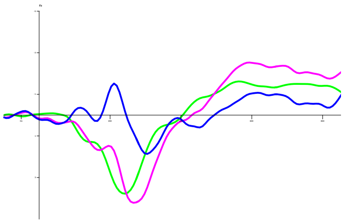
green:plus fa magenta:min fa
blue:MMN

F1



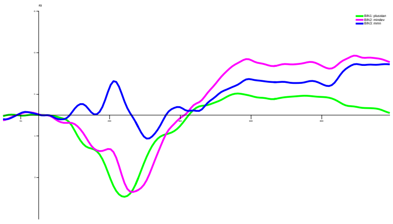
green:plus fa magenta:min fa
blue:MMN

Fz



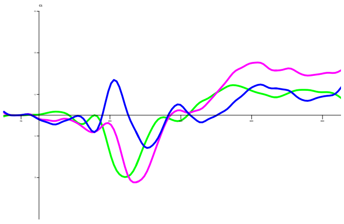
green:plus fa magenta:min fa
blue:MMN

F2



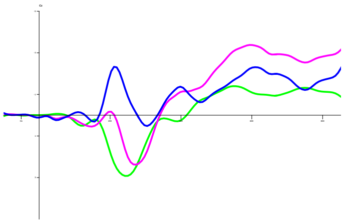
green:plus fa magenta:min fa
blue:MMN

C1



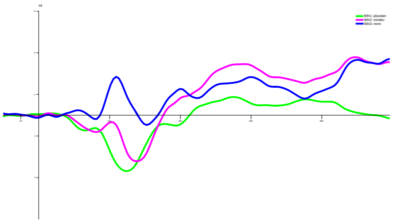
green:plus fa magenta:min fa
blue:MMN

Cz



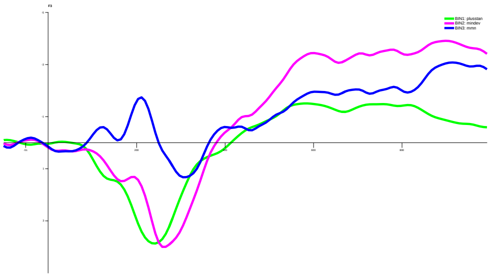
green:plus fa magenta:min fa
blue:MMN

C2



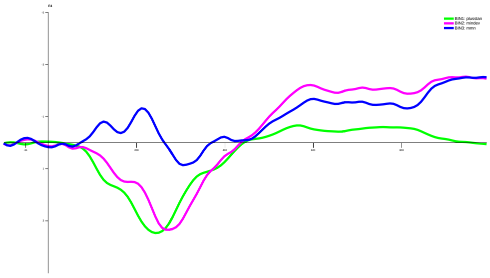
green:plus fa magenta:min fa
blue:MMN

F3



green:plus fa magenta:min fa blue:MMN

F4

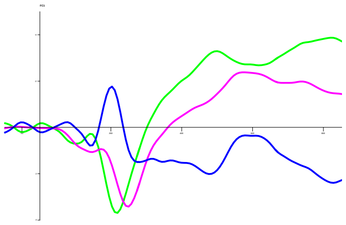


green:plus fa magenta:min fa blue:MMN

Figure C.3: MMN-FA — raw ERP's for standard +fa and deviant -fa for ONLY paradis.

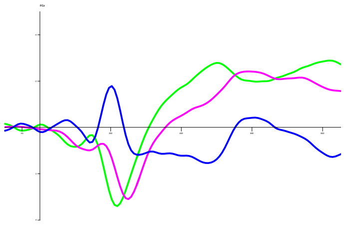
C.1.3.4 Deviant with final accent

FC1



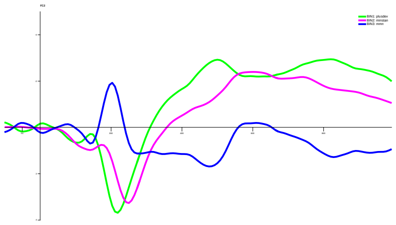
green:plus fa magenta:min fa blue:MMN

FCz



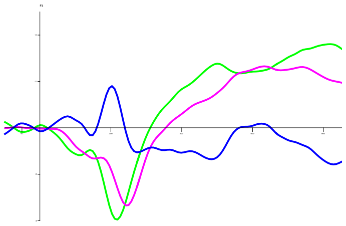
green:plus fa magenta:min fa blue:MMN

FC2



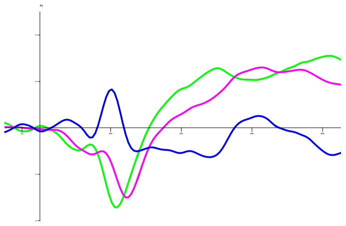
green:plus fa magenta:min fa blue:MMN

F1



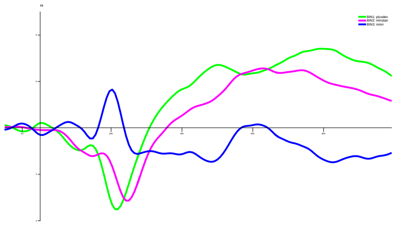
green:plus fa magenta:min fa blue:MMN

Fz



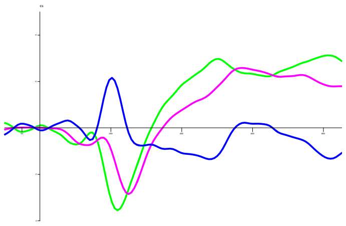
green:plus fa magenta:min fa blue:MMN

F2



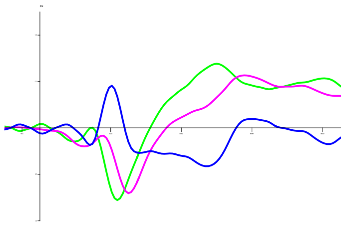
green:plus fa magenta:min fa blue:MMN

C1



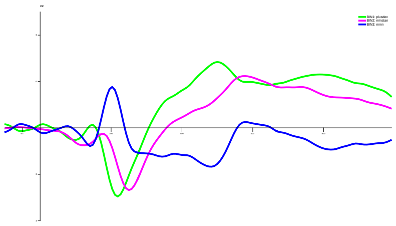
green:plus fa magenta:min fa blue:MMN

Cz



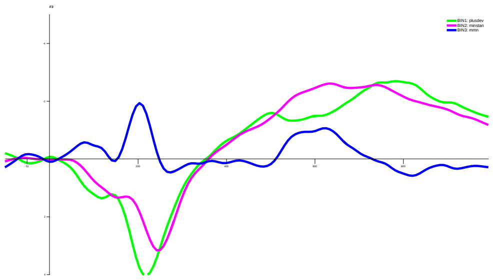
green:plus fa magenta:min fa blue:MMN

C2



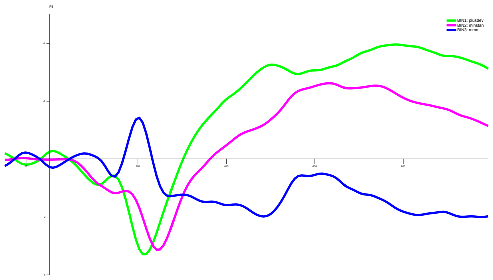
green:plus fa magenta:min fa blue:MMN

F3



green:plus fa magenta:min fa blue:MMN

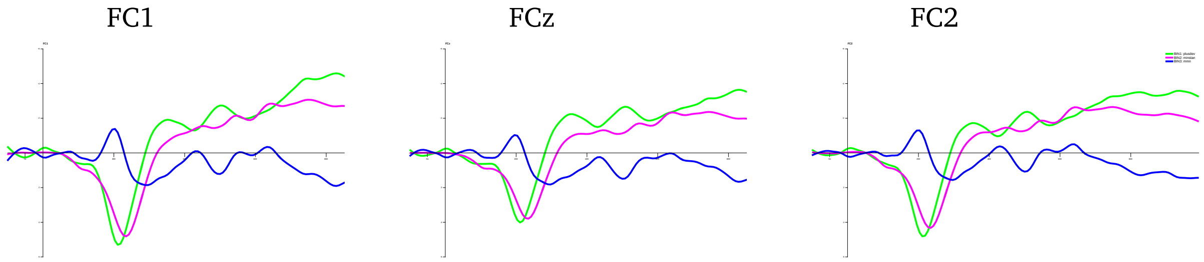
F4



green:plus fa magenta:min fa blue:MMN

Figure C.4: MMN-FA — raw ERP's for standard -fa and deviant +fa.

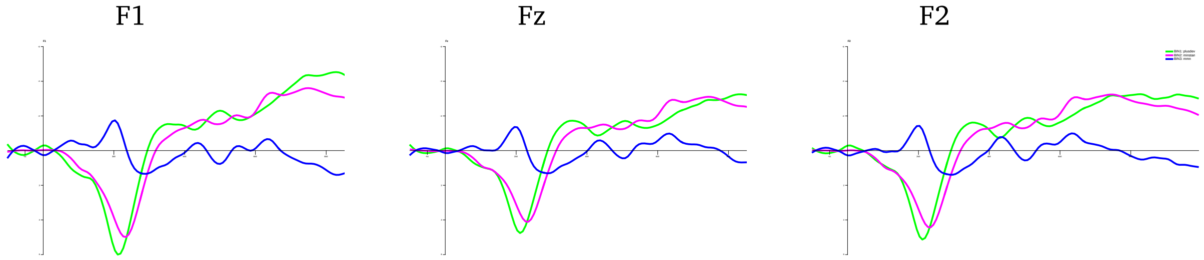
C.1.3.5 Deviant with final accent — casino (4 subjects)



green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

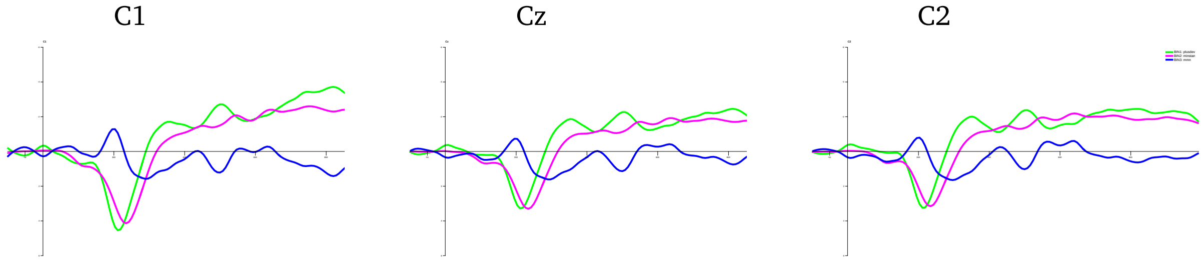
green:plus fa magenta:min fa blue:MMN



green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

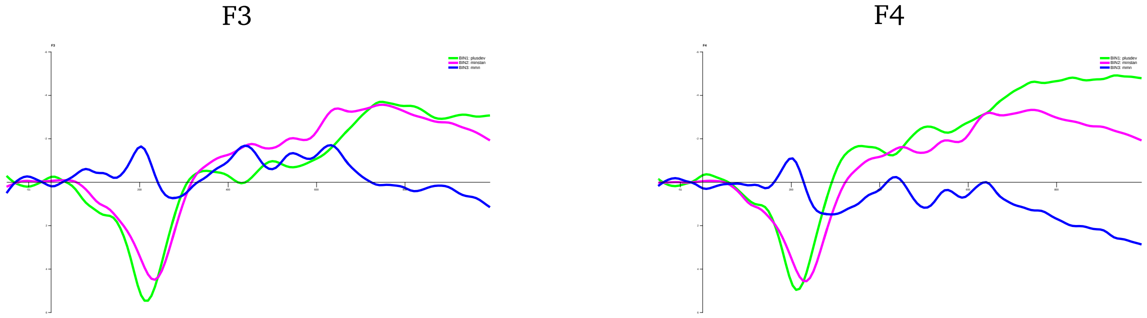
green:plus fa magenta:min fa blue:MMN



green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

green:plus fa magenta:min fa blue:MMN

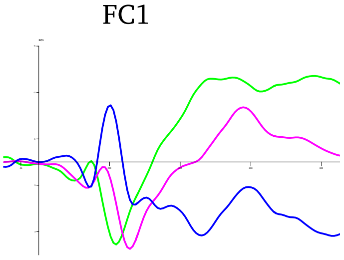


green:plus fa magenta:min fa blue:MMN

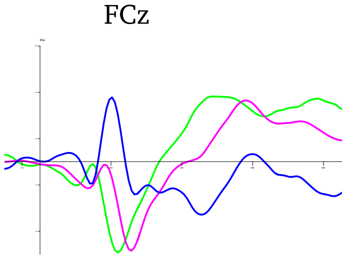
green:plus fa magenta:min fa blue:MMN

Figure C.5: MMN-FA — raw ERP's for standard -fa and deviant +fa for ONLY casino.

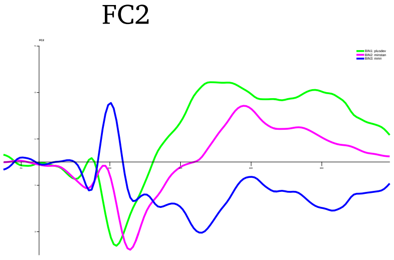
C.1.3.6 Deviant with final accent — paradis (7 subjects)


FC1

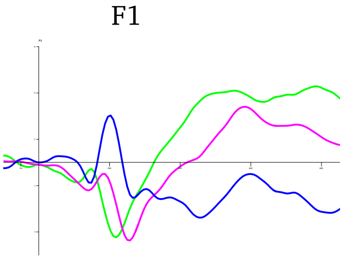
green:plus fa magenta:min fa blue:MMN


FCz

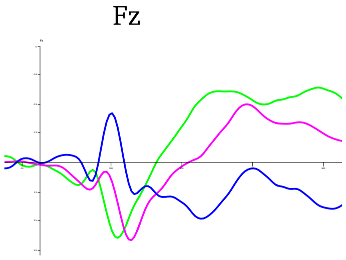
green:plus fa magenta:min fa blue:MMN


FC2

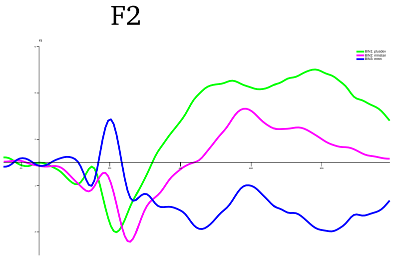
green:plus fa magenta:min fa blue:MMN


F1

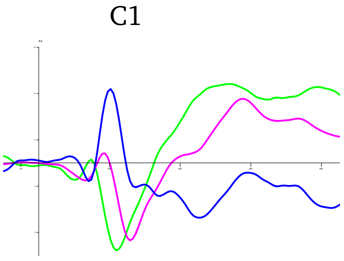
green:plus fa magenta:min fa blue:MMN


Fz

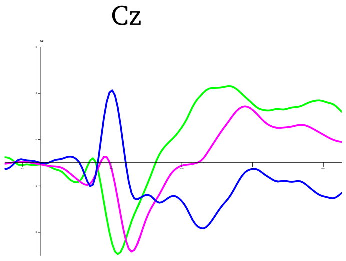
green:plus fa magenta:min fa blue:MMN


F2

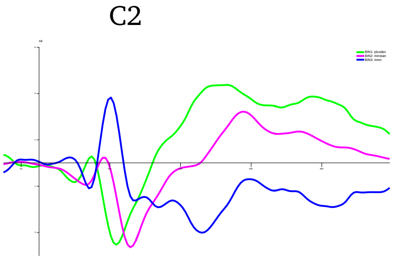
green:plus fa magenta:min fa blue:MMN


C1

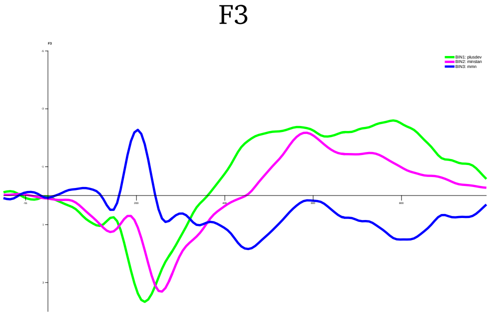
green:plus fa magenta:min fa blue:MMN


Cz

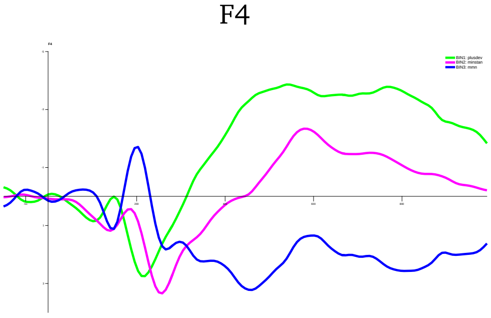
green:plus fa magenta:min fa blue:MMN


C2

green:plus fa magenta:min fa blue:MMN


F3

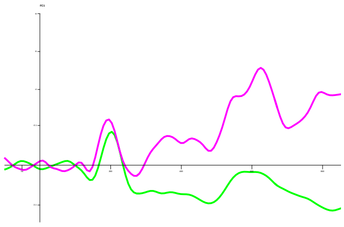
green:plus fa magenta:min fa blue:MMN


F4

green:plus fa magenta:min fa blue:MMN

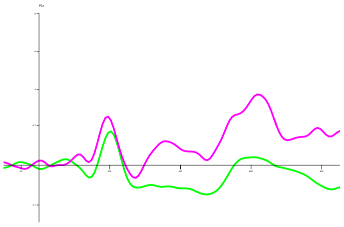
Figure C.6: MMN-FA — raw ERP's for standard -fa and deviant +fa for ONLY paradis.

C.1.3.7 Both mmn's — mmn--fa and mmn-+fa



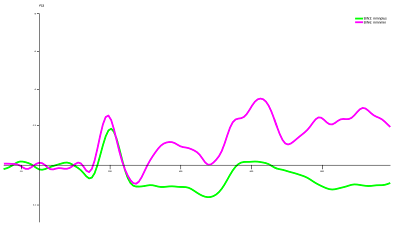
FC1

green:mmmn-plus
magenta:mmn-min



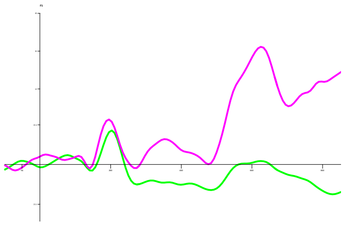
FCz

green:mmmn-plus
magenta:mmn-min



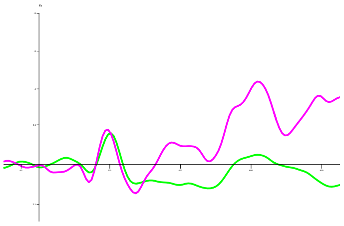
FC2

green:mmmn-plus
magenta:mmn-min



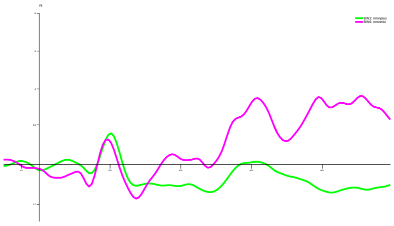
F1

green:mmmn-plus
magenta:mmn-min



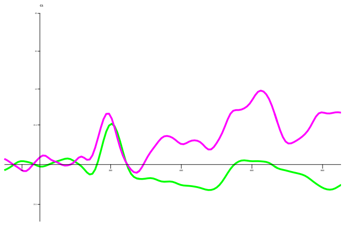
Fz

green:mmmn-plus
magenta:mmn-min



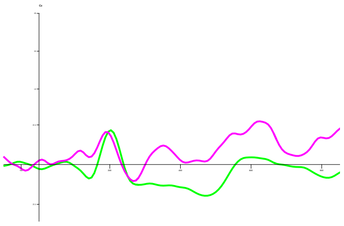
F2

green:mmmn-plus
magenta:mmn-min



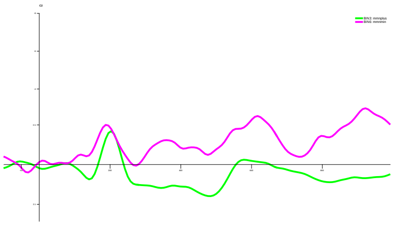
C1

green:mmmn-plus
magenta:mmn-min



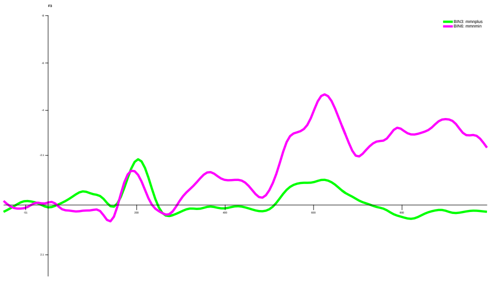
Cz

green:mmmn-plus
magenta:mmn-min



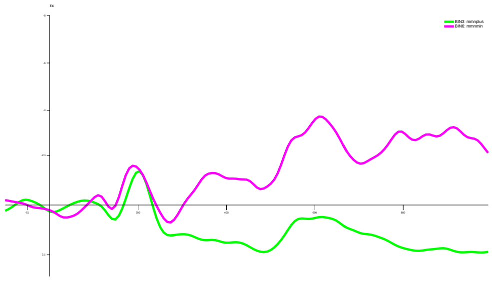
C2

green:mmmn-plus
magenta:mmn-min



F3

green:mmmn-plus magenta:mmn-min



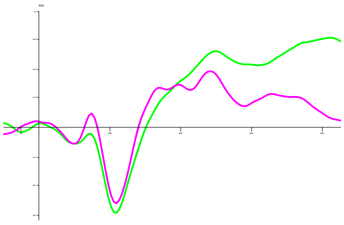
F4

green:mmmn-plus magenta:mmn-min

Figure C.7: MMN-FA — raw ERP's for both mmn- −FA and mmn- +FA.

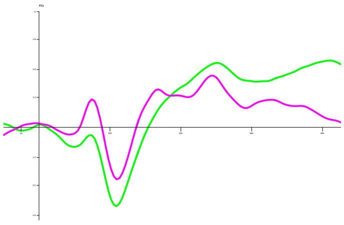
C.1.3.8 Both +fas — standard+fa and deviant+fa

FC1



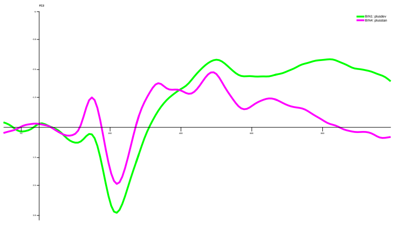
green:plus-dev
magenta:plus-stan

FCz



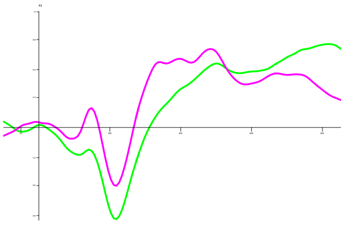
green:plus-dev
magenta:plus-stan

FC2



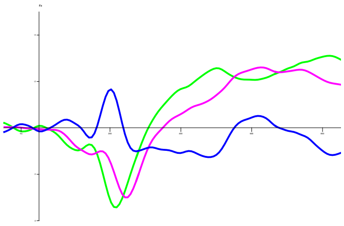
green:plus-dev
magenta:plus-stan

F1



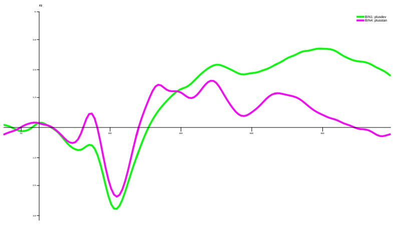
green:plus-dev
magenta:plus-stan

Fz



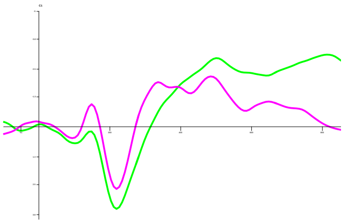
green:plus-dev
magenta:plus-stan

F2



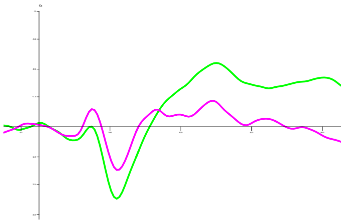
green:plus-dev
magenta:plus-stan

C1



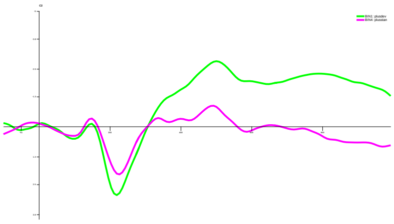
green:plus-dev
magenta:plus-stan

Cz



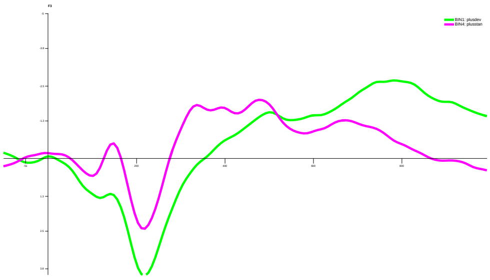
green:plus-dev
magenta:plus-stan

C2



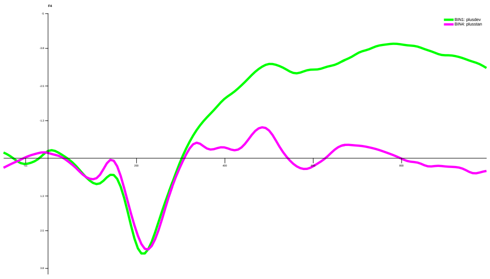
green:plus-dev
magenta:plus-stan

F3



green:plus-dev magenta:plus-stan

F4

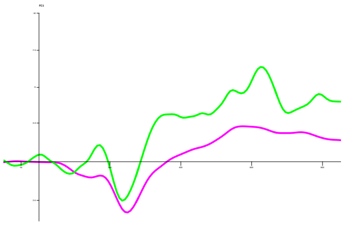


green:plus-dev magenta:plus-stan

Figure C.8: MMN-FA — raw ERP's for standard +fa and deviant +fa.

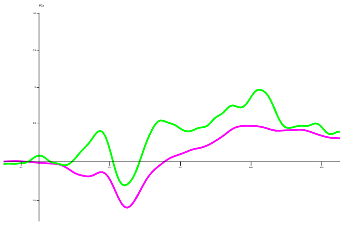
C.1.3.9 Both -fas — standard-fa and deviant-fa

FC1



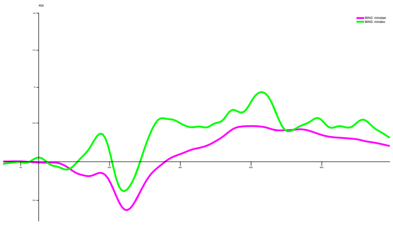
green:min-dev magenta:min-stan

FCz



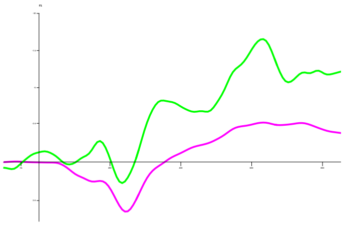
green:min-dev magenta:min-stan

FC2



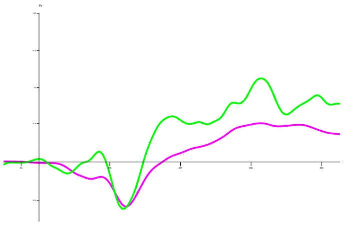
green:min-dev magenta:min-stan

F1



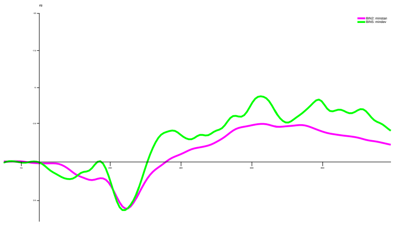
green:min-dev magenta:min-stan

Fz



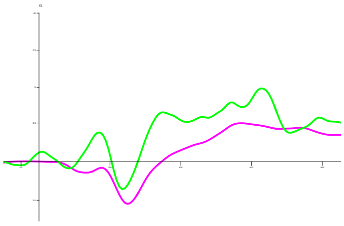
green:min-dev magenta:min-stan

F2



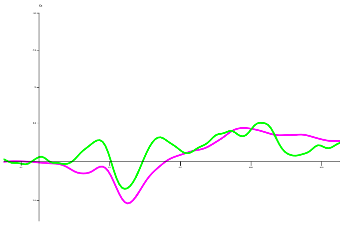
green:min-dev magenta:min-stan

C1



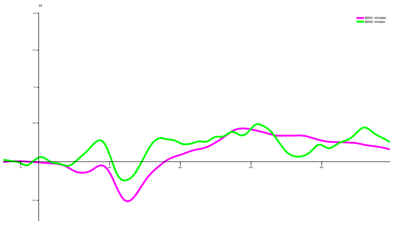
green:min-dev magenta:min-stan

Cz



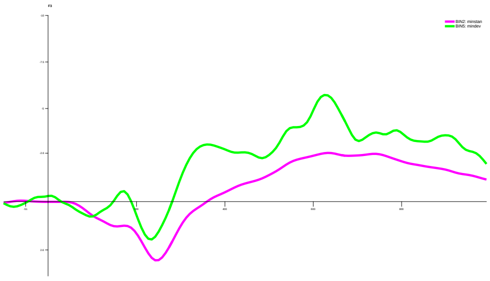
green:min-dev magenta:min-stan

C2



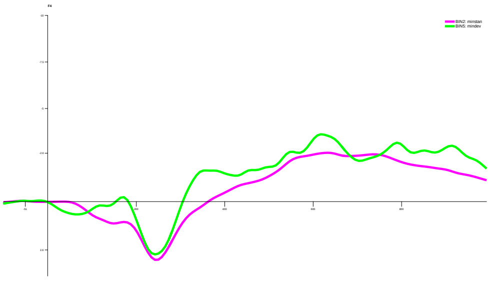
green:min-dev magenta:min-stan

F3



green:min-dev magenta:min-stan

F4



green:min-dev magenta:min-stan

Figure C.9: MMN-FA — raw ERP's for standard -fa and deviant -fa.

C.2 MMN—IA and FA mixed

C.2.1 Study overview

RQ:	Are both IA and FA expected by listeners? How does the MMN resulting from deviants —IA compare to the MMN resulting from deviants —FA?
Hypotheses:	Both deviants will result in an MMN. The MMN's will be approximately equal indicating similar expectations for both accents....

C.2.1.1 Procedure

Nr participants:	20 listeners (0 excluded)
Nr stimuli per condition:	1000 standards, 100 —IA deviants, 100 —FA deviants
Task:	No
ISI:	600 ms

C.2.1.2 Stimuli

Manipulation:	Duration for —FA and f_0 for —IA (see section 5.1.3)
Some descriptives:	

Table C.2: Durations and f_0 values of 'casino' and 'paradis': total duration, 1st syllable, 2nd syllable, 3rd syllable and the onset of the 3rd (FA) syllable.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd syllable
	ms	f_0	*ms*	f_0	*ms*	f_0	*ms*	f_0	onset
CASINO									
standard	503.3	117.8	112.9	125.8	157.1	122.2	233.3	111.0	269.0
dev—IA	503.3	116.0	112.9	120.5	157.1	121.6	233.3	109.9	269.0
dev—FA	500.8	119.0	146.9	125.1	178.1	122.3	178.8	110.7	325.0
PARADIS									
standard	459.0	106.5	121.5	121.7	111.6	118.9	225.9	93.4	233.1
dev—IA	459.0	104.1	121.5	114.4	111.6	114.4	225.9	93.4	233.1
dev—FA	456.5	110.9	168.8	123.3	139.3	119.1	148.3	95.1	308.1

C.2.1.3 EEG preprocessing

Nr electrodes: 64

Reference: Average

Filter and down-sampling: 0.01 − 30 bandpass, 128Hz

Epoch length: −200 − 1000

C.2.1.4 Analysis

Descriptive statistic:

Table C.3: MMN-mix — Descriptive statistics of peak amplitude latency variability.

Peak latencies for —IA and 'casino'											
chlabel	F1	F5h	FC1	C1	Fz	F2	F6h	FC2	FCz	Cz	C2
subjects	9	9	9	9	9	9	9	9	9	9	9
mean	243.9237	271.7013	251.7361	244.7919	256.0764	266.4931	250.0000	246.5278	253.4724	250.0002	251.7362
sd	32.86318	24.63685	27.28178	35.37275	30.45310	28.82279	30.75778	28.73451	34.52352	33.82920	34.89005
max	289.062	304.688	289.062	304.688	289.062	296.875	296.875	289.062	296.875	304.688	304.688
min	203.125	226.562	218.750	203.125	210.938	210.938	210.938	210.938	210.938	210.938	210.938
Peak latencies for —IA and 'paradis'											
chlabel	F1	F5h	FC1	C1	Fz	F2	F6h	FC2	FCz	Cz	C2
subjects	11	11	11	11	11	11	11	11	11	11	11
mean	247.1591	251.4205	238.6364	223.7216	241.4774	241.4775	253.5513	230.8238	225.8523	242.1875	234.3748
sd	34.45888	39.19025	36.37047	27.12471	33.95618	41.55422	37.20001	20.18089	18.30679	34.93872	22.09716
max	304.688	304.688	304.688	296.875	304.688	304.688	304.688	265.625	265.625	304.688	273.438
min	203.125	203.125	203.125	203.125	203.125	203.125	203.125	203.125	203.125	210.938	203.125
Peak latencies for —FA and 'casino'											
chlabel	F1	F5h	FC1	C1	Fz	F2	F6h	FC2	FCz	Cz	C2
subjects	9	9	9	9	9	9	9	9	9	9	9
mean	567.7084	596.3542	578.1250	551.2153	595.4862	568.5763	575.5208	566.8404	565.9723	562.4999	562.5002
sd	44.19411	42.43267	47.19932	44.21338	45.85095	42.39281	35.58761	45.23676	46.23768	42.79096	39.64413
max	632.812	640.625	632.812	617.188	640.625	632.812	640.625	632.812	632.812	617.188	617.188
min	515.625	515.625	515.625	484.375	515.625	515.625	523.438	515.625	515.625	507.812	515.625
Peak latencies for —FA and 'paradis'											
chlabel	F1	F5h	FC1	C1	Fz	F2	F6h	FC2	FCz	Cz	C2
subjects	11	11	11	11	11	11	11	11	11	11	11
mean	558.9487	546.8749	571.7329	562.4999	562.4998	540.4828	535.5112	560.3692	542.6135	560.3693	546.8749
sd	76.09595	73.37093	75.16402	77.57639	72.45014	80.34486	70.86266	79.14116	82.33674	80.89567	82.45815
max	648.438	640.625	648.438	648.438	640.625	640.625	640.625	648.438	648.438	648.438	648.438
min	453.125	468.750	460.938	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125

Statistical analysis:	t_{max} mass univariate permutation test, 2500 permutations in Matlab
Electrodes:	11 electrodes (Fz, Cz, FC1, FC2, C3, C4, F1, F2, F5h, F6h, FCz)
Time-windows:	$201-301$ for IA; $451-651$ for FA

C.2.1.5 Results

ia: *marginal effect of* —IA *as deviant* (critical t-score: ± 3.368, test-wise α: 0.0323, $p =$ 0.066).

fa: *effect of* —FA *as deviant, significant* (critical t-score: ± 3.4322, test-wise α: 0.002794, $p =$ between $0.0488-0.02$).

C.2.2 All non-parametric results

C.2.2.1 t_{max} permutation tests

C.2.2.1.1 Deviant initial accent

standard +IA *and* +FA; *deviant* −IA

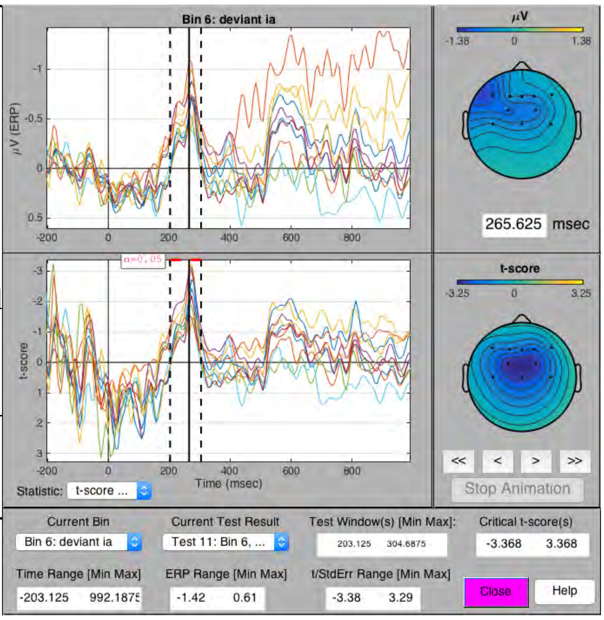
```
GND=tmaxGND(GND,6 , 'time_wind',[201  301], 'include_chans',
            {'Fz', 'Cz', 'FC1', 'FC2', 'C3', 'C4', 'F1', 'F2', 'F5h', 'F6h', 'FCz'});
```

Testing null hypothesis that the grand average ERPs in Bin 6 (deviant ia) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	11
Nr time points	14
Nr comparisons	140
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.368 and
Test-wise alpha level	0.003230
Bonferroni test-wise alpha	0.000357

ERPs are NOT significantly different from zero ($\alpha = 0$

All $p - values >= 0.066000$



Statistic: t-score ...

Current Bin: Bin 6: deviant ia

Current Test Result: Test 11: Bin 6, ...

Test Window(s) [Min Max]: 203.125 304.6875

Critical t-score(s): -3.368 3.368

Time Range [Min Max]: -203.125 992.1875

ERP Range [Min Max]: -1.42 0.61

t/StdErr Range [Min Max]: -3.38 3.29

C.2.2.1.2 Deviant final accent

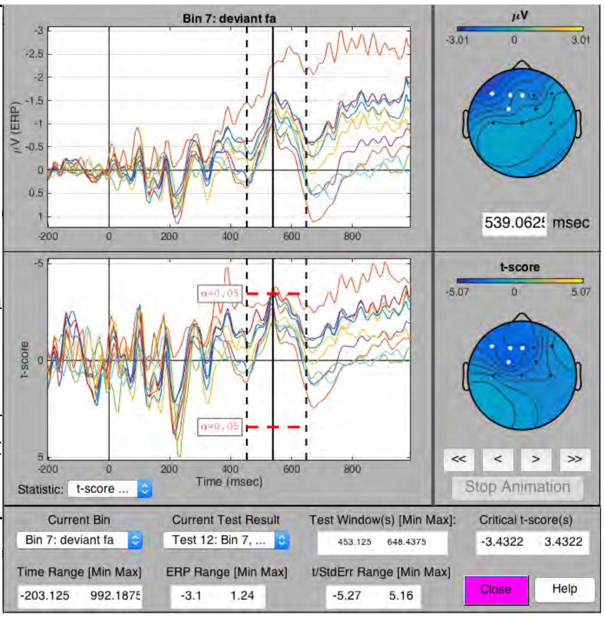
```
GND=tmaxGND(GND,7 , 'time_wind',[451  651], 'include_chans',
            {'Fz', 'Cz', 'FC1', 'FC2', 'C3', 'C4', 'F1', 'F2', 'F5h', 'F6h', 'FCz'});
```

standard +IA *and* +FA; *deviant* −FA

Testing null hypothesis that the grand average ERPs in Bin 7 (deviant fa) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	10
Nr time points	26
Nr comparisons	260
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.4322 an
Test-wise alpha level	0.002794.
Bonferroni test-wise alpha	0.000192

Significant differences from zero: $531 - 601$ ms, elect
$p - value$ are between 0.048800 and 0.020000



Statistic: t-score ...

Current Bin: Bin 7: deviant fa

Current Test Result: Test 12: Bin 7, ...

Test Window(s) [Min Max]: 453.125 648.4375

Critical t-score(s): -3.4322 3.4322

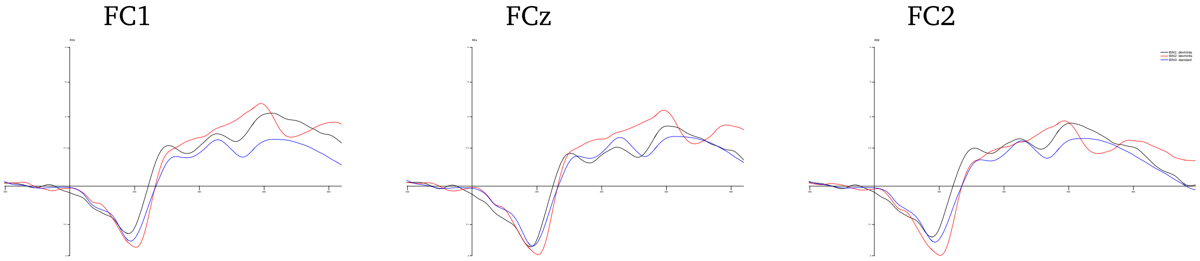
Time Range [Min Max]: -203.125 992.1875

ERP Range [Min Max]: -3.1 1.24

t/StdErr Range [Min Max]: -5.27 5.16

C.2.3 All erp plots

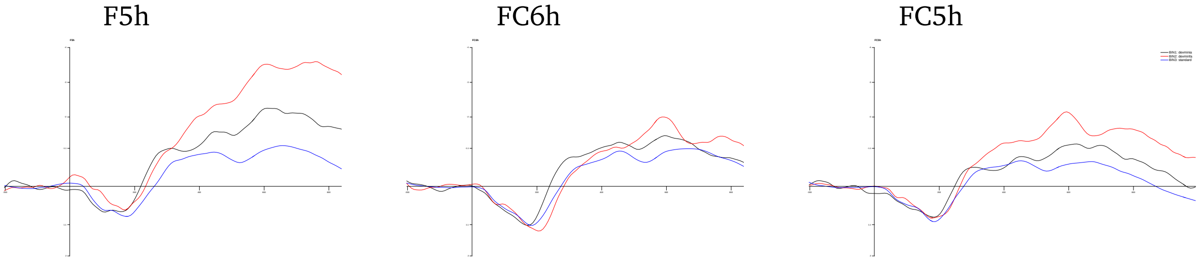
C.2.3.1 ERP's: no difference wave


FC1 FCz FC2

red:min fa black:min ia blue:standard

red:min fa black:min ia blue:standard

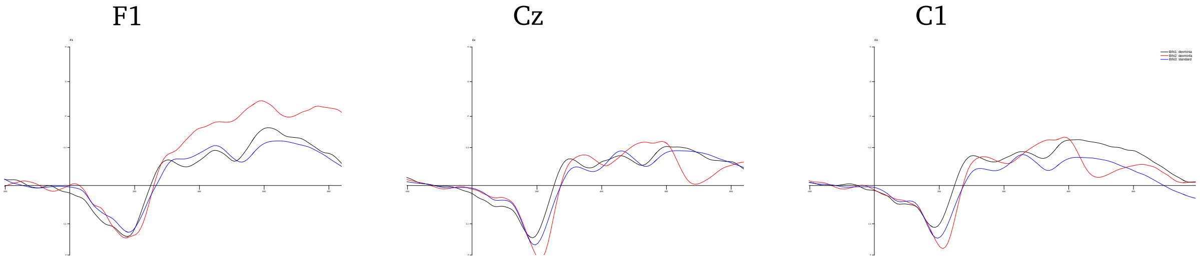
red:min fa black:min ia blue:standard


F5h FC6h FC5h

red:min fa black:min ia blue:standard

red:min fa black:min ia blue:standard

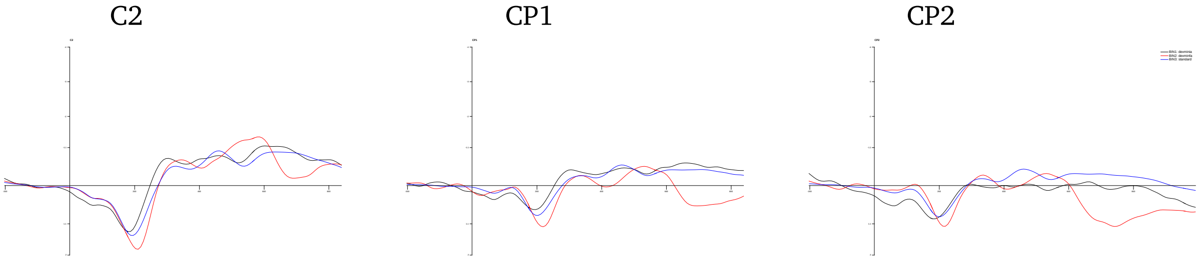
red:min fa black:min ia blue:standard


F1 Cz C1

red:min fa black:min ia blue:standard

red:min fa black:min ia blue:standard

red:min fa black:min ia blue:standard


C2 CP1 CP2

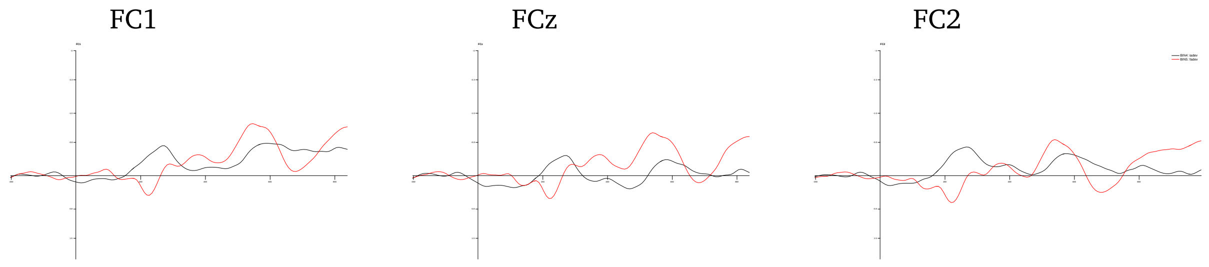
red:min fa black:min ia blue:standard

red:min fa black:min ia blue:standard

red:min fa black:min ia blue:standard

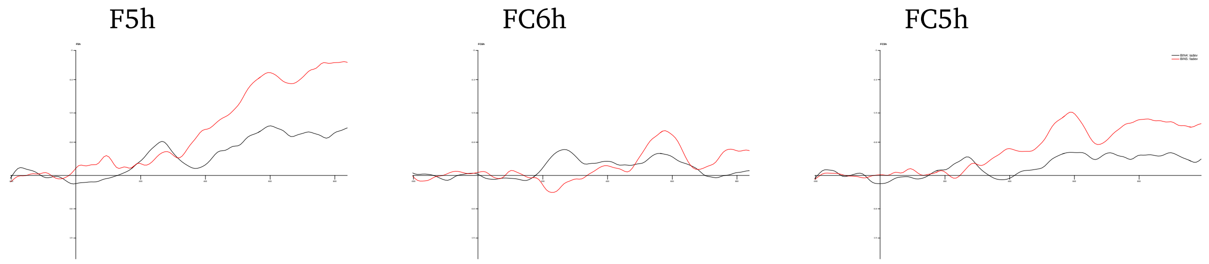
Figure C.10: MMN-mix — raw ERP's for standard and —IA deviant and —FA deviant.

C.2.3.2 ERP's: difference waves — deviant – standard



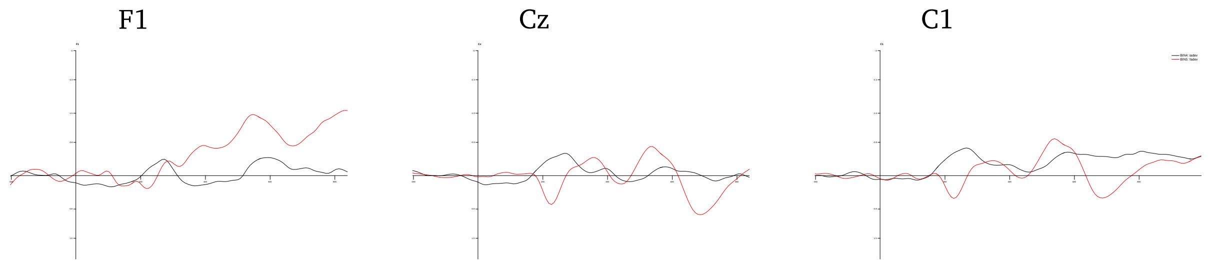
FC1 FCz FC2

red:min fa black:min ia red:min fa black:min ia red:min fa black:min ia



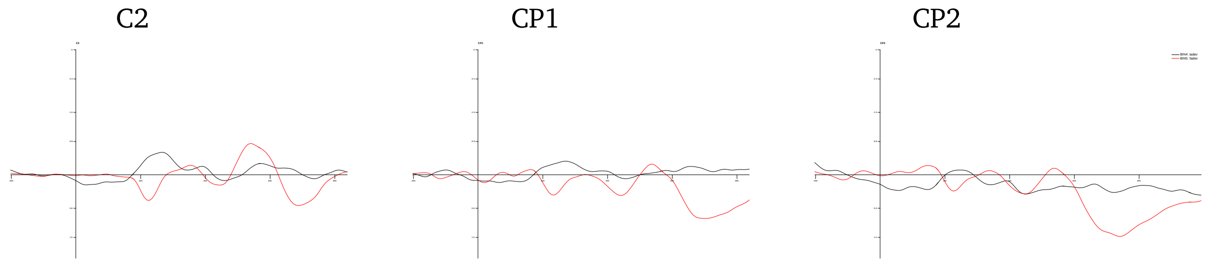
F5h FC6h FC5h

red:min fa black:min ia red:min fa black:min ia red:min fa black:min ia



F1 Cz C1

red:min fa black:min ia red:min fa black:min ia red:min fa black:min ia



C2 CP1 CP2

red:min fa black:min ia red:min fa black:min ia red:min fa black:min ia

Figure C.11: MMN-mix — ERP's difference waves —IA deviant and —FA deviant.

C.3 Lexical decision results—IA

C.3.1 Study overview

RQ: Is IA expected by listeners and does presenting stress template —IA hamper stress extraction in a way similar as in Böcker et al. (1999)?

Hypotheses: The N325 will be bigger when stimuli are presented —IA. If stress extraction is pre-lexical, there should be no interaction with lexicality.

C.3.1.1 Procedure

Nr participants: 23 listeners (3 excluded, 26 completed task)

Nr stimuli per condition: 120 (4 lists of 240 phrases, 4 conditions)

Task: Lexical decision task. Always right hand on arrow keys, random left-/right vs lexicality assignment.

ISI: ITI varied depending on subject's RT. Trials durations were fixed on 2900 ms (stimulus duration + response time + flexible ISI + fixed ISI of 400 ms)

C.3.1.2 Stimuli

Manipulation: f_0 exclusively (see section 5.1.2)

Some descriptives:

Table C.4: Durations of words and pseudowords: total duration, 1st syllable, 2nd syllable, 3rd syllable and f_0 value 1st syllable.

	Total duration		1st syllable ms		1st vowel ms		1st vowel f_0		Det vowel f_0	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
WORD										
+IA	647.7	62.5	176.6	36.1	79.0	22.4	263.1	19.2	199.4	11.7
—IA	647.7	62.5	176.2	35.8	79.0	22.4	217.6	13.4	196.7	11.8
PSEUDOWORD										
+IA	658.6	44.8	169.6	25.3	75.2	18.2	272.7	28.9	197.5	29.5
—IA	658.6	44.8	169.6	25.3	75.2	18.2	217.4	10.5	196.3	8.5

C.3.1.3 EEG preprocessing

Nr electrodes: 32

Reference: Mastoids

Filter and down-sampling: 0.01 − 30 bandpass, 125Hz

Epoch length: −200 − 2000

C.3.1.4 Analysis

Descriptive statistic behavioral:

Table C.5: Overview reaction times and first syllable duration per condition

Condition	RT *m* (ms)	RT *sd* (ms)	max RT	min RT	first syll *m* ms	first syll *sd* (ms)
word −IA	732.6273	165.4320	1774	325	174.9611	32.8876
word +IA	734.4650	160.2627	1679	306	175.0686	32.8383
pseudoword −IA	810.7314	179.1319	1793	358	172.1855	26.2040
pseudoword +IA	812.9181	182.7916	1754	436	172.2983	26.1950

Statistical analysis: Linear Mixed Effects Model in R

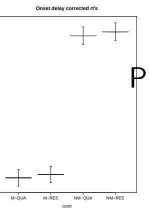
Statistical analysis: $t_{\max}$ mass univariate permutation test, 2500 permutations in Matlab

Electrodes: 8 electrodes (Fz, Cz, FC1, FC2, F3, F4, C3, C4)

Time-windows: 151 − 251 for P2; 201 − 431 for N325

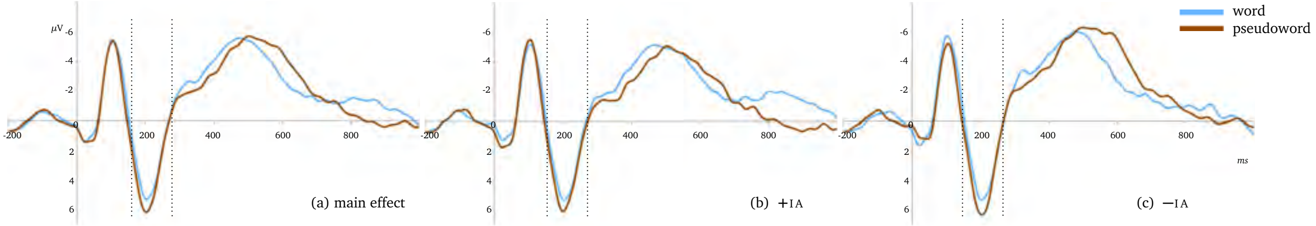
C.3.1.5 Results

Behavioral: Effect of lexicality; participants were slower to respond to pseudowords than to lexical words.



P200: *Effect of lexical congruency; main effect* (critical t-score: ± 3.5589, test-wise α: 0.001757, $p = 0.0292$).

The right frontal central electrode (FC2) is significantly more positive for pseudowords than words at 182 ms after stimulus presentation add ref to figure. Words and pseudowords are supposed to differ only in their lexicality, but it's too early for any linguistic processes to play a role. I see no differences at the N100 and for the N400 to overlap with the P2 is also a little bit improbable, since they are topographically distant and an N400 would be unexpected because stimuli were presented in isolation. The difference could theoretically be explained by an overlap between the P2 and the N325 if the presence/absence of an initial accent were to elicit an N325 only in the word condition and not in the pseudoword condition. As a matter of fact, presence/absence of initial accent had a bigger effect on the N325 in the word condition. Therefore, the lexicality effect on the P2 could have been caused by a temporal overlap between the N325 and the P2 (see also below). Böcker et al. report a similar overlap between the N325 and the P2 at the fronto-central electrodes. Finding an overlap between the N325 and the P2 implies that the process of stress extraction starts before our predefined N325 time-window (201 − 431 ms) and during the P2 time-window (151 − 251 ms). Such an early latency confirms that lexical access crucially involved an automatic, pre-lexical extraction of the French initial accent.



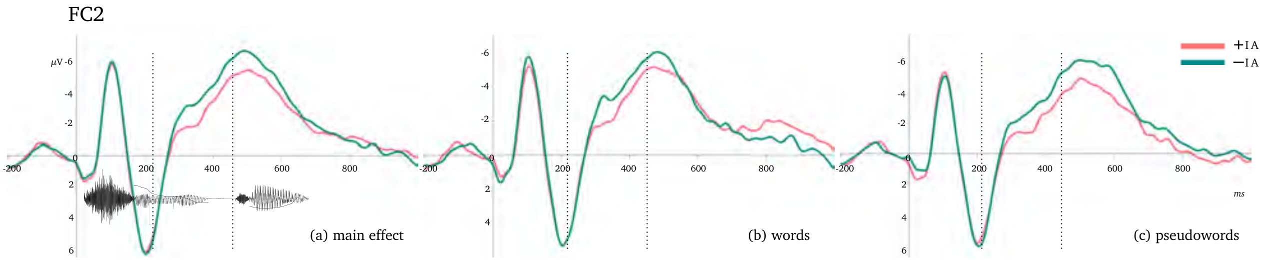
(a) main effect (b) +IA (c) −IA

Effect lexical congruency; only for without initial accent (critical t-score: ± 3.575, test-wise α: 0.000446, $p = 0.04$).
The left central electrode (C3) is significantly more positive for pseudowords than words at 206 ms after stimulus presentation for stimuli without initial accent. This again indicates a possible overlap with the N325; words without initial accent elicit an early N325, while pseudowords without initial accent do not. Note however that this electrode is different from the significant electrode in the main effect of lexicality on the P2 (which was significant at FC2).

N325: *Parametric: interaction initial accent * electrode * hemisphere* ($F(3,66) = 3.3, p < .05$). This is of course what we hoped for; apparently the presence of an initial accent affected a specific region in a specific hemisphere. The non-parametric analysis provides more details on this result.

Effect of initial accent; main effect (critical t-score: ± 3.6887, test-wise α: 0.001285, p between 0.0464 and 0.0156).

Stimuli (words AND pseudowords) without initial accent resulted in a larger negativity in frontal central electrodes (FC2 (right hemisphere) and Cz) from 318 − 358 ms after stimulus presentation (actually, Cz is only significant at the 350 ms time sample), compared to stimuli with initial accent . This indicates that processing of the stimuli was more demanding when there was no initial accent.



(a) main effect (b) words (c) pseudowords

Effect of initial accent; for lexical words (critical t-score: ± 3.8546, test-wise α: 0.000859, $p = 0.0372$).

In the lexical words condition, the right frontal central electrode (FC2) is significantly more negative for words without initial accent than for words with initial accent at 318 ms after stimulus presentation . This effect did not reach signicance in the pseudowords condition ($p = 0.0968$), but the trend in ERPs is very similar .

C.3.1.6 Conclusion

Listeners expect words irrespective of lexical congruency to be marked with the initial accent. Presenting words without IA affects pre-lexical processing as reflected by the early onset of the N325 (i.e. during the P2 time-window) and by the independence to lexical congruity.

C.3.2 Behavioral analyses

Table C.6: Overview reaction times and first syllable duration per condition

Condition	RT m (ms)	RT sd (ms)	max RT	min RT	first syll m ms	first syll sd (ms)
word −IA	732.6273	165.4320	1774	325	174.9611	32.8876
word +IA	734.4650	160.2627	1679	306	175.0686	32.8383
pseudoword −IA	810.7314	179.1319	1793	358	172.1855	26.2040
pseudoword +IA	812.9181	182.7916	1754	436	172.2983	26.1950

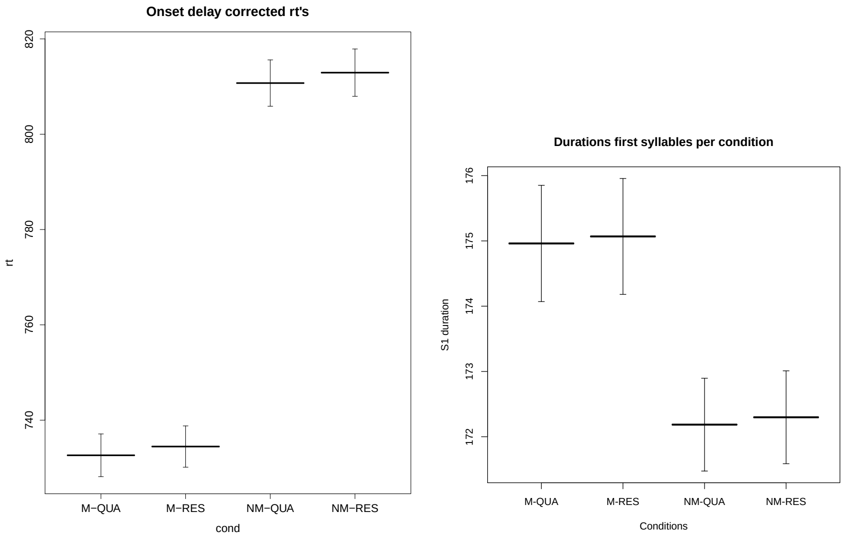


Figure C.12: Elaborate caption because there's quite a lot to say about these figures.

C.3.3 Parametric results

Separate two-tailed, repeated measures ANOVAs were conducted in *R* to test for differences in mean amplitude - at the P2 time window, the N325 time window and the N400 time window.
Time windows were calculated with the method used by Böcker et al (1999). For the P2 and the N400, a 100 ms time window was draped around the average latency peak at electrodes C3 and C4. For the N325, the time window was defined as the period between the peak latencies of the P2 and N400. This came down to a time window of 151 − 251 for the P2 (for Böcker et al. it was 171 − 271), a time

window of 381 − 481 for the N400 (same as Böcker et al.) and a time window of 201 − 431 (for Böcker et al. it was 221 − 431).
Factors were words vs pseudowords (2), +IA vs −IA (2), left vs right hemisphere (2) and electrode site (F3/F4, C3/C4, P3/P4, FC1/FC2; 4). Reported ANOVA p-values are after epsilon correction (Huynh-Feldt) for repeated measures with more than one degree of freedom.

C.3.3.1 P200

```
p200anov <- ezANOVA(data=ld_P200, dv=.(amplitude), wid=.(subject), within=.(motpseu,plus
```

ANOVA						
Effect	*DFn*	*DFd*	*F*	*p*	$p < .05$	*ges*
motpseu	1	22	8.11913342	9.325169e-03	*	3.493015e-03
plusmin	1	22	0.21697935	6.459299e-01		1.922549e-04
elect	3	66	63.15601812	2.273777e-19	*	1.952418e-01
hemisphere	1	22	0.22765108	6.379766e-01		8.981623e-05
motpseu:plusmin	1	22	0.04787033	8.288284e-01		3.867928e-05
motpseu:elect	3	66	0.14746395	9.309607e-01		2.178637e-05
plusmin:elect	3	66	1.48814399	2.258528e-01		1.778153e-04
motpseu:hemisphere	1	22	0.44744843	5.105081e-01		1.637023e-05
plusmin:hemisphere	1	22	0.38486180	5.413864e-01		2.165945e-05
elect:hemisphere	3	66	4.71109884	4.855063e-03	*	1.007000e-03
motpseu:plusmin:elect	3	66	1.55347879	2.090172e-01		2.780846e-04
motpseu:plusmin:hemisphere	1	22	0.09174516	7.648157e-01		7.533465e-06
motpseu:elect:hemisphere	3	66	1.29129910	2.847395e-01		4.689382e-05
plusmin:elect:hemisphere	3	66	1.53592822	2.134174e-01		5.534029e-05
motpseu:plusmin:elect:hemisphere	3	66	0.13201032	9.406928e-01		7.566418e-06

Mauchly’s Test for Sphericity			
Effect	*W*	*p*	*p* < .05
elect	0.07717244	3.747232e-10	*
motpseu:elect	0.37713351	1.171280e-03	*
plusmin:elect	0.52639063	2.098165e-02	*
elect:hemisphere	0.84198788	6.143249e-01	
motpseu:plusmin:elect	0.30550257	1.746129e-04	*
motpseu:elect:hemisphere	0.81296387	5.090433e-01	
plusmin:elect:hemisphere	0.65818211	1.237210e-01	
motpseu:plusmin:elect:hemisphere	0.65616178	1.209085e-01	

Sphericity Corrections						
Effect	*GGe*	*p*[*GG*]	< .05	*HFe*	*p*[*HF*]	< .0
elect	0.4385098	9.953529e-10	*	0.4553687	5.090769e-10	*
motpseu:elect	0.6007567	8.423331e-01		0.6510998	8.586712e-01	
plusmin:elect	0.7207768	2.354985e-01		0.8020700	2.329923e-01	
elect:hemisphere	0.9013746	6.626706e-03	*	1.0398569	4.855063e-03	*
motpseu:plusmin:elect	0.5669609	2.258907e-01		0.6095552	2.247283e-01	
motpseu:elect:hemisphere	0.8776168	2.857327e-01		1.0078158	2.847395e-01	
plusmin:elect:hemisphere	0.7865860	2.222964e-01		0.8871995	2.182364e-01	
motpseu:plusmin:elect:hemisphere	0.8217288	9.128791e-01		0.9333642	9.316303e-01	

C.3.3.2 N325

```
n325anov <- ezANOVA(data=ld_N325, dv=.(amplitude), wid=.(subject),
                    within=.(motpseu,plusmin,elect,hemisphere))
```

ANOVA						
Effect	*DFn*	*DFd*	*F*	*p*	$p < .05$	*ges*
motpseu	1	22	3.3540465077	0.0806167734		1.008736e-03
plusmin	1	22	1.3152016458	0.2637751634		7.320749e-04
elect	3	66	8.1077724141	0.0001126347	*	2.708524e-02
hemisphere	1	22	0.0010350879	0.9746243664		2.964679e-07
motpseu:plusmin	1	22	0.2244384112	0.6403463312		1.344262e-04
motpseu:elect	3	66	0.1334241371	0.9398153223		1.462541e-05
plusmin:elect	3	66	2.1575679258	0.1013497658		1.308371e-04
motpseu:hemisphere	1	22	1.2595682884	0.2738327053		2.576927e-05
plusmin:hemisphere	1	22	0.4249255956	0.5212420471		1.288862e-05
elect:hemisphere	3	66	2.4800131716	0.0687058796		5.380227e-04
motpseu:plusmin:elect	3	66	1.6416168889	0.1882101931		1.775035e-04
motpseu:plusmin:hemisphere	1	22	0.0003240414	0.9858001977		1.056624e-08
motpseu:elect:hemisphere	3	66	0.8385335279	0.4775776412		2.795498e-05
plusmin:elect:hemisphere	3	66	3.2921952138	0.0258800228	*	1.025990e-04
motpseu:plusmin:elect:hemisphere	3	66	0.1387274413	0.9365000897		6.918871e-06

Mauchly's Test for Sphericity			
Effect	*W*	*p*	$p < .05$
elect	0.09120626	1.887553e-09	*
motpseu:elect	0.26181876	4.222273e-05	*
plusmin:elect	0.96525428	9.811540e-01	
elect:hemisphere	0.36549778	8.851196e-04	*
motpseu:plusmin:elect	0.53053096	2.239178e-02	*
motpseu:elect:hemisphere	0.63879785	9.872854e-02	
plusmin:elect:hemisphere	0.76712002	3.594332e-01	
motpseu:plusmin:elect:hemisphere	0.76407758	3.504770e-01	

Sphericity Corrections						
Effect	*GGe*	$p[GG]$	$< .05$	*HFe*	$p[HF]$	$< .0$
elect	0.4628846	0.003912931	*	0.4841837	0.003391952	*
motpseu:elect	0.5402466	0.832838497		0.5770084	0.846767580	
plusmin:elect	0.9773600	0.102965424		1.1439430	0.101349766	
elect:hemisphere	0.6640243	0.095623590		0.7300053	0.089631157	
motpseu:plusmin:elect	0.7408088	0.201721978		0.8278031	0.197285395	
motpseu:elect:hemisphere	0.7996909	0.456532152		0.9043564	0.468149198	
plusmin:elect:hemisphere	0.8437828	0.034022395	*	0.9625904	0.027625909	*
motpseu:plusmin:elect:hemisphere	0.8574407	0.914540486		0.9807898	0.933937865	

C.3.3.3 N400

```
n400anov <- ezANOVA(data=ld_N400, dv=.(amplitude), wid=.(subject),
                    within=.(motpseu, plusmin, elect, hemisphere))
```

ANOVA						
Effect	*DFn*	*DFd*	*F*	p	$p < .05$	*ges*
motpseu	1	22	3.47786956	0.07559017		1.579431e-03
plusmin	1	22	3.70519261	0.06727042		1.641860e-03
elect	3	66	3.83291699	0.01359582	*	1.170883e-02
hemisphere	1	22	0.03718982	0.84884829		1.396858e-05
motpseu:plusmin	1	22	1.17576849	0.28995964		5.902383e-04
motpseu:elect	3	66	2.36770405	0.07867053		2.168498e-04
plusmin:elect	3	66	2.09249964	0.10961156		1.159821e-04
motpseu:hemisphere	1	22	0.34903132	0.56068301		9.838439e-06
plusmin:hemisphere	1	22	0.03010553	0.86383757		7.246786e-07
elect:hemisphere	3	66	3.42054312	0.02220025	*	7.952286e-04
motpseu:plusmin:elect	3	66	1.56615594	0.20589339		1.264122e-04
motpseu:plusmin:hemisphere	1	22	0.02149694	0.88476853		4.471090e-07
motpseu:elect:hemisphere	3	66	1.17359386	0.32654079		4.335431e-05
plusmin:elect:hemisphere	3	66	3.66388887	0.01661451	*	1.370810e-04
motpseu:plusmin:elect:hemisphere	3	66	0.73480970	0.53496949		2.899678e-05

Mauchly's Test for Sphericity *Effect*	W	p	$p < .05$
elect	0.1139963	1.618845e-08	*
motpseu:elect	0.4384903	4.423375e-03	*
plusmin:elect	0.9462651	9.501766e-01	
elect:hemisphere	0.5103955	1.621467e-02	*
motpseu:plusmin:elect	0.6464586	1.080833e-01	
motpseu:elect:hemisphere	0.6462954	1.078770e-01	
plusmin:elect:hemisphere	0.8718238	7.247190e-01	
motpseu:plusmin:elect:hemisphere	0.9317118	9.171405e-01	

Sphericity Corrections *Effect*	GGe	$p[GG]$	$< .05$	HFe	$p[HF]$	$< .05$
elect	0.5139769	0.04162540	*	0.5452520	0.03870930	*
motpseu:elect	0.6368216	0.10830665		0.6958961	0.10286850	
plusmin:elect	0.9631133	0.11224229		1.1242384	0.10961156	
elect:hemisphere	0.7228692	0.03740323	*	0.8047506	0.03203433	*
motpseu:plusmin:elect	0.8191315	0.21409187		0.9299352	0.20912738	
motpseu:elect:hemisphere	0.8143442	0.32342965		0.9236222	0.32554176	
plusmin:elect:hemisphere	0.9267195	0.01939210	*	1.0743003	0.01661451	*
motpseu:plusmin:elect:hemisphere	0.9552365	0.52934095		1.1133819	0.53496949	

C.3.4 Non-parametric results

Separate repeated measures, two-tailed permutation tests based on the t-max statistic were conducted in Matlab for each of the time windows described above and each of the comparisons of interest, i.e.:

Time-windows:

- P2 → $151 - 251$ ms
- N325 → $201 - 431$ ms
- N400 → $381 - 481$ ms

Comparisons:

- main effect ±IA
- main effect lexicality
- ±IA $*$ lexicality (4 comparisons)

13 electrodes were selected for analysis: Fz, Cz, FC1, FC2, F3, F4, C3, C4, Pz, P3, P4, CP1, CP2. The data was down-sampled to 125 Hz to reduce the number of comparisons. Also, 2500 permutations were used to estimate the distribution of the null hypothesis for the customary family-wise α level of 0.05. All significant effects found with the parametric tests were replicated in the non-parametric test. The non-parametric tests did prove to be a little bit more powerful, as they revealed some additional effects.

C.3.4.1 P200

Effect initial accent; main effect

```
GND=tmaxGND(GND,9,'time_wind',[151 251],'include_chans',
{'Fz','Cz','FC1','FC2','F3','F4','C3','C4'});
```

Testing null hypothesis that the grand average ERPs in Bin 9 (minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	14
Nr comparisons	112
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.4959 and 3.4959
Test-wise alpha level	0.002044
Bonferroni test-wise alpha	0.000446

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed

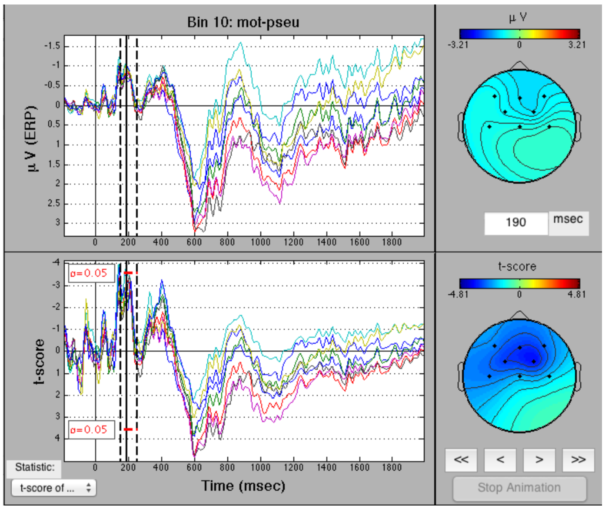
All $p - values >= 0.766000$

Effect lexical congruency; main effect Testing null hypothesis that the grand average ERPs in Bin 10 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	14
Nr comparisons	112
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.5589 and 3.3.5589
Test-wise alpha level	0.001757.
Bonferroni test-wise alpha	0.000446

Significant differences from zero: 182 ms, electrode(s): FC2

$p - value = 0.029200$



Effect initial accent; condition → lexical words Testing null hypothesis that the grand average ERPs in Bin 11 (mot: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	14
Nr comparisons	112
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6968 and 3.6968
Test-wise alpha level	0.001260
Bonferroni test-wise alpha	0.000446

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.852800$

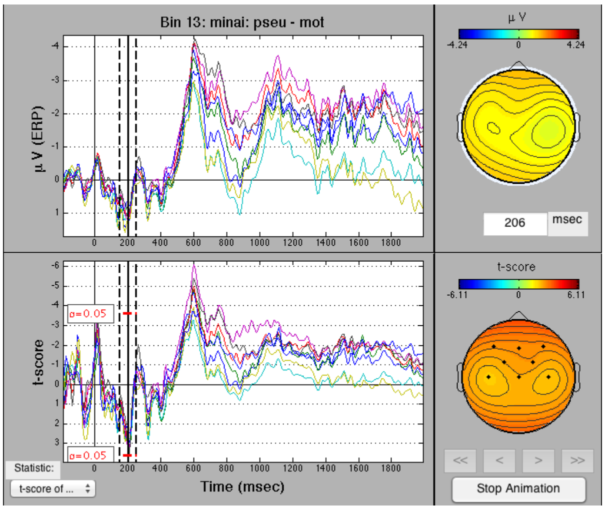
Effect initial accent; condition → pseudowords Testing null hypothesis that the grand average ERPs in Bin 12 (pseu: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	14
Nr comparisons	112
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.3521 and 3.3521
Test-wise alpha level	0.002883
Bonferroni test-wise alpha	0.000446

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.429600$

Effect lexical congruency; condition → min initial accent PLAATJE Testing null hypothesis that the grand average ERPs in Bin 13 (minai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	14
Nr comparisons	112
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.575 and 3.575
Test-wise alpha level	0.001690
Bonferroni test-wise alpha	0.000446
Significant differences from zero:206 ms, electrode(s): C3 $p - value = 0.040000$	



Effect lexical congruency; condition → with initial accent Testing null hypothesis that the grand average ERPs in Bin 14 (plusai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	14
Nr comparisons	112
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.4365 and 3.4365
Test-wise alpha level	0.002356
Bonferroni test-wise alpha	0.000446
ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed All $p - values >= 0.407600$	

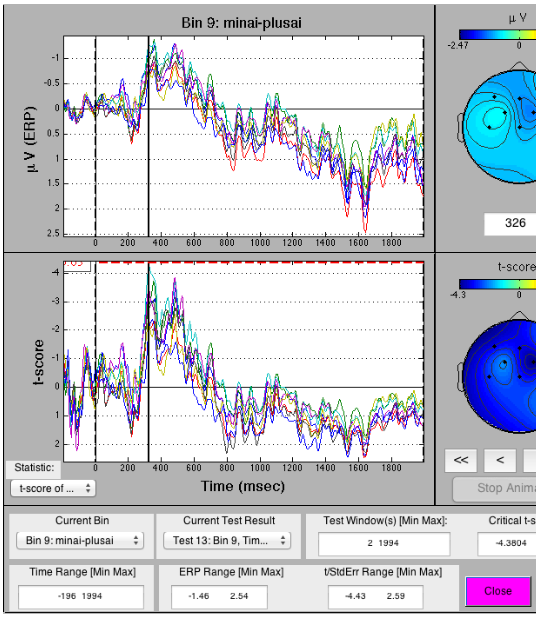
C.3.4.2 N325

Effect initial accent; main effect Testing null hypothesis that the grand average ERPs in Bin 9 (minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	27
Nr comparisons	216
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6887 and 3.6887
Test-wise alpha level	0.001285
Bonferroni test-wise alpha	0.000231

Significant differences from zero: 318 ms, electrode(s): FC2; 326 ms, electrode(s): FC2; 334 ms, electrode(s): FC2; 342 ms, electrode(s): FC2.350 ms, electrode(s): Cz, FC2; 358 ms, electrode(s): FC2

$p - values$ between 0.046400 and 0.015600



Current Bin: Bin 9: minai-plusai

Current Test Result: Test 13: Bin 9, Tim...

Test Window(s) [Min Max]: 2 1994

Critical t-s -4.3804

Time Range [Min Max]: -196 1994

ERP Range [Min Max]: -1.46 2.54

t/StdErr Range [Min Max]: -4.43 2.59

Effect lexical congruency; main effect Testing null hypothesis that the grand average ERPs in Bin 10 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

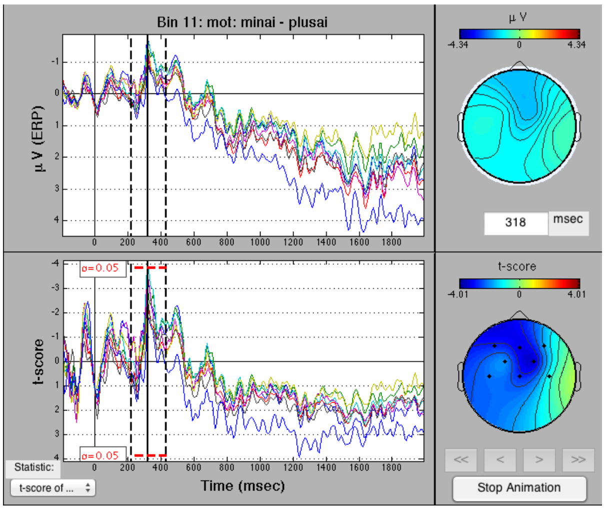
Nr channels	8
Nr time points	27
Nr comparisons	216
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8435 and 3.8435
Test-wise alpha level	0.000883
Bonferroni test-wise alpha	0.000231

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.139600$

Effect initial accent; condition → lexical words Testing null hypothesis that the grand average ERPs in Bin 11 (mot: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	27
Nr comparisons	216
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8546 and 3.8546
Test-wise alpha level	0.000859
Bonferroni test-wise alpha	0.000231

Significant differences from zero: 318 ms, electrode(s): FC2
$p - value = 0.037200$



Effect initial accent; condition → pseudowords Testing null hypothesis that the grand average ERPs in Bin 12 (pseu: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	27
Nr comparisons	216
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.5866 and 3.5866
Test-wise alpha level	0.001644
Bonferroni test-wise alpha	0.000231

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p-values >= 0.096800$

Effect lexical congruency; condition → min initial accent Testing null hypothesis that the grand average ERPs in Bin 13 (minai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	27
Nr comparisons	216
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0048 and 4.0048
Test-wise alpha level	0.000596
Bonferroni test-wise alpha	0.000231

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p-values >= 0.356800$

Effect lexical congruency; condition → with initial accent Testing null hypothesis that the grand average ERPs in Bin 14 (plusai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	8
Nr time points	27
Nr comparisons	216
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.5819 and 3.5819
Test-wise alpha level	0.001662
Bonferroni test-wise alpha	0.000231

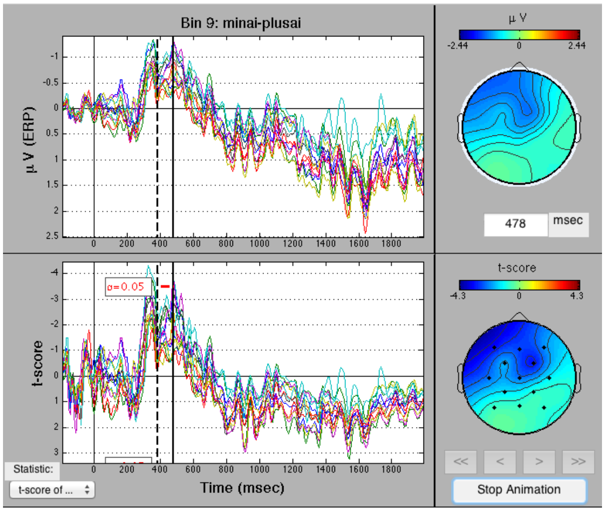
ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.184400$

C.3.4.3 N400

Effect initial accent; main effect Testing null hypothesis that the grand average ERPs in Bin 9 (minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

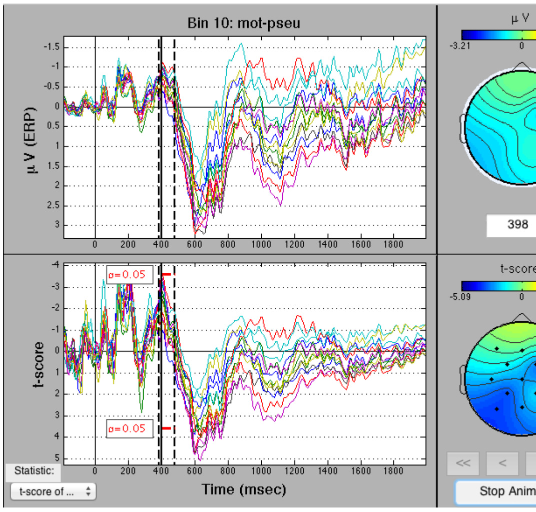
Nr channels	13
Nr time points	13
Nr comparisons	169
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.4964 and 3.4964
Test-wise alpha level	0.002041
Bonferroni test-wise alpha	0.000296

Significant differences from zero: 478 ms, electrode(s): FC2, F3
$p - values$ between 0.036000 and 0.028000



Effect lexical congruency; main effect Testing null hypothesis that the grand average ERPs in Bin 10 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	13
Nr time points	13
Nr comparisons	169
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.5991 and 3.5991
Test-wise alpha level	0.001595
Bonferroni test-wise alpha	0.000296
Significant differences from zero: 398 ms, electrode(s): P3 $p-value=0.047600$	



Effect initial accent; condition → lexical words Testing null hypothesis that the grand average ERPs in Bin 11 (mot: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	13
Nr time points	13
Nr comparisons	169
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.7039 and 3.7039
Test-wise alpha level	0.001238
Bonferroni test-wise alpha	0.000296
ERPs are NOT significantly different from zero ($\alpha=0.05$) at any time point or window analyzed All $p-values>=0.672800$	

Effect initial accent; condition → pseudowords Testing null hypothesis that the grand average ERPs in Bin 12 (pseu: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	13
Nr time points	13
Nr comparisons	169
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.3864 and 3.3864
Test-wise alpha level	0.002656.
Bonferroni test-wise alpha	0.000296

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.072400$

Effect lexical congruency; condition → min initial accent Testing null hypothesis that the grand average ERPs in Bin 13 (minai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	13
Nr time points	13
Nr comparisons	169
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6818 and 3.6818
Test-wise alpha level	0.001306
Bonferroni test-wise alpha	0.000296

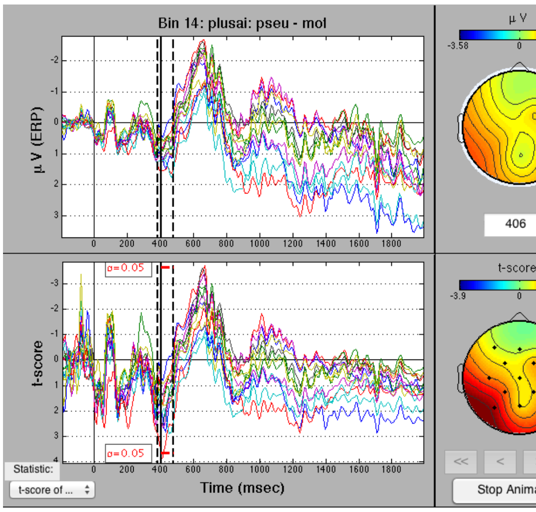
ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.640400$

Effect lexical congruency; condition → with initial accent Testing null hypothesis that the grand average ERPs in Bin 14 (plusai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	13
Nr time points	13
Nr comparisons	169
t-score degrees of freedom	22
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6572 and 3.6572
Test-wise alpha level	0.001386
Bonferroni test-wise alpha	0.000296

Significant differences from zero: 406 ms, electrode(s): P3; 414 ms, electrode(s): P3

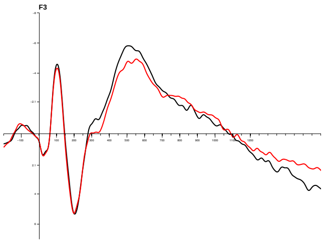
$p - values$ between 0.037600 and 0.034000



C.3.5 All erp plots

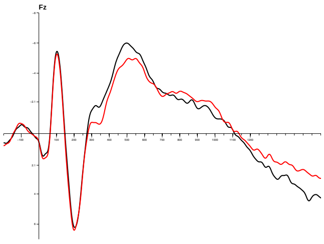
C.3.5.1 Effect initial accent; main effect

F3



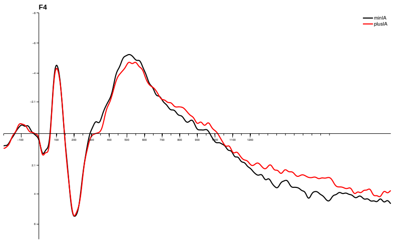
red:plus ia black:min ia

FZ



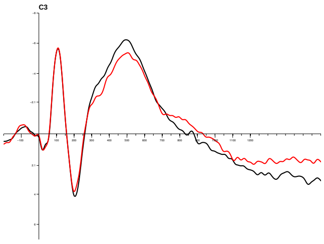
red:plus ia black:min ia

F4



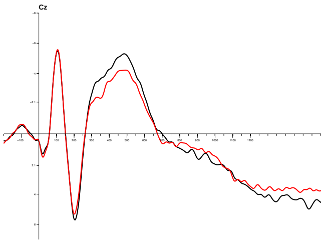
red:plus ia black:min ia

C3



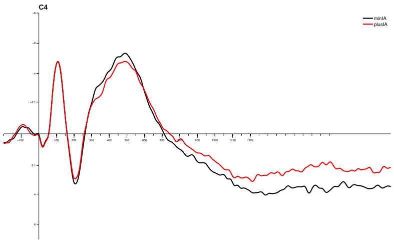
red:plus ia black:min ia

CZ



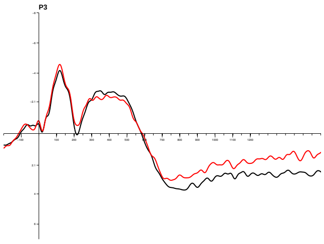
red:plus ia black:min ia

C4



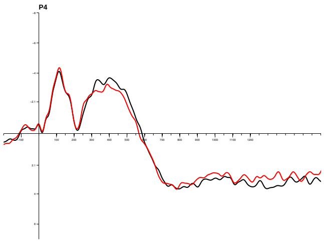
red:plus ia black:min ia

P3



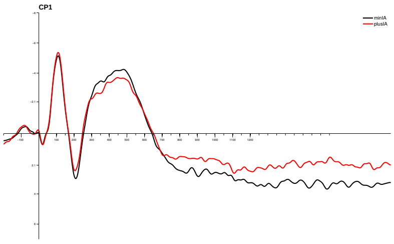
red:plus ia black:min ia

P4



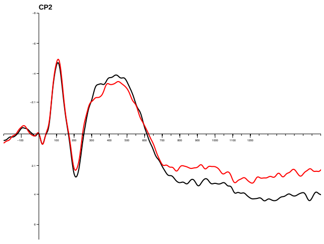
red:plus ia black:min ia

CP1



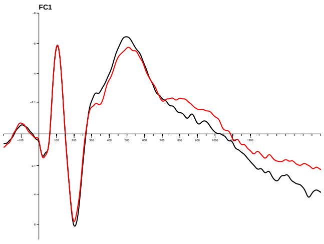
red:plus ia black:min ia

CP2



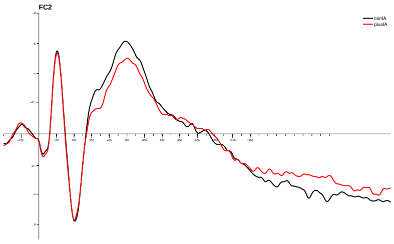
red:plus ia black:min ia

FC1



red:plus ia black:min ia

FC2

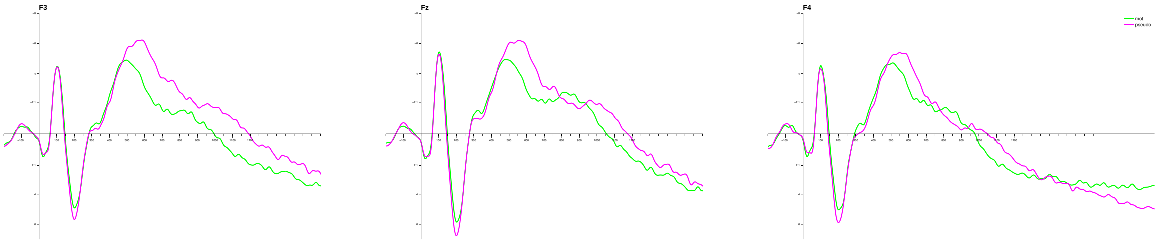


red:plus ia black:min ia

Figure C.13: LD-IA — ERP's main effect initial accent: so all plus initial accent word versus all min initial accent stimuli.

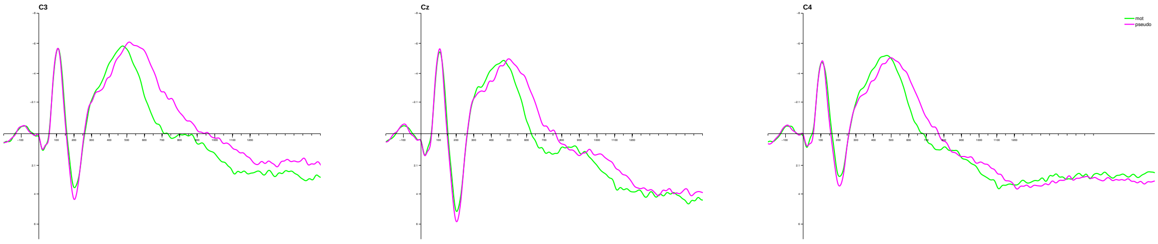
C.3.5.2 Effect lexical congruency; main effect

F3 FZ F4



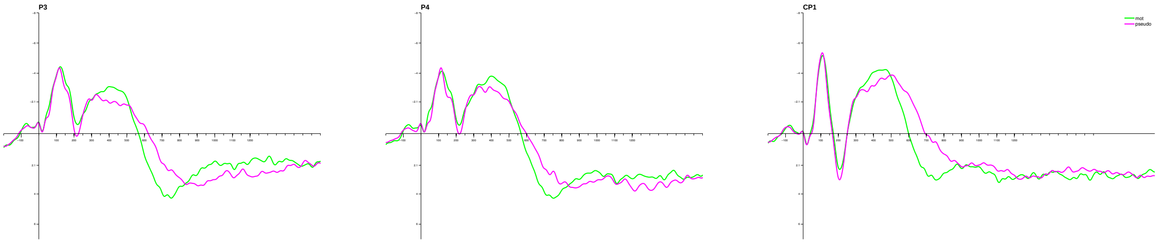
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

C3 CZ C4



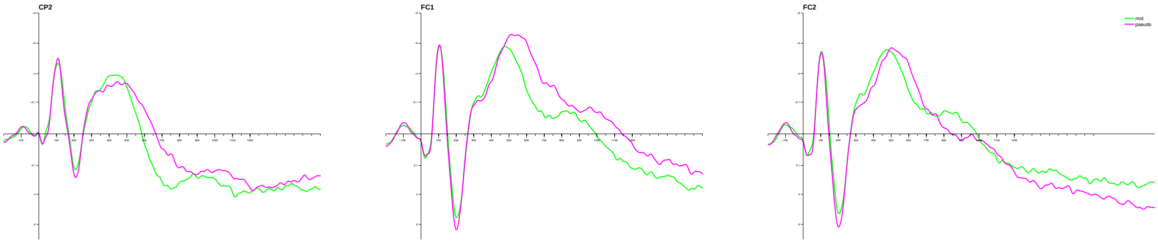
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

P3 P4 CP1



green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

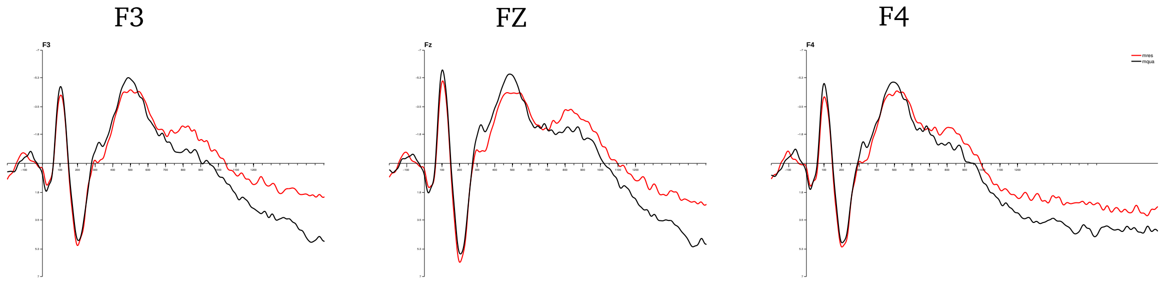
CP2 FC1 FC2



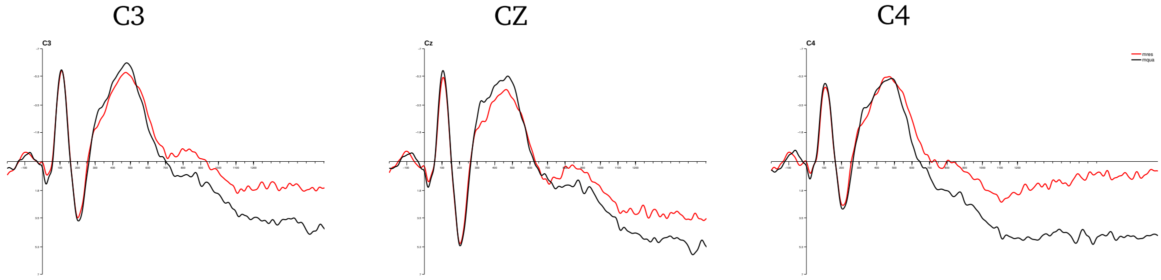
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

Figure C.14: LD-IA — ERP's main effect lexical congruency: so all congruent words versus all pseudowords.

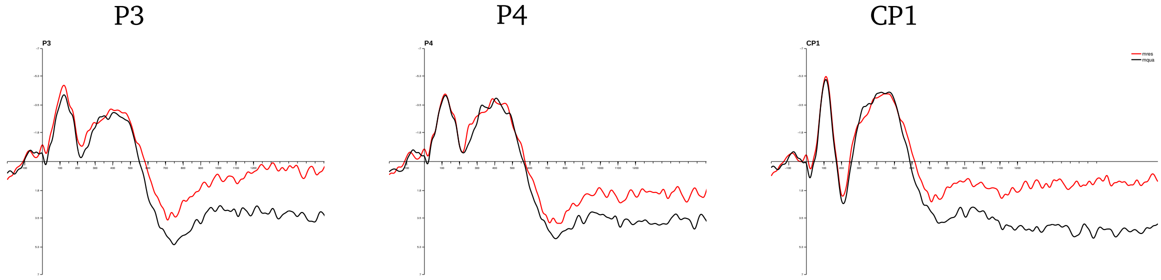
C.3.5.3 Effect initial accent; condition → lexical words



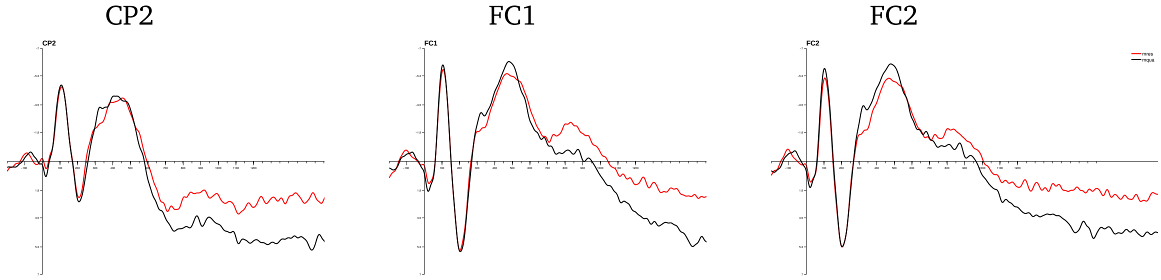
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

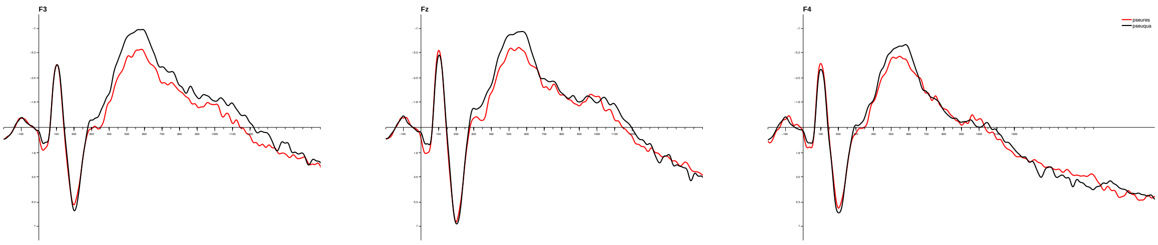


red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

Figure C.15: LD-IA — ERP's effect initial accent but ONLY for lexical words.

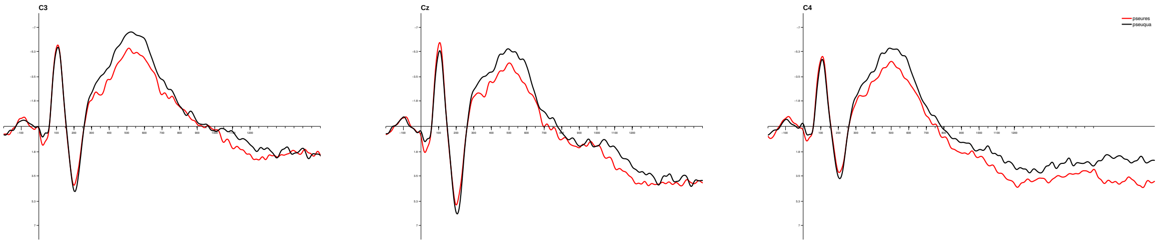
C.3.5.4 Effect initial accent; condition → pseudowords

F3 FZ F4



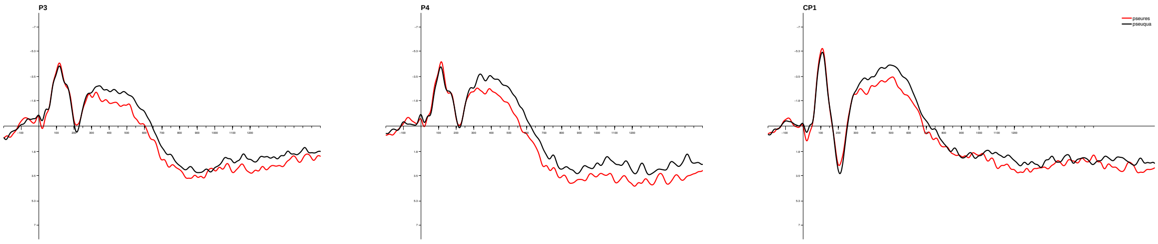
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

C3 CZ C4



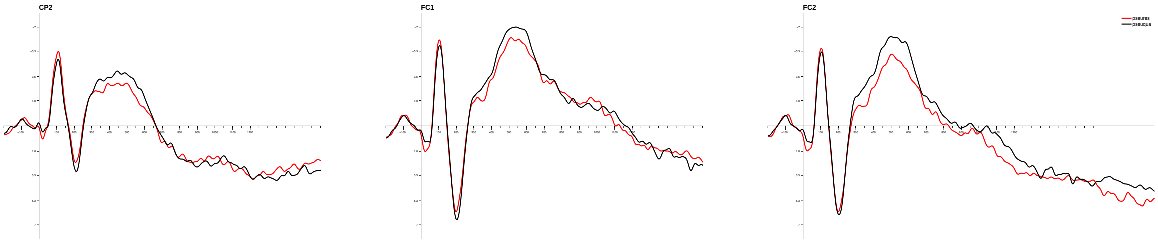
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

P3 P4 CP1



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

CP2 FC1 FC2

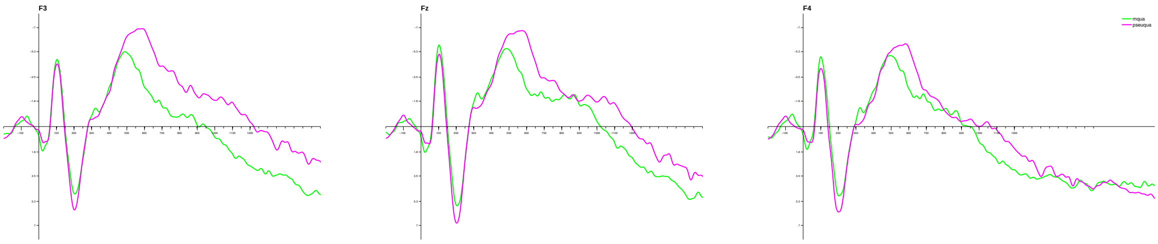


red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

Figure C.16: LD-IA — ERP's effect initial accent but ONLY for pseudowords.

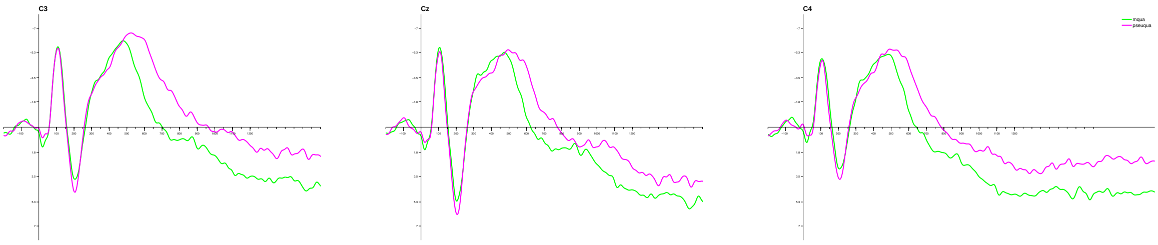
C.3.5.5 Effect lexical congruency; condition → min initial accent

F3 FZ F4



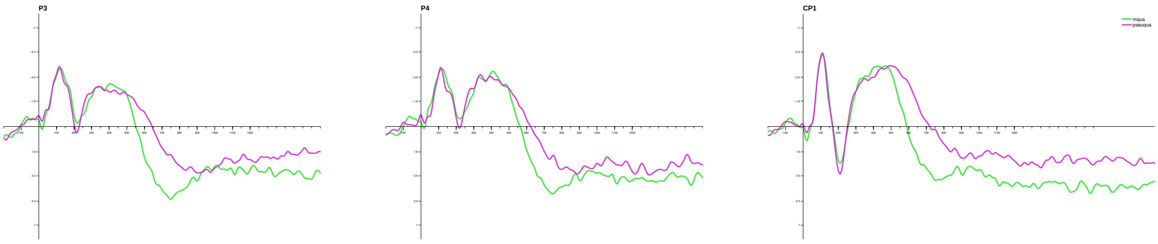
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

C3 CZ C4



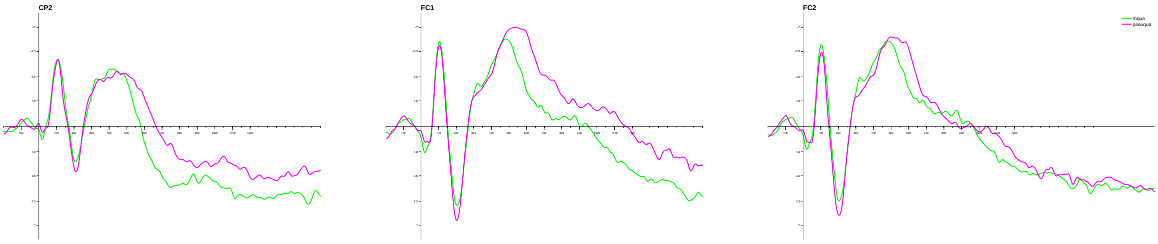
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

P3 P4 CP1



green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

CP2 FC1 FC2

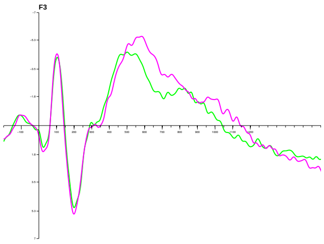


green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

Figure C.17: LD-IA — ERP's effect lexical congruency but ONLY for items MINUS initial accent

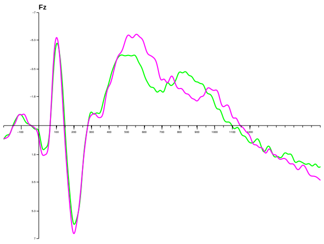
C.3.5.6 Effect lexical congruency; condition → with initial accent

F3



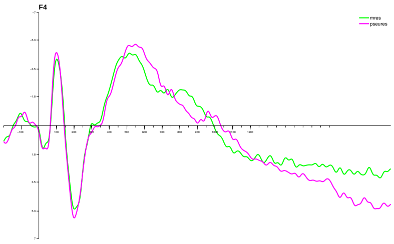
green:word magenta:pseudoword

FZ



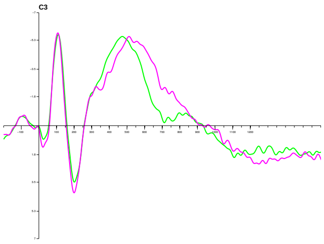
green:word magenta:pseudoword

F4



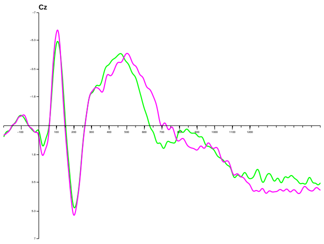
green:word magenta:pseudoword

C3



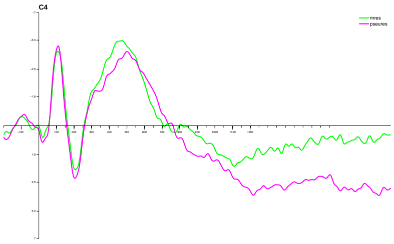
green:word magenta:pseudoword

CZ



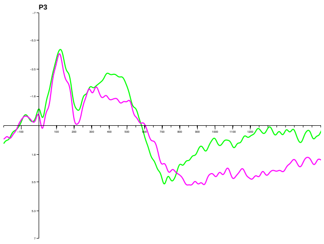
green:word magenta:pseudoword

C4



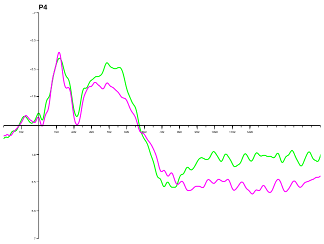
green:word magenta:pseudoword

P3



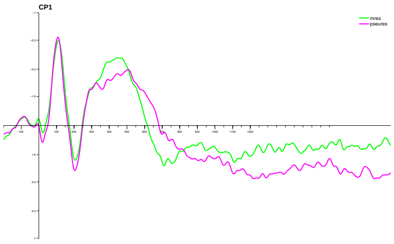
green:word magenta:pseudoword

P4



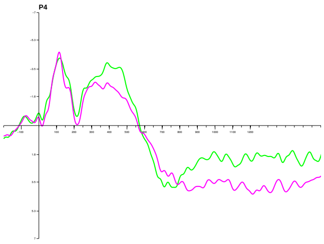
green:word magenta:pseudoword

CP1



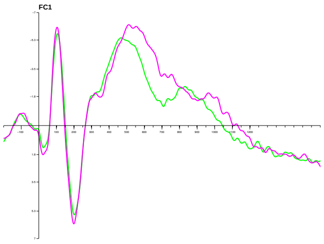
green:word magenta:pseudoword

CP2



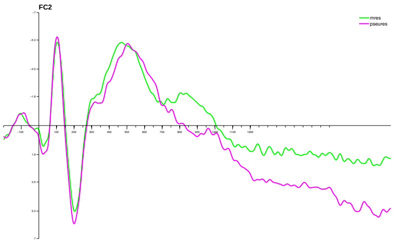
green:word magenta:pseudoword

FC1



green:word magenta:pseudoword

FC2



green:word magenta:pseudoword

Figure C.18: LD-IA — ERP's effect lexical congruency but ONLY for items PLUS initial accent

C.4 Lexical decision results—FA

C.4.1 Study overview

RQ:	Is FA expected by listeners and does presenting stress template —FA hamper stress extraction in a way similar as in Böcker et al. (1999); te Rietmolen et al. (2016)?
Hypotheses:	The N325 will be bigger when stimuli are presented —FA. The stress manipulation is later than in te Rietmolen et al. (2016) so that we could now observe an interaction between metrical processing and lexicality. In the event of an interaction, we'd expect a bigger N325 for words —FA than for pseudowords —FA.

C.4.1.1 Procedure

Nr participants:	20 listeners (0 excluded)
Nr stimuli per condition:	80 (4 lists of 160 phrases, 4 conditions)
Task:	Lexical decision task (random left/right assignment)
ISI:	Participants had a maximum of 1s post stimulus-offset to respond. An ITI of 600 ms followed either the respons or the stimulus offset (whichever was last).

C.4.1.2 Stimuli

Manipulation:	mostly duration, f_0 when necessary (see section 5.1.3)
Some descriptives:	

Table C.7: Durations ofwords andpseudowords: totalduration, 1stsyllable, 2ndsyllable, 3rdsyllable andthe onsetofthe 3rd(FA) syllable.

	Total duration		1st syllable		2nd syllable		3rd syllable		3rd onset	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
WORD										
+FA	647.9	62.5	176.6	36.1	168.3	30.1	303.6	54.6	345.2	40.5
−FA	565.9	69.0	176.2	35.8	167.7	30.5	221.7	58.7	346.5	56.4
PSEUDOWORD										
+FA	658.9	44.8	169.6	25.3	175.3	31.3	314.0	47.2	345.0	31.9
−FA	564.0	49.0	169.6	25.3	175.3	31.3	219.0	47.2	345.0	31.9

Uniqueness point:

C.4.1.3 EEG preprocessing

Nr electrodes: 64

Reference: Average

Filter and down-sampling: 0.01 − 30 bandpass, 128Hz

Epoch length: −200 − 1500

C.4.1.4 Analysis

Descriptive statistic behavioral:

Table C.8: Overview reaction times per condition

Condition	mean	sd	max	min	Condition	mean	sd	max	min
minword	1011.292	224.2460	2290.721	612.2959	plus	1064.061	264.6636	2474.524	612.2959
minpseu	1102.840	274.5048	2461.479	623.5240	min	1058.895	255.6739	2461.479	651.0589
plusword	1001.813	206.7938	2444.667	651.0589	word	1006.579	215.7004	2444.667	612.2959
pluspseu	1121.363	297.4497	2474.524	658.1290	pseu	1112.193	286.2076	2474.524	623.5240

Statistical analysis: Linear Mixed Effects Model in R

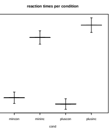
Descriptive statistic ERP:

Statistical analysis: $t_{\max}$ mass univariate permutation test, 2500 permutations in Matlab

Electrodes: 9 electrodes (Fz, Cz, FC1, FC2, CPz, AFz, Fpz, F1, F2)

Time-windows: 151 − 251 for P2; 546 − 776 for N325[1]

C.4.1.5 Results



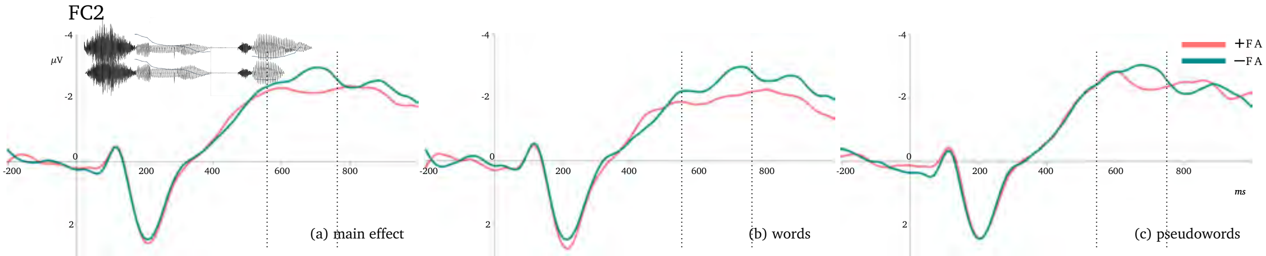
Behavioral: *Effect lexicality*

Lexicality was a significant predictor of reaction times, participants were slower to respond to pseudowords than to lexical words.

P2: No differences. Whatever significant differences we find in the later components cannot be attributed to differences on the early P2. Also, failing to replicate the P2 lexicality effect we found in the lexical decision study with IA indicates that it indeed was caused by our ±IA manipulation and the temporal overlap between the P2 and the N325.

N325 *Effect of final accent; main effect* ($t_{\max}$ p between 0.037 and 0.03 at FC2 between 734 − 740ms).

Stimuli (words and pseudowords) without final accent resulted in a larger negativity compared to stimuli with final accent in frontal central electrodes and in the N325 time-window (546 − 776) . This indicates an expectation for stimuli to be presented with final accent, although, as is described below, the effect seems to come mainly from increased processing difficulties within the word condition, with no effect within the pseudoword condition .



(a) main effect (b) words (c) pseudowords

Marginal effect of final accent in cluster analysis; for lexical words ($t_{\max}$ $p = 0.15$, so no effect).

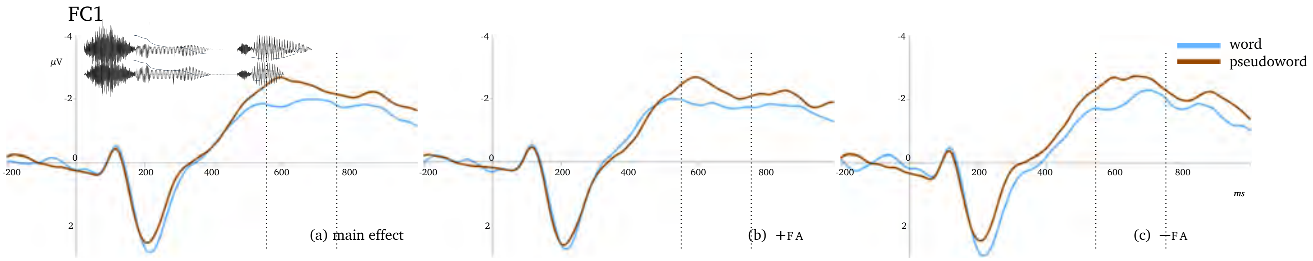
Words without final accent resulted in a bigger N325 than did words with final accent, although this effect was only marginal.

Effect of lexical congruency; main effect ($t_{\max}$ $p = 0.03$ at FC1 and Fz between 585—601).

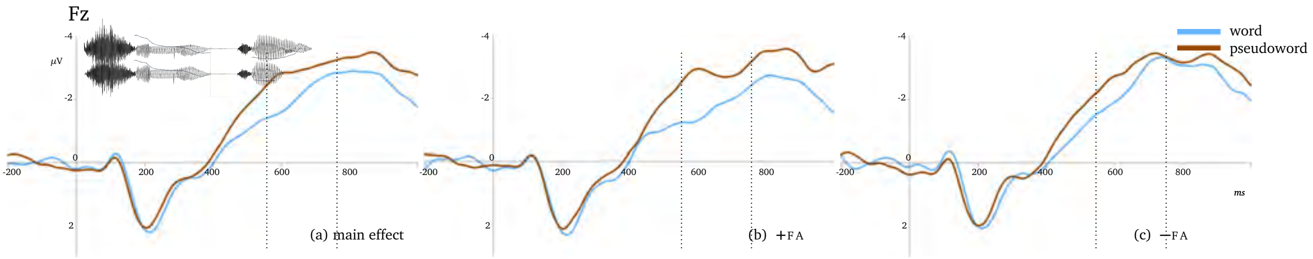
---

[1] *reminder:* I also tested the N400 time-window, but because we no expectations for the N400, I am leaving the results out of this study summary.

There was a bigger amplitude for pseudowords than lexical words. This could be an N400 effect, although it is frontal. It could also be the result of an interaction between the lexical and metrical manipulations.



(a) main effect (b) +FA (c) −FA



(a) main effect (b) +FA (c) −FA

Effect lexical congruency; only for with final accent ($t_{\max}$ p between 0.0007 and 0.04 at Fz between $578-600$). Fz is significantly more negative for words than for pseudowords between 578 ms and 600 ms after stimulus presentation in the condition with final accent.

C.4.1.6 Conclusion

Again we find that listeners expect words to be marked now with the final accent. However, because the final accent is heard late in word processing, this time we found an interaction between lexical congruency and presence of FA. FA interacts with lexical processing such that it facilitates word processing, but not pseudoword processing as reflected by the effect of FA only for words and by the effect of lexicality only for +FA stimuli.

C.4.2 Behavioral analyses

C.4.2.1 Descriptives

Table C.9: Overview reaction times per condition

Condition	mean	sd	max	min	Condition	mean	sd	max	min
minword	1011.292	224.2460	2290.721	612.2959	plus	1064.061	264.6636	2474.524	612.2959
minpseu	1102.840	274.5048	2461.479	623.5240	min	1058.895	255.6739	2461.479	651.0589
plusword	1001.813	206.7938	2444.667	651.0589	word	1006.579	215.7004	2444.667	612.2959
pluspseu	1121.363	297.4497	2474.524	658.1290	pseu	1112.193	286.2076	2474.524	623.5240

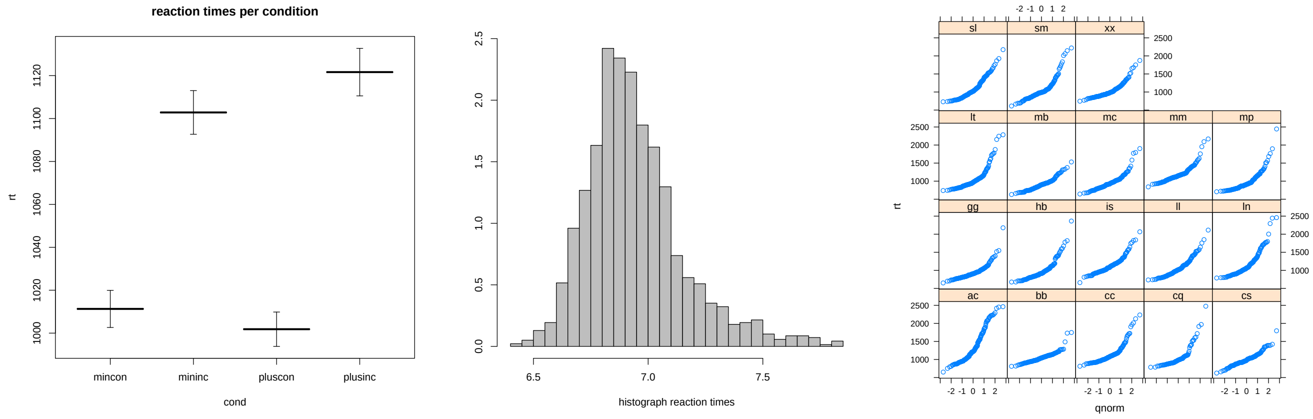


Figure C.19: Elaborate caption because there's quite a lot to say about these figures.

C.4.2.2 Regression models

Table C.10: Overview linear mixed models. The model fitting the data best takes lexicality as fixed factor and subjects and stimuli variability as random factors. Presence of final accent does not significantly contribute as a prectictor of reaction times.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	1061.94***	1059.83***	1112.79***	1054.66***	1107.89***	1101.76***
	(24.46)	(25.47)	(26.07)	(25.85)	(26.44)	(26.77)
lexicality			−108.71***		−108.71***	−96.05***
			(13.82)		(13.84)	(16.34)
±FA				10.40	9.83	22.14
				(8.80)	(8.75)	(12.19)
±FA:lexicality						−25.42
						(17.51)
AIC	38591.28	38415.85	38358.03	38410.27	38352.59	38344.92
BIC	38609.08	38457.39	38405.50	38457.74	38406.00	38404.26
Log Likelihood	−19292.64	−19200.93	−19171.01	−19197.13	−19167.29	−19162.46
Num. obs.	2792	2792	2792	2792	2792	2792
Num. groups: subj	18	18	18	18	18	18
Var: subj (Intercept)	10389.67	10491.87	10566.46	10499.48	10573.56	10575.08
Var: Residual	57741.78	50200.34	50185.95	50185.00	50171.06	50157.33
Num. groups: stimuli		160	160	160	160	160
Num. groups: lex:stimuli		160	160	160	160	160
Var: stimuli (Intercept)		1.17	2103.46	7371.37	1394.45	1227.17
Var: lex.stimuli (Intercept)		7552.73	2208.74	209.28	0.00	3466.80

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.11: ANOVA: lexicality as significant predictor for reaction times

ANOVA	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
ld—intercept	3	38599	38617	-19297	38593			
ld—random structure	7	38424	38466	-19205	38410	183.3	4	0.00000
ld— ±FA	8	38425	38472	-19204	38409	1.4	1	0.23766
ld—lexicality	8	38373	38421	-19179	38357	51.4	0	0.00000
ld—interaction	10	38374	38433	-19177	38354	2.1	1	0.14627

C.4.3 Non-parametric results

Separate repeated measures, two-tailed permutation tests based on the t-max statistic were conducted in Matlab for each of the time windows described above and each of the comparisons of interest, i.e.:

Time-windows:

- P2 → 151 − 251 ms
- N400 → 546 − 776 ms (this is the LD-IA N325 time-window + 345, third syllable onset)
- N325 → 381 − 481 ms

Comparisons:

- main effect ±FA
- main effect lexicality
- ±FA ∗ lexicality (4 comparisons)

11 electrodes were selected for analysis: Fz, Cz, FC1, FC2, CPz, AFz, Fpz, F1, F2, Cp1, CP3. The data were down-sampled to 128 Hz versus 125 in LD-IA. EEG was referenced to the average of the electrodes instead of the mastoids in LD-IA (I re-analyzed the data referenced to the mastoids (see **??**) to see if reference had a big effect, it did not.). 2500 permutations were used to estimate the distribution of the null hypothesis for the customary family-wise α level of 0.05.

C.4.3.1 P200

Effect final accent; main effect

```
GND=tmaxGND(GND,9,’time_wind’,[151 251],’include_chans’, {’Fz’,’Cz’,’FC1’,’FC2’,’CPz’,’A
```

Testing null hypothesis that the grand average ERPs in Bin 14 (plusfa-minfa) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8962 and 3.8962
Test-wise alpha level	0.000971
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p-values >= 0.249200$

Effect lexical congruency; main effect Testing null hypothesis that the grand average ERPs in Bin 14 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.82 and 3.82
Test-wise alpha level	0.001165.
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p-value >= 0.169200$

Effect final accent; condition → lexical words Testing null hypothesis that the grand average ERPs in Bin 11 (mot: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.793 and 3.793
Test-wise alpha level	0.001229
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.538800$

Effect final accent; condition → pseudowords Testing null hypothesis that the grand average ERPs in Bin 12 (pseu: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.3521 and 3.3521
Test-wise alpha level	0.002883
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.429600$

Effect lexical congruency; condition → min final accent Testing null hypothesis that the grand average ERPs in Bin 13 (minai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.9188 and 3.9188
Test-wise alpha level	0.000922
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - value >= 0.665600$

Effect lexical congruency; condition → with final accent Testing null hypothesis that the grand average ERPs in Bin 14 (plusai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8173 and 3.8173
Test-wise alpha level	0.001163
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.291200$

C.4.3.2 N325

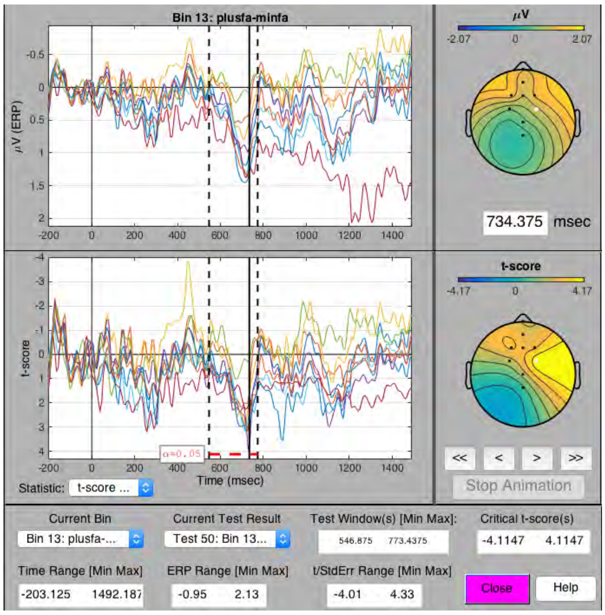
```
GND=tmaxGND(GND,14,'time_wind',[546 776],'include_chans', {'Fz','Cz','FC1','FC2',
```

Effect final accent; main effect Testing null hypothesis that the grand average ERPs in Bin 14 (plusfa-minfa) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	30
Nr comparisons	270
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0428 and 4.0428
Test-wise alpha level	0.000695
Bonferroni test-wise alpha	0.000185

Significant differences from zero: 734-740 ms, electrode(s): FC2

$p-values$ between 0.037000 and 0.030000



734.375 msec

Statistic: t-score ...

Current Bin: Bin 13: plusfa-...

Current Test Result: Test 50: Bin 13...

Test Window(s) [Min Max]: 546.875 773.4375

Critical t-score(s): -4.1147 4.1147

Time Range [Min Max]: -203.125 1492.187

ERP Range [Min Max]: -0.95 2.13

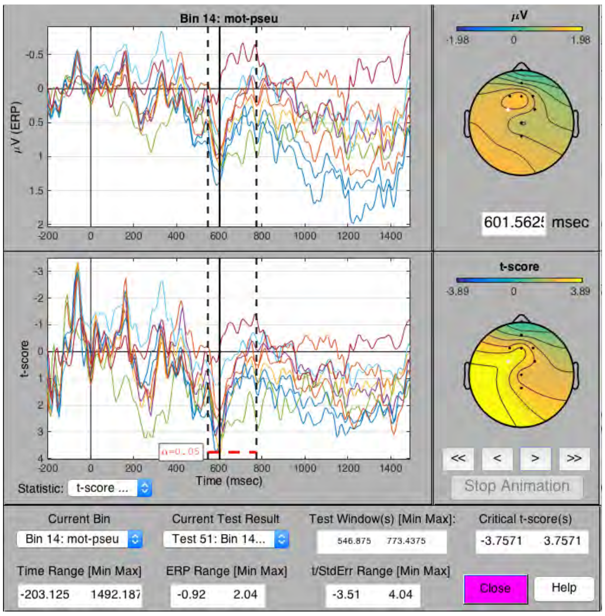
t/StdErr Range [Min Max]: -4.01 4.33

Effect lexical congruency; main effect Testing null hypothesis that the grand average ERPs in Bin 14 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	30
Nr comparisons	270
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.7571 and 3.7571
Test-wise alpha level	0.001220
Bonferroni test-wise alpha	0.000185

Significant differences from zero: 585-601 ms, electrode(s): Fz and FC1

$p-values = 0.037000$



601.562 msec

Statistic: t-score ...

Current Bin: Bin 14: mot-pseu

Current Test Result: Test 51: Bin 14...

Test Window(s) [Min Max]: 546.875 773.4375

Critical t-score(s): -3.7571 3.7571

Time Range [Min Max]: -203.125 1492.187

ERP Range [Min Max]: -0.92 2.04

t/StdErr Range [Min Max]: -3.51 4.04

Effect final accent; condition → lexical words Testing null hypothesis that the grand average ERPs in Bin 11 (mot: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	30
Nr comparisons	270
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0239 and 4.0239
Test-wise alpha level	0.000725
Bonferroni test-wise alpha	0.000185

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed.
All $p-values >= 0.150800$

Effect final accent; condition → pseudowords Testing null hypothesis that the grand average ERPs in Bin 12 (pseu: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	30
Nr comparisons	270
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.9563 and 3.9563
Test-wise alpha level	0.000847
Bonferroni test-wise alpha	0.000185

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p-values >= 0.456800$

Effect lexical congruency; condition → min final accent Testing null hypothesis that the grand average ERPs in Bin 13 (minai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

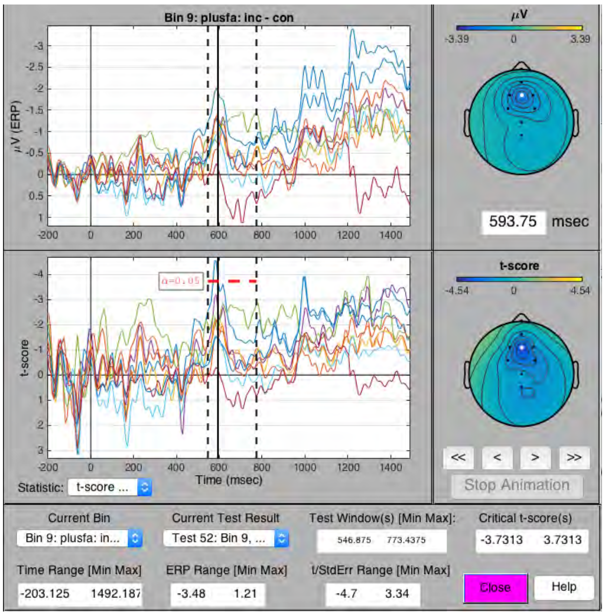
Nr channels	9
Nr time points	30
Nr comparisons	270
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8915 and 3.8915
Test-wise alpha level	0.000982
Bonferroni test-wise alpha	0.000185

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.148800$

Effect lexical congruency; condition → with final accent Testing null hypothesis that the grand average ERPs in Bin 14 (plusai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	30
Nr comparisons	270
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.7313 and 3.7313
Test-wise alpha level	0.001149
Bonferroni test-wise alpha	0.000185

Significant differences from zero: 578-600 ms, electrode(s): Fz
$p - values$ between 0.007200 and 0.043600



Statistic: t-score ...
Current Bin: Bin 9: plusfa: in...
Current Test Result: Test 52: Bin 9, ...
Test Window(s) [Min Max]: 546.875 773.4375
Critical t-score(s): -3.7313 3.7313
Time Range [Min Max]: -203.125 1492.187
ERP Range [Min Max]: -3.48 1.21
t/StdErr Range [Min Max]: -4.7 3.34

C.4.3.3 N400

Effect final accent; main effect Testing null hypothesis that the grand average ERPs in Bin 14 (plusfa-minfa) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

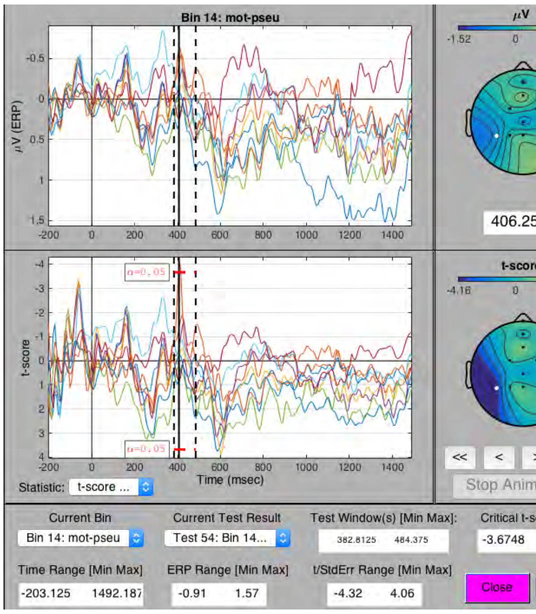
Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.9249 and 3.9249
Test-wise alpha level	0.000910
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.061200$

Effect lexical congruency; main effect Testing null hypothesis that the grand average ERPs in Bin 14 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6369 and 3.6369
Test-wise alpha level	0.001755
Bonferroni test-wise alpha	0.000397

Significant differences from zero: 406-414 ms, electrode(s): CP3
$p - values$ between 0.028000 and 0.015600



406.25

Statistic: t-score ...

Current Bin: Bin 14: mot-pseu

Current Test Result: Test 54: Bin 14...

Test Window(s) [Min Max]: 382.8125 484.375

Critical t-s: -3.6748

Time Range [Min Max]: -203.125 1492.187

ERP Range [Min Max]: -0.91 1.57

t/StdErr Range [Min Max]: -4.32 4.06

Effect final accent; condition → lexical words Testing null hypothesis that the grand average ERPs in Bin 11 (mot: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.7766 and 3.7766
Test-wise alpha level	0.001276
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.415200$

Effect final accent; condition → pseudowords Testing null hypothesis that the grand average ERPs in Bin 12 (pseu: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6908 and 3.6908
Test-wise alpha level	0.001552.
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.718400$

Effect lexical congruency; condition → min final accent Testing null hypothesis that the grand average ERPs in Bin 13 (minai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8382 and 3.8382
Test-wise alpha level	0.001109
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.822800$

Effect lexical congruency; condition → with final accent Testing null hypothesis that the grand average ERPs in Bin 14 (plusai: pseu - mot) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

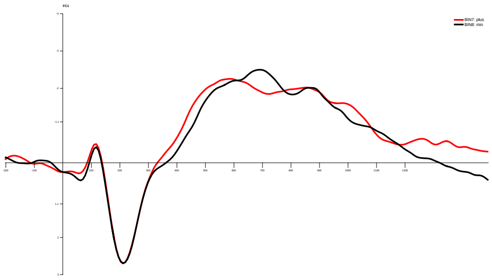
Nr channels	9
Nr time points	14
Nr comparisons	126
t-score degrees of freedom	19
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.6851 and 3.6851
Test-wise alpha level	0.001572
Bonferroni test-wise alpha	0.000397

ERPs are NOT significantly different from zero ($\alpha = 0.05$) at any time point or window analyzed
All $p - values >= 0.162400$

C.4.4 All erp plots — average

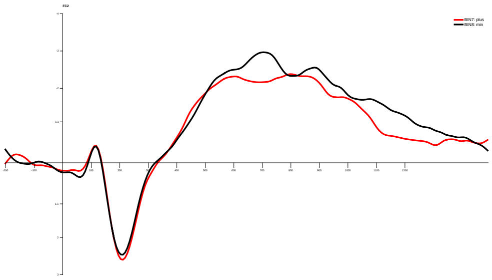
C.4.4.1 Effect final accent; main effect

FC1



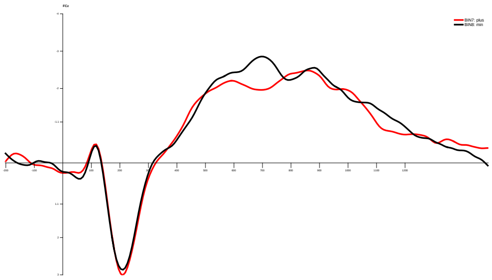
red:plus FA black:min FA

FC2



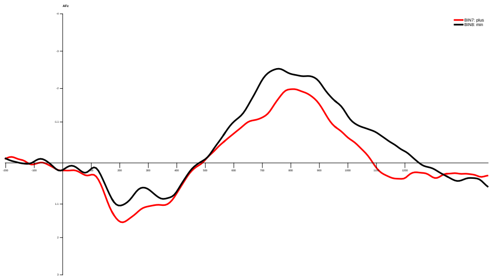
red:plus FA black:min FA

FCz



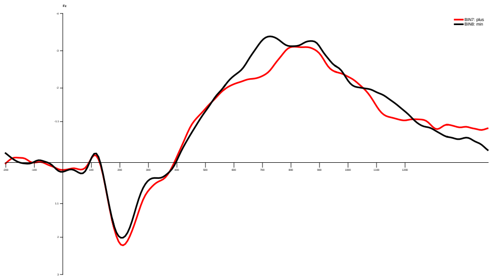
red:plus FA black:min FA

AFz



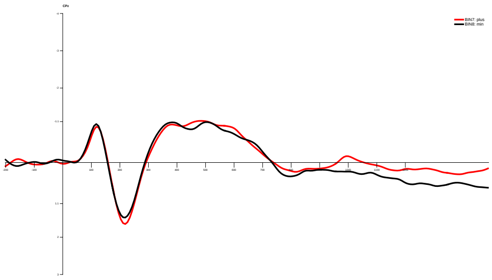
red:plus FA black:min FA

Fz



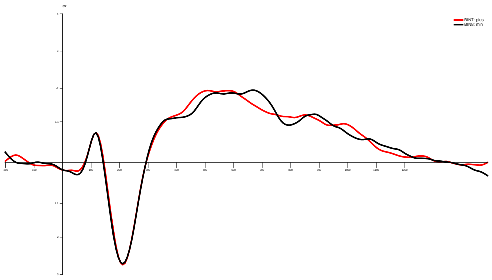
red:plus FA black:min FA

CPz



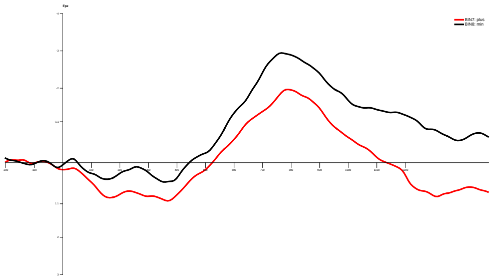
red:plus FA black:min FA

Cz



red:plus FA black:min FA

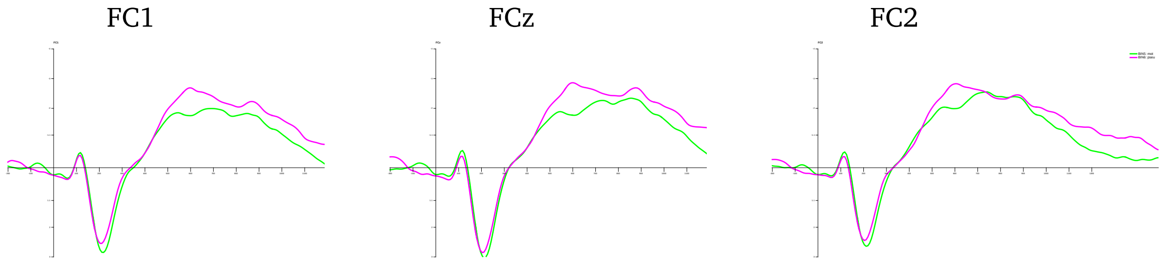
FPz



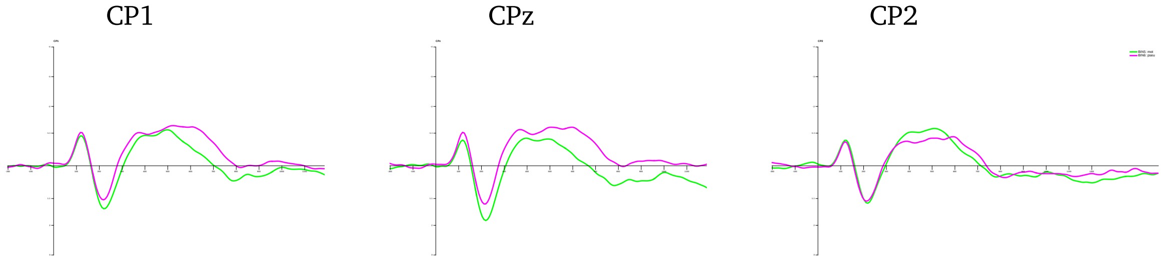
red:plus FA black:min FA

Figure C.20: LD-FA — ERP's main effect final accent: all plus FA word versus all min FA stimuli.

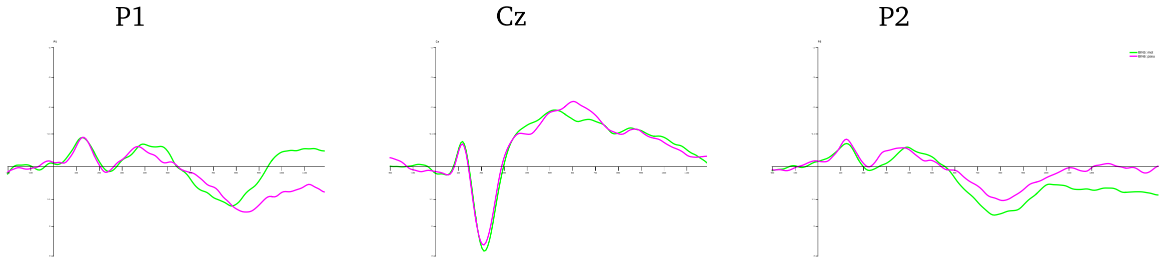
C.4.4.2 Effect lexical congruency; main effect



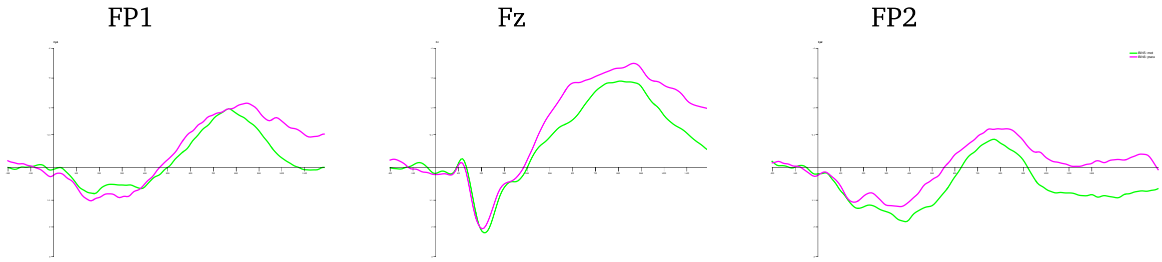
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword



green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword



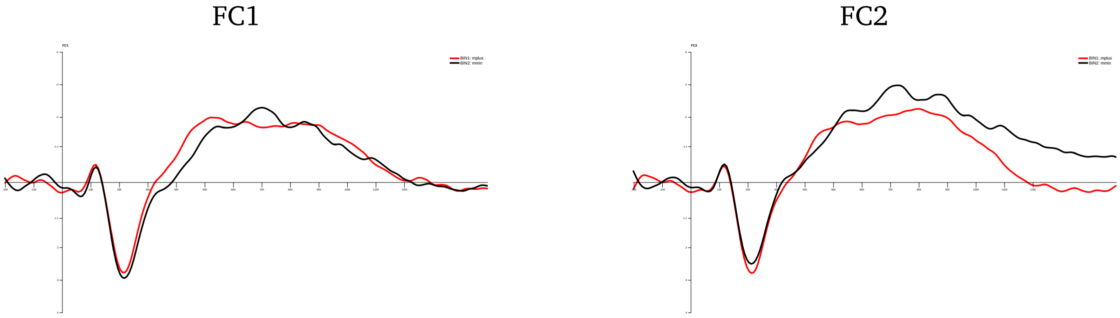
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword



green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

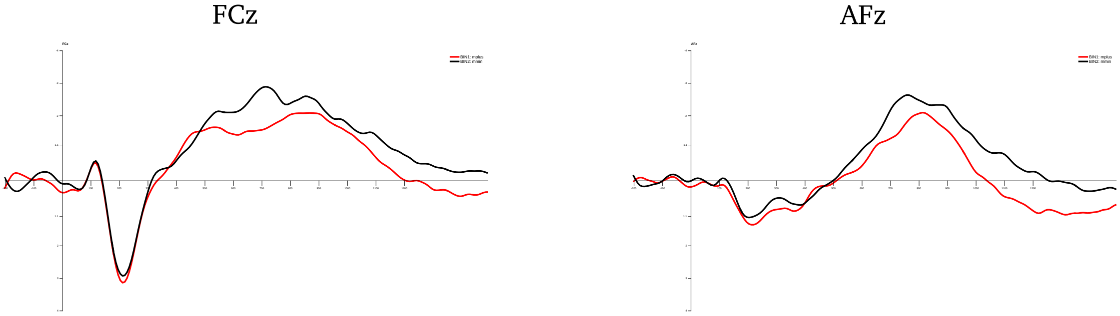
Figure C.21: LD-FA — ERP's main effect lexical congruency: all congruent words versus all pseudowords.

C.4.4.3 Effect final accent; condition → lexical words



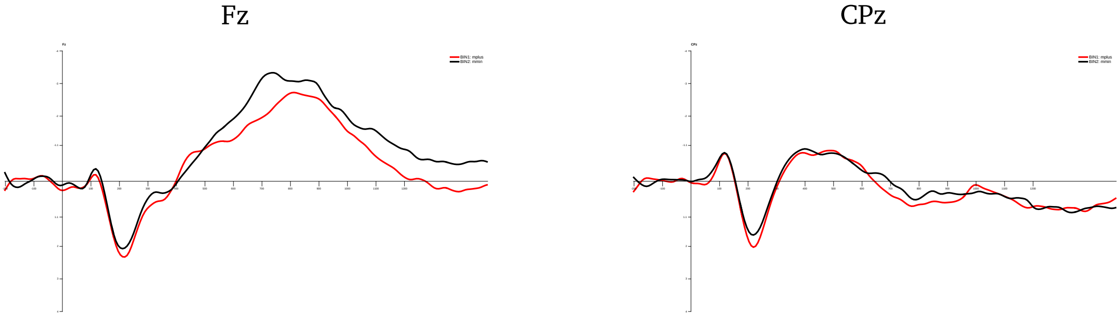
FC1 FC2

red:plus FA black:min FA red:plus FA black:min FA



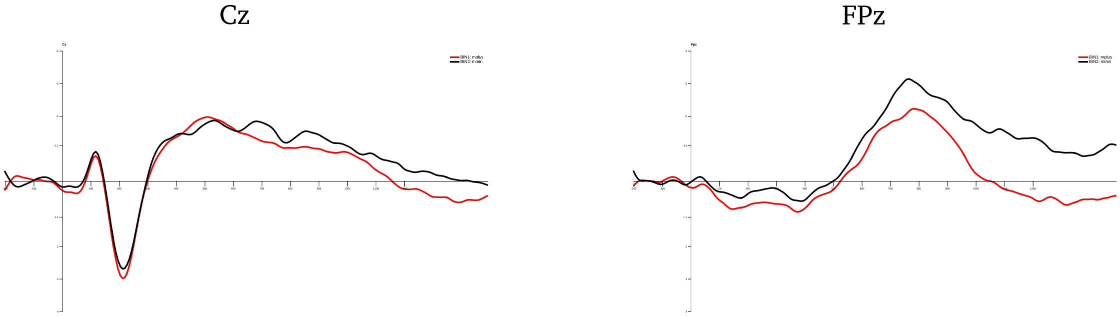
FCz AFz

red:plus FA black:min FA red:plus FA black:min FA



Fz CPz

red:plus FA black:min FA red:plus FA black:min FA

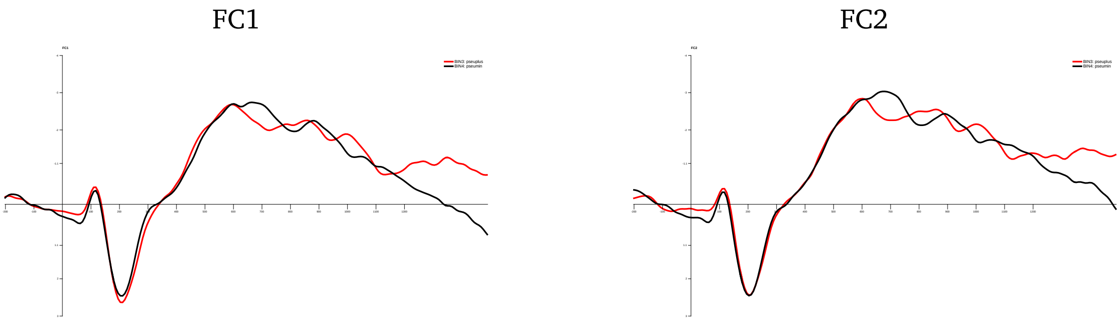


Cz FPz

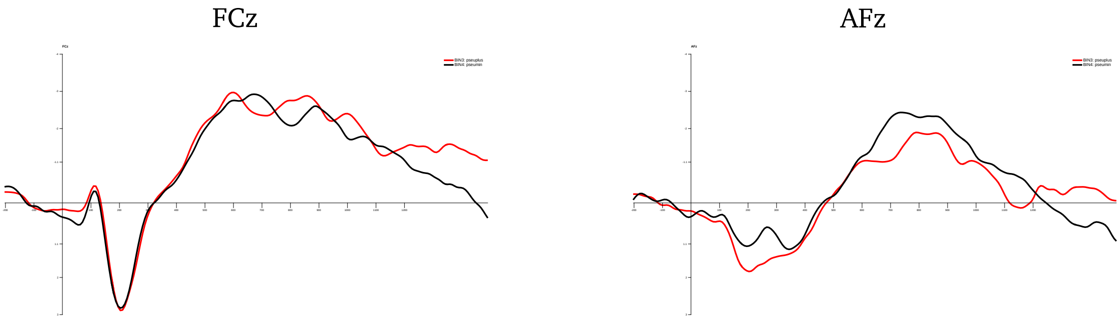
red:plus FA black:min FA red:plus FA black:min FA

Figure C.22: LD-FA — ERP's effect final accent but ONLY for lexical words.

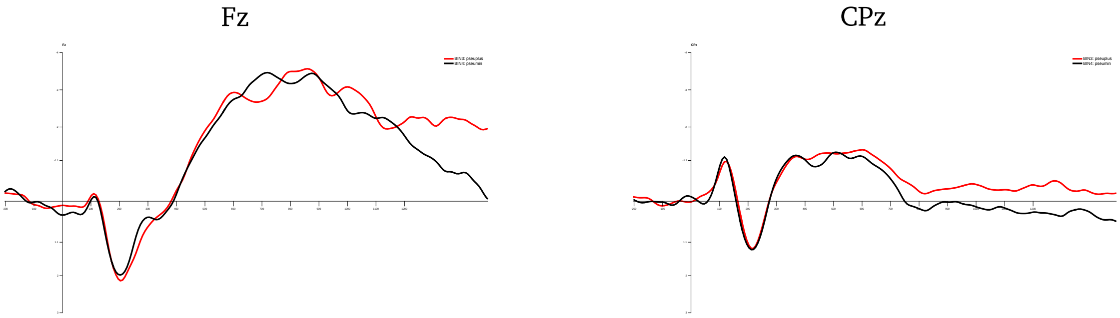
C.4.4.4 Effect final accent; condition → pseudowords



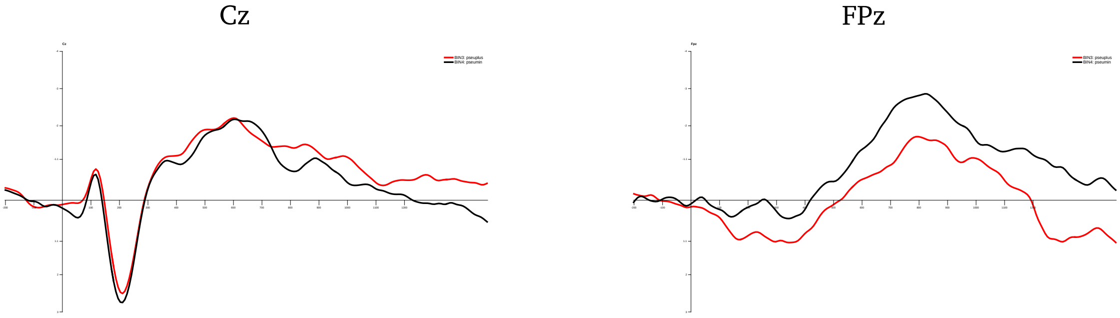
red:plus FA black:min FA red:plus FA black:min FA



red:plus FA black:min FA red:plus FA black:min FA



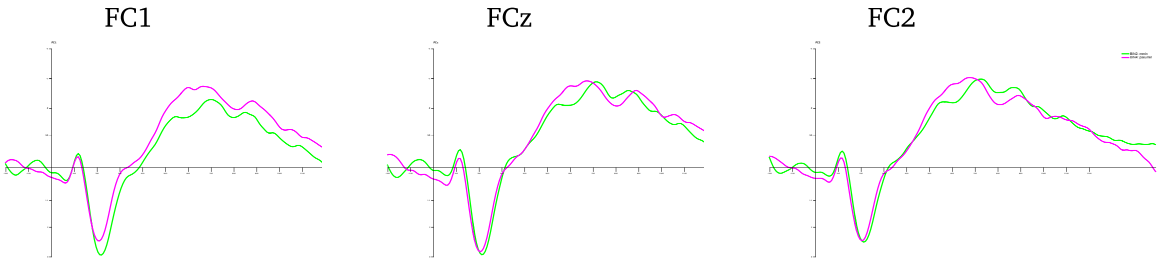
red:plus FA black:min FA red:plus FA black:min FA



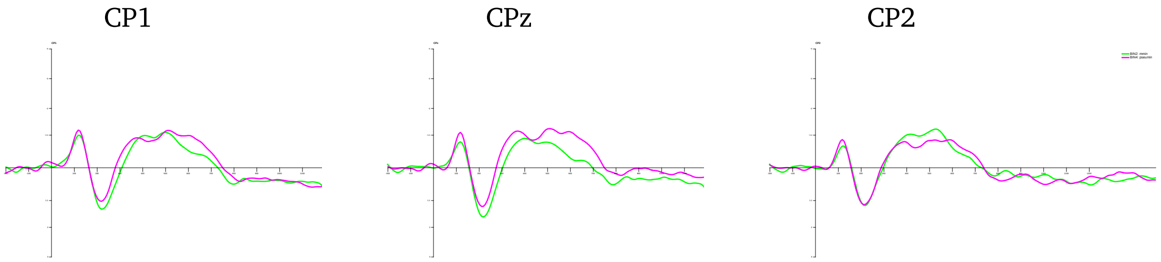
red:plus FA black:min FA red:plus FA black:min FA

Figure C.23: LD-FA — ERP's effect final accent but ONLY for pseudowords.

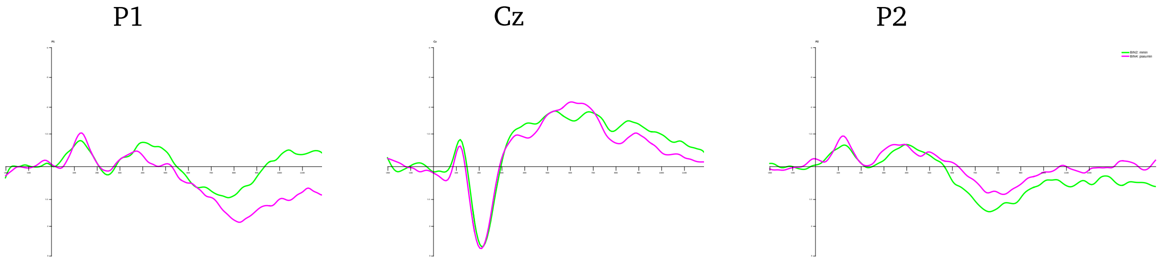
C.4.4.5 Effect lexical congruency; condition → min final accent



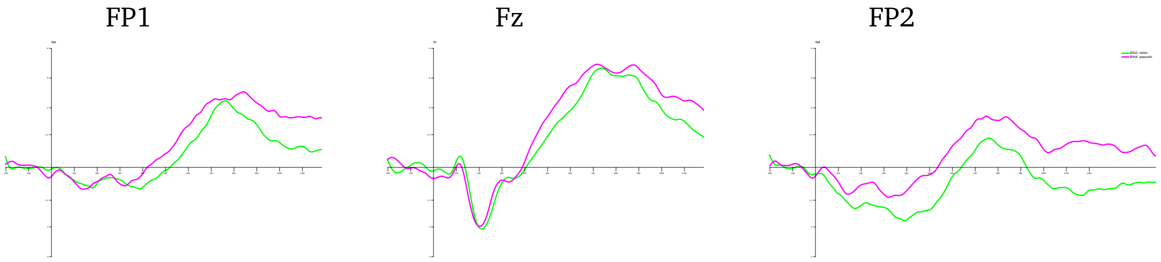
green:word magenta:pseudoword
green:word magenta:pseudoword
green:word magenta:pseudoword



green:word magenta:pseudoword
green:word magenta:pseudoword
green:word magenta:pseudoword



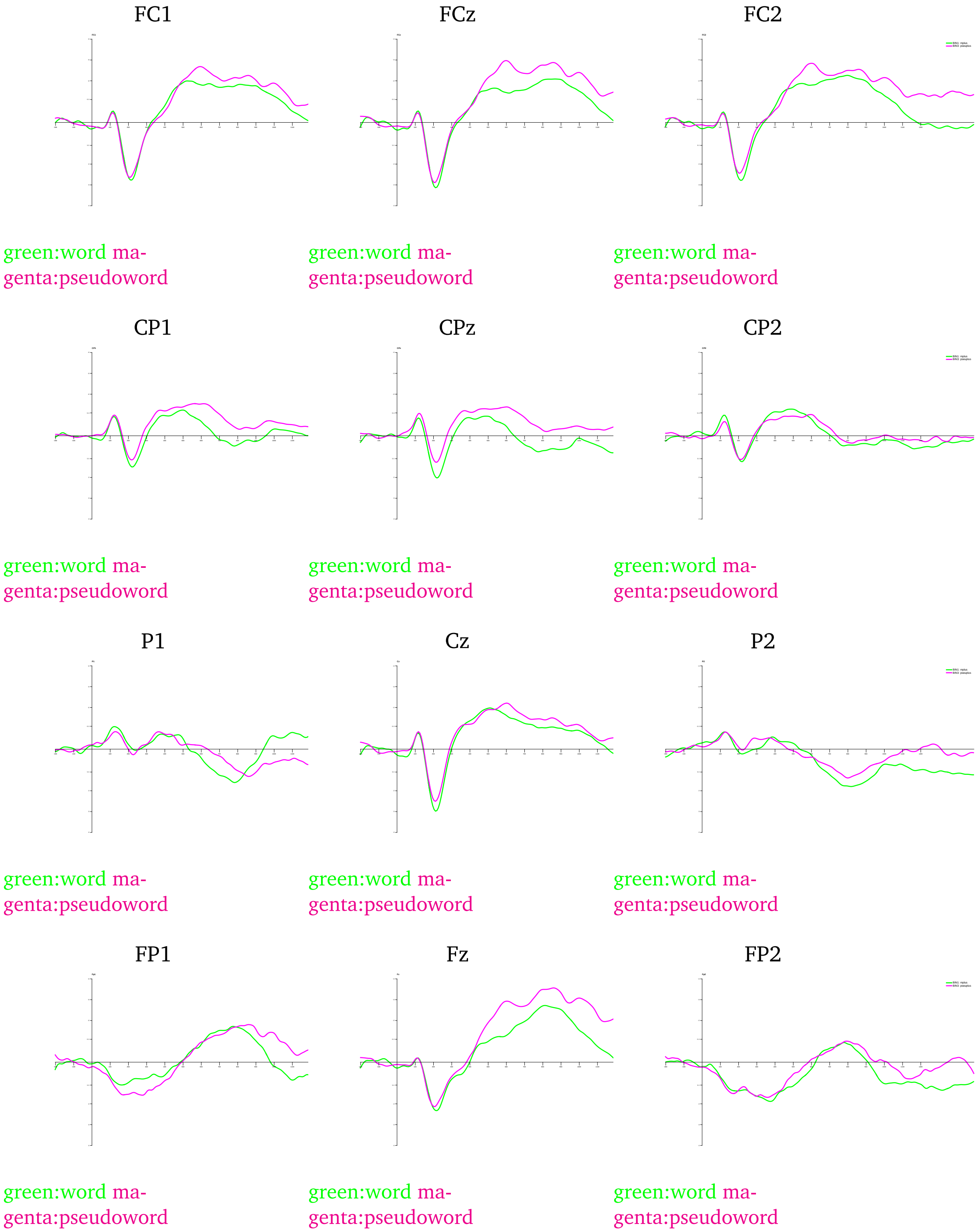
green:word magenta:pseudoword
green:word magenta:pseudoword
green:word magenta:pseudoword



green:word magenta:pseudoword
green:word magenta:pseudoword
green:word magenta:pseudoword

Figure C.24: LD-FA — ERP's effect lexical congruency but ONLY for items MINUS final accent

C.4.4.6 Effect lexical congruency; condition → with final accent



FC1

green:word magenta:pseudoword

FCz

green:word magenta:pseudoword

FC2

green:word magenta:pseudoword

CP1

green:word magenta:pseudoword

CPz

green:word magenta:pseudoword

CP2

green:word magenta:pseudoword

P1

green:word magenta:pseudoword

Cz

green:word magenta:pseudoword

P2

green:word magenta:pseudoword

FP1

green:word magenta:pseudoword

Fz

green:word magenta:pseudoword

FP2

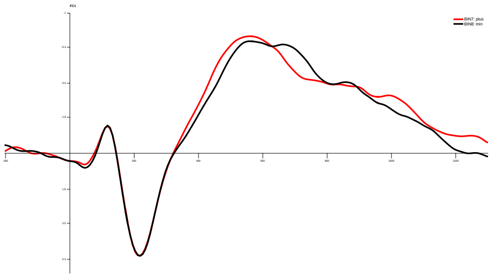
green:word magenta:pseudoword

Figure C.25: LD-FA — ERP's effect lexical congruency but ONLY for items PLUS final accent

C.4.5 All erp plots — mastoids

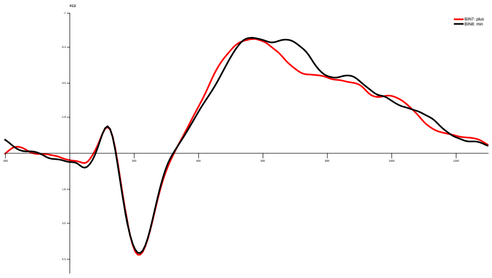
C.4.5.1 Effect final accent; main effect

FC1



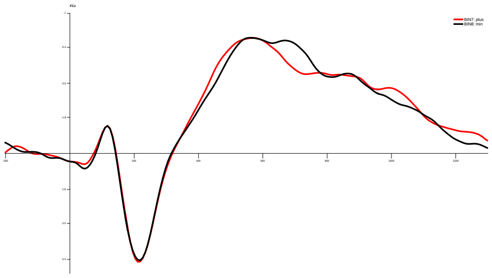
red:plus FA black:min FA

FC2



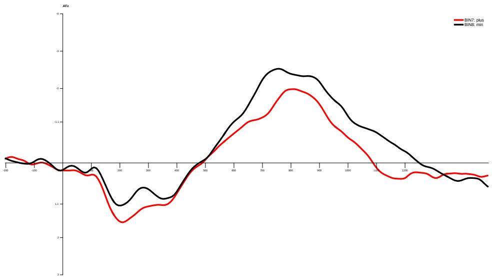
red:plus FA black:min FA

FCz



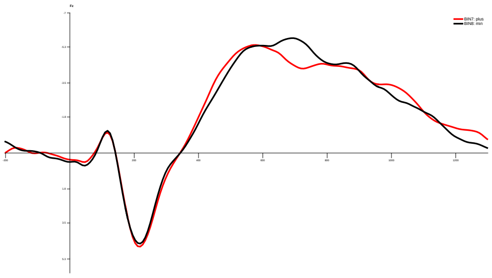
red:plus FA black:min FA

AFz



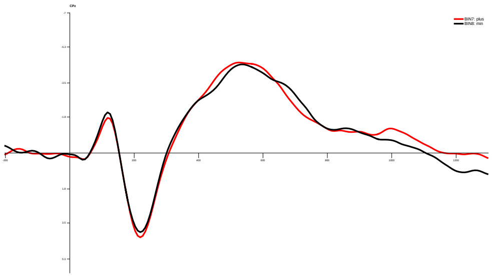
red:plus FA black:min FA

Fz



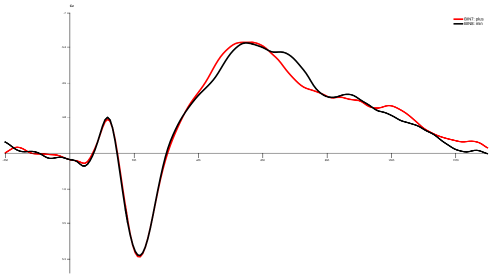
red:plus FA black:min FA

CPz



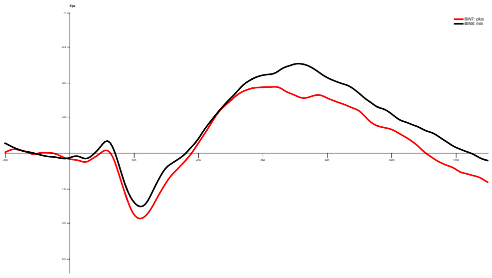
red:plus FA black:min FA

Cz



red:plus FA black:min FA

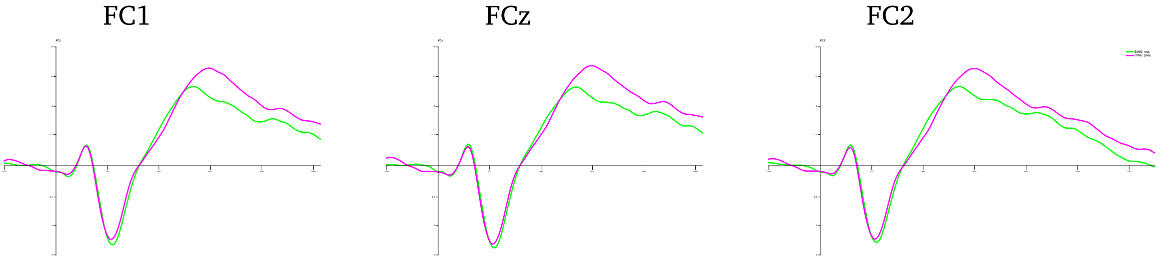
FPz



red:plus FA black:min FA

Figure C.26: LD-FA — ERP's main effect final accent: all plus FA word versus all min FA stimuli.

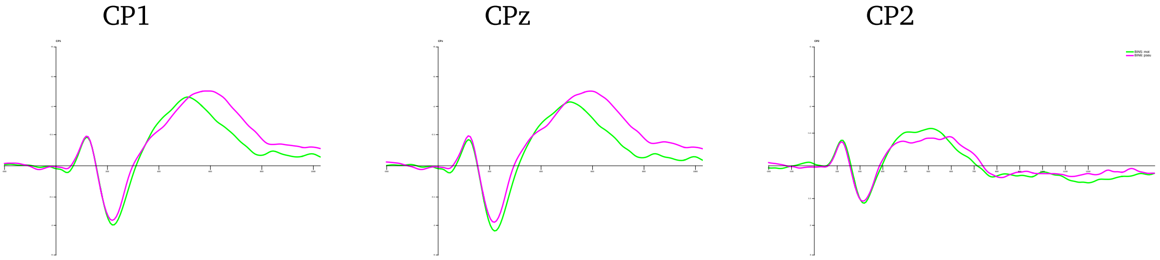
C.4.5.2 Effect lexical congruency; main effect



green:word magenta:pseudoword

green:word magenta:pseudoword

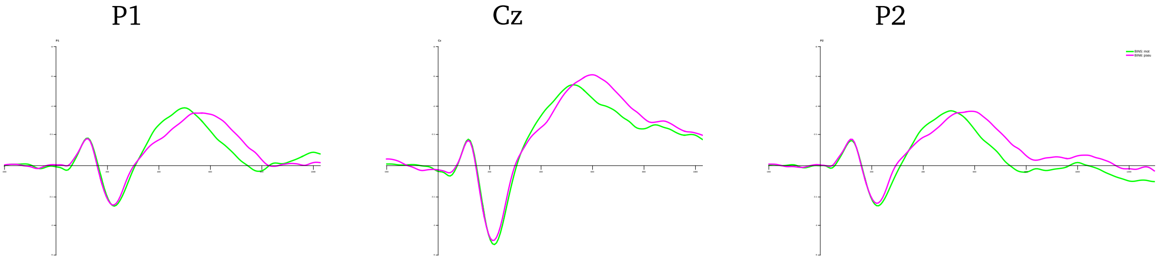
green:word magenta:pseudoword



green:word magenta:pseudoword

green:word magenta:pseudoword

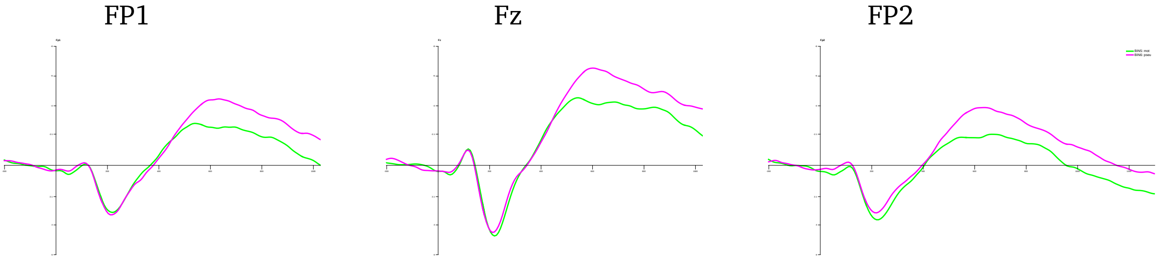
green:word magenta:pseudoword



green:word magenta:pseudoword

green:word magenta:pseudoword

green:word magenta:pseudoword



green:word magenta:pseudoword

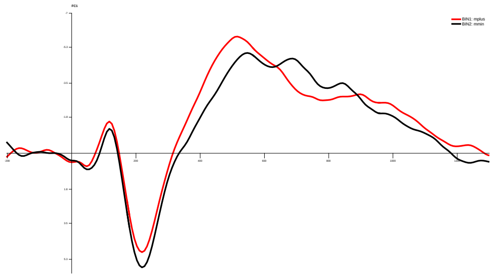
green:word magenta:pseudoword

green:word magenta:pseudoword

Figure C.27: LD-FA — ERP's main effect lexical congruency: all congruent words versus all pseudowords.

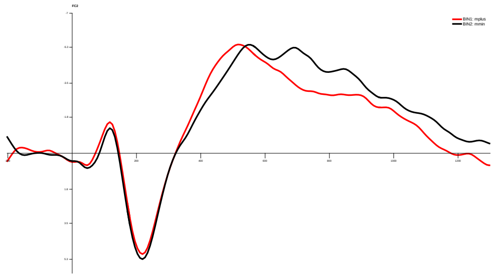
C.4.5.3 Effect final accent; condition → lexical words

FC1



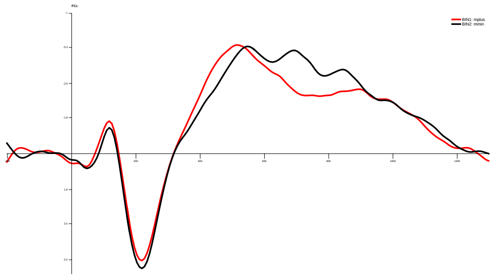
red:plus FA black:min FA

FC2



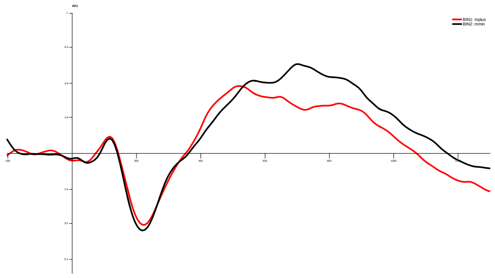
red:plus FA black:min FA

FCz



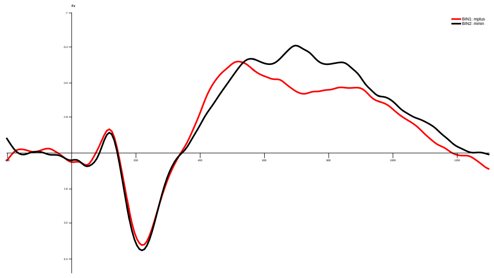
red:plus FA black:min FA

AFz



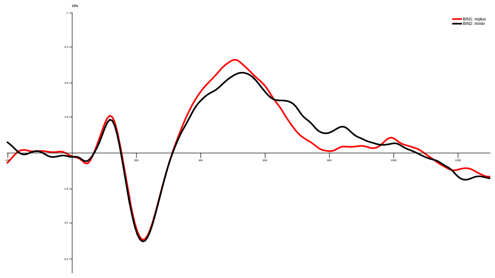
red:plus FA black:min FA

Fz



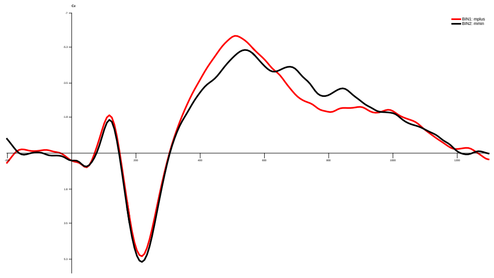
red:plus FA black:min FA

CPz



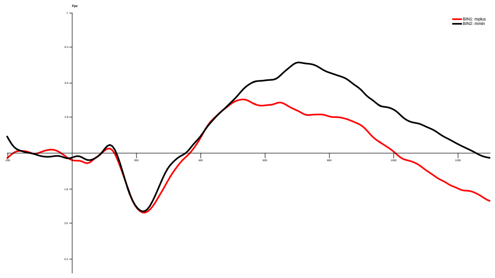
red:plus FA black:min FA

Cz



red:plus FA black:min FA

FPz

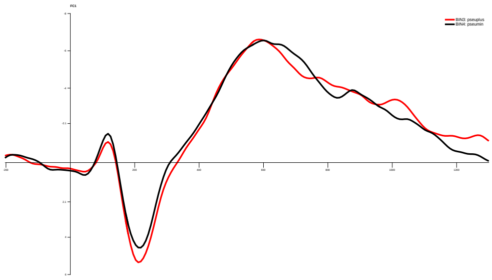


red:plus FA black:min FA

Figure C.28: LD-FA — ERP's effect final accent but ONLY for lexical words.

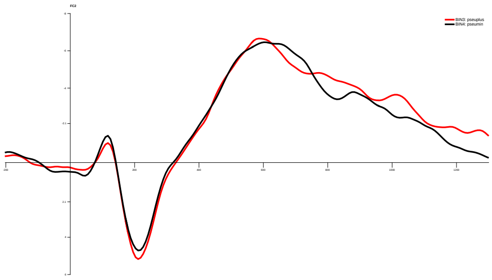
C.4.5.4 Effect final accent; condition → pseudowords

FC1



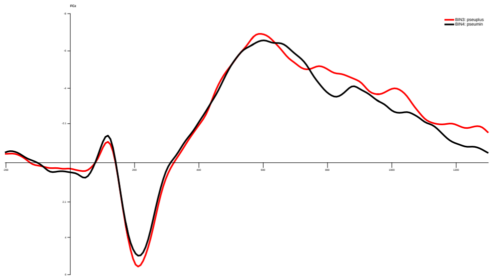
red:plus FA black:min FA

FC2



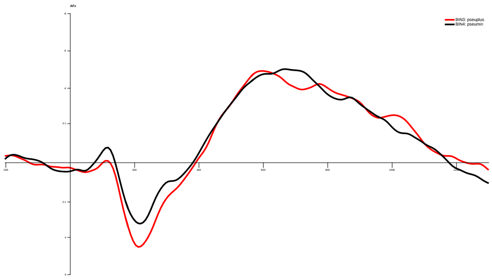
red:plus FA black:min FA

FCz



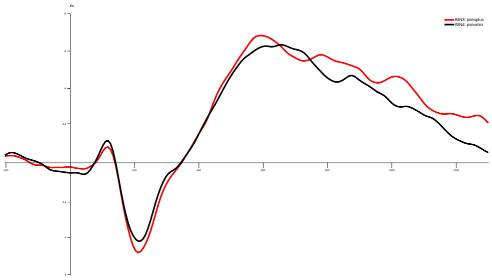
red:plus FA black:min FA

AFz



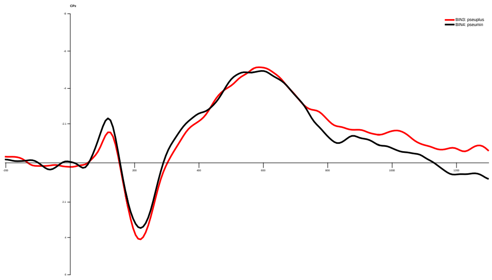
red:plus FA black:min FA

Fz



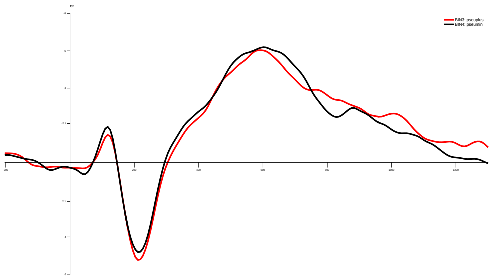
red:plus FA black:min FA

CPz



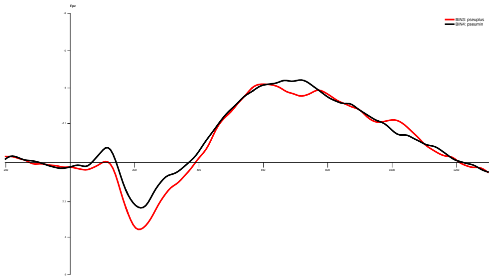
red:plus FA black:min FA

Cz



red:plus FA black:min FA

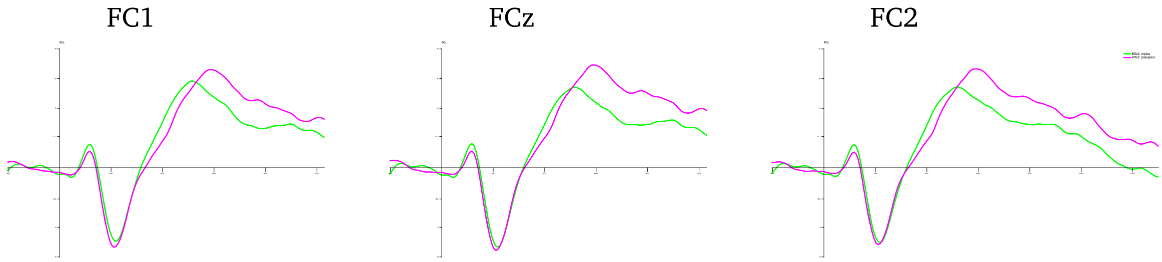
FPz



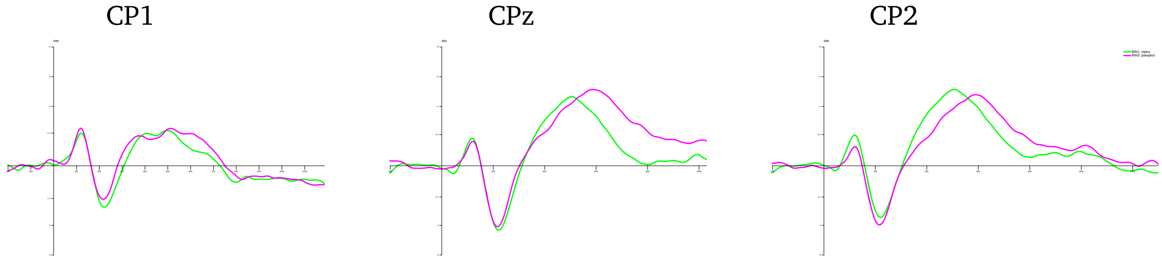
red:plus FA black:min FA

Figure C.29: LD-FA — ERP's effect final accent but ONLY for pseudowords.

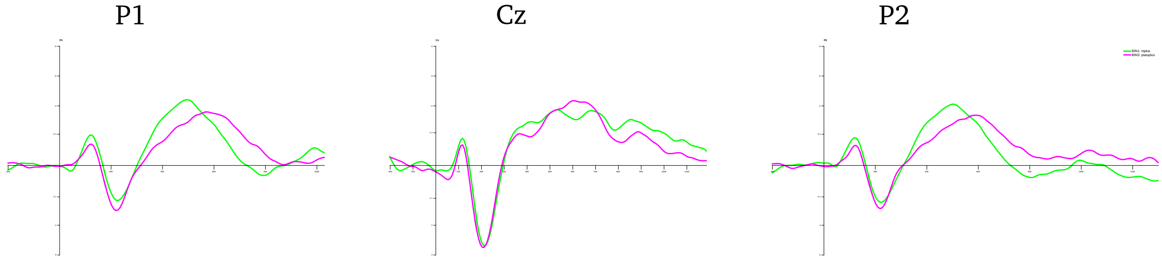
C.4.5.5 Effect lexical congruency; condition → min final accent


FC1 FCz FC2

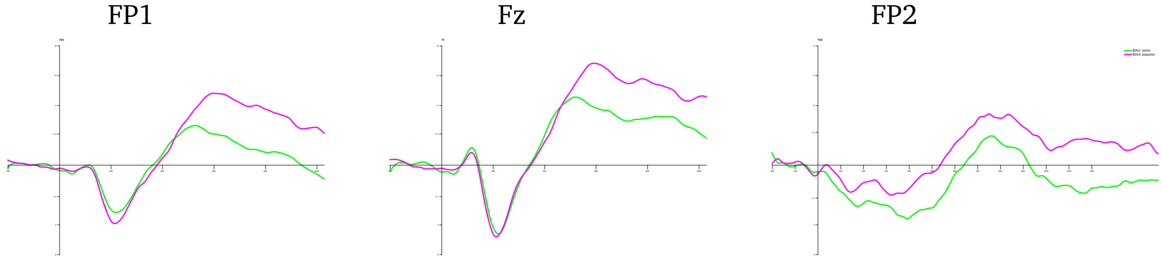
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword


CP1 CPz CP2

green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword


P1 Cz P2

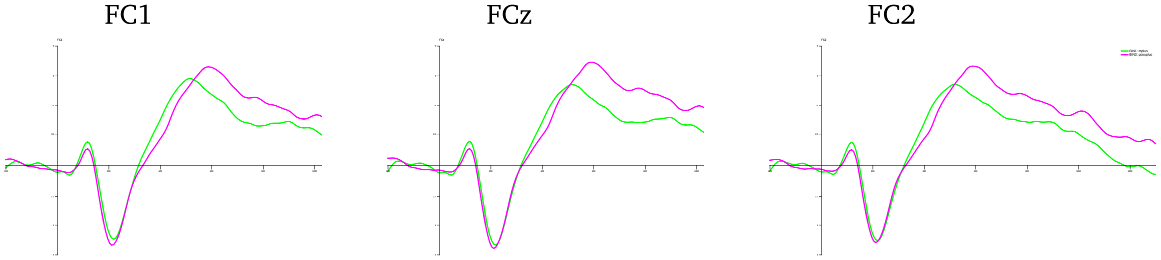
green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword


FP1 Fz FP2

green:word magenta:pseudoword green:word magenta:pseudoword green:word magenta:pseudoword

Figure C.30: LD-FA — ERP's effect lexical congruency but ONLY for items MINUS final accent

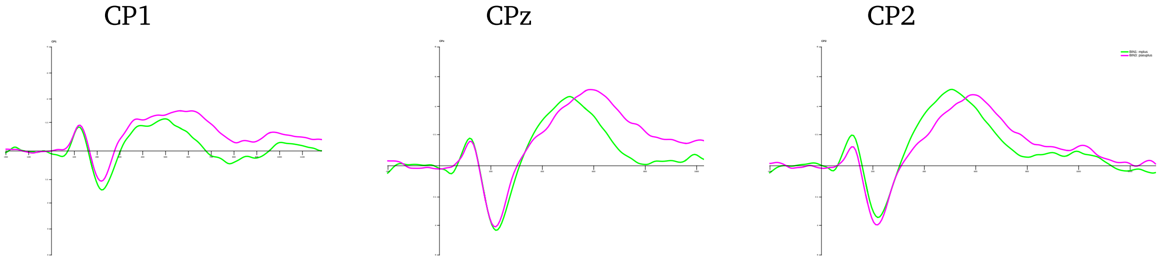
C.4.5.6 Effect lexical congruency; condition → with final accent



green:word magenta:pseudoword

green:word magenta:pseudoword

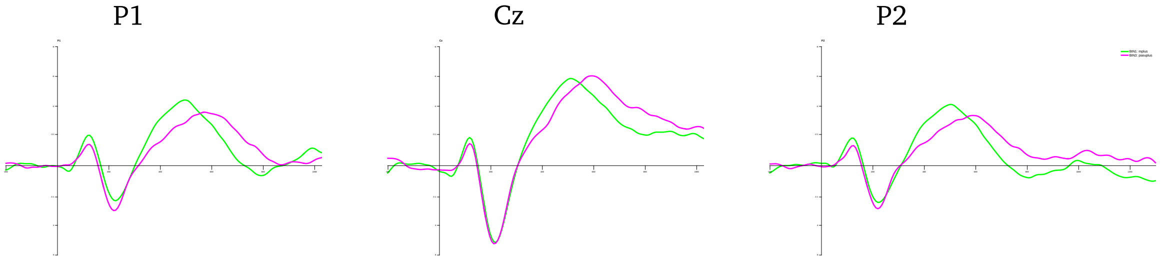
green:word magenta:pseudoword



green:word magenta:pseudoword

green:word magenta:pseudoword

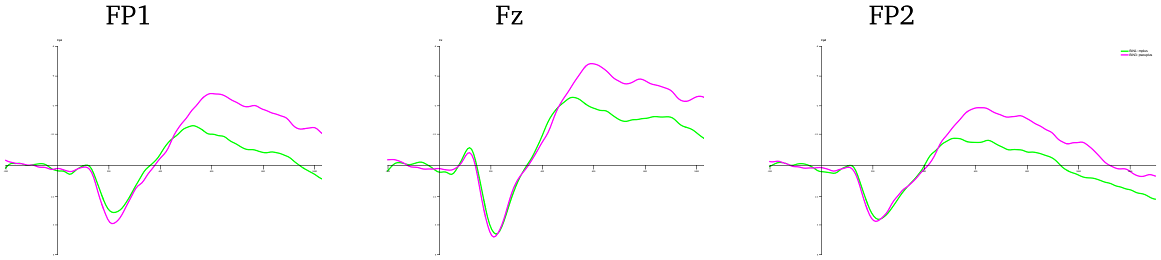
green:word magenta:pseudoword



green:word magenta:pseudoword

green:word magenta:pseudoword

green:word magenta:pseudoword



green:word magenta:pseudoword

green:word magenta:pseudoword

green:word magenta:pseudoword

Figure C.31: LD-FA — ERP's effect lexical congruency but ONLY for items PLUS final accent

C.5 N400 results—IA

C.5.1 Study overview

RQ: Does ±IA affect speech processing when it is embedded in a sentence and away from phrase boundaries? Does IA also affect later post-lexical stages of speech processing such as lexico-semantic processing?

Hypotheses: IA is lexically encoded and will affect speech processing even when it is not utterance initial. The N400 will be bigger when target-words are presented −IA. Further, there will be an interaction between metrical expectancy (metrical N400) and semantic expectancy (semantic N400).

C.5.1.1 Procedure

Nr participants: 18 listeners (2 excluded, 20 completed task)

Nr stimuli per condition: 20 (4 lists of 80 phrases, 4 conditions)

Task: Judge semantic congruency (random left/right assignment)

ISI: 600 ms

C.5.1.2 Stimuli

Manipulation: f_0 exclusively (see section 5.1.2)

Some descriptives:

Table C.12: Durations of target-words: total duration, 1st syllable duration, 1st vowel duration and first vowel mean f_0.

	Sentence *ms*		Target word *ms*		1st syllable *ms*		1st vowel *ms*		1st vowel f_0	
	m	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*	*m*	*sd*
SEMANTICALLY CONGRUENT										
−IA	2097.07	402.81	552.88	96.98	157.23	28.76	72.16	25.9	116.56	11.73
+IA	2092.07	402.81	552.88	96.98	157.23	28.76	72.16	25.9	126.38	12.2
SEMANTICALLY INCONGRUENT										
−IA	2122.59	411.72	583.45	61.72	160.57	32.16	77.86	27.23	123.02	42.18
+IA	2122.59	411.72	583.45	61.72	160.57	32.16	77.86	27.23	140.28	44.26

C.5.1.3 EEG preprocessing

Nr electrodes: 64

Reference: Mastoids

Filter and down-sampling: 0.01 − 30 bandpass, 128Hz

Epoch length: −200 − 1000

C.5.1.4 Analysis

Descriptive statistic behavioral:

Table C.13: Overview reaction times per condition

Condition	mean	sd	max	min	Condition	mean	sd	max	min
mincon	939.09	202.87	1494.07	359.01	plus	951.75	226.89	1564.43	25.01
mininc	1011.12	209.89	1511.31	572.25	min	975.32	207.89	1511.31	359.01
pluscon	915.70	239.43	1561.71	25.01	con	1000.31	207.04	1564.43	572.92
plusinc	989.14	206.93	1564.43	573.43	inc	927.31	222.18	1561.71	25.10

Statistical analysis: Linear Mixed Effects Model in R

Descriptive statistic ERP:

Table C.14: N400-IA — Descriptive statistics of peak amplitude latency variability for metrical N400.

Peak latencies for −IA and congruent

chlabel	Fp1	AF3	F3	FC1	Fpz	Fp2	AF4	Afz	Fz	F4	FC2	FCz
subjects	18	18	18	18	18	18	18	18	18	18	18	18
mean	394.0971	400.6075	402.3437	411.8923	402.3436	399.7394	414.0624	411.4583	418.4027	407.1179	414.0626	407.5521
sd	31.27565	31.68411	28.64224	35.02133	32.95557	33.06412	24.11686	27.19559	24.14998	30.18507	30.78698	33.06437
max	445.312	445.312	453.125	453.125	445.312	445.312	453.125	453.125	453.125	453.125	453.125	453.125
min	351.562	351.562	359.375	351.562	351.562	351.562	367.188	351.562	367.188	351.562	359.375	351.562

Peak latencies for +IA and congruent

chlabel	Fp1	AF3	F3	FC1	Fpz	Fp2	AF4	Afz	Fz	F4	FC2	FCz
subjects	18	18	18	18	18	18	18	18	18	18	18	18
mean	406.2500	417.1007	406.6841	407.5521	412.3263	400.1736	410.1562	405.8159	404.0798	403.6458	409.7222	401.9096
sd	32.04404	27.21387	32.20866	33.17265	28.30255	35.50482	33.28067	33.62649	35.22566	33.78924	35.47113	31.04508
max	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125
min	367.188	367.188	359.375	351.562	367.188	351.562	351.562	359.375	351.562	367.188	351.562	351.562

Peak latencies for −IA and incongruent

chlabel	Fp1	AF3	F3	FC1	Fpz	Fp2	AF4	Afz	Fz	F4	FC2	FCz
subjects	18	18	18	18	18	18	18	18	18	18	18	18
mean	384.5486	394.5312	393.6632	401.0417	387.1528	392.7951	397.5694	396.2673	393.2291	397.1355	399.3055	396.7013
sd	31.31381	26.96347	25.99942	25.42144	26.28562	28.83676	27.31259	28.58645	27.97647	24.30227	29.82612	29.90625
max	445.312	445.312	445.312	445.312	421.875	445.312	445.312	445.312	445.312	437.500	437.500	445.312
min	351.562	351.562	351.562	359.375	351.562	351.562	351.562	351.562	351.562	359.375	351.562	351.562

Peak latencies for +IA and incongruent

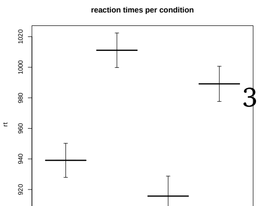
chlabel	Fp1	AF3	F3	FC1	Fpz	Fp2	AF4	Afz	Fz	F4	FC2	FCz
subjects	18	18	18	18	18	18	18	18	18	18	18	18
mean	398.4376	396.7014	394.9653	396.2675	397.5694	398.4375	394.5312	395.3993	407.1182	387.1528	393.6631	397.1354
sd	33.36155	35.20017	33.49285	32.79774	32.69734	34.00117	29.86951	29.37134	29.58442	28.38710	28.50275	33.70955
max	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	429.688	453.125	453.125
min	359.375	351.562	351.562	359.375	351.562	351.562	351.562	359.375	367.188	351.562	351.562	351.562

Table C.15: N400-IA — Descriptive statistics of peak amplitude latency variability for semantic N400.

Peak latencies for −IA and congruent													
chlabel	FC1	C3	CP1	P1	P3	Pz	Fz	FC2	Cz	C4	CP2	P2	P4
subjects	18	18	18	18	18	18	18	18	18	18	18	18	18
mean	535.1563	523.8715	526.4757	530.3820	519.0973	508.2466	513.8889	506.9445	535.5902	532.1182	533.4201	538.6285	541.6667
sd	47.95403	50.73490	45.64399	49.47507	47.12116	47.97059	46.61044	44.99035	53.20466	40.89359	44.82366	40.68076	44.35628
max	601.562	601.562	593.750	601.562	593.750	601.562	601.562	601.562	601.562	593.750	601.562	601.562	601.562
min	468.750	453.125	453.125	453.125	453.125	453.125	453.125	453.125	453.125	468.750	460.938	453.125	453.125
Peak latencies for +IA and congruent													
chlabel	FC1	C3	CP1	P1	P3	Pz	Fz	FC2	Cz	C4	CP2	P2	P4
subjects	18	18	18	18	18	18	18	18	18	18	18	18	18
mean	516.9271	519.0973	532.9861	546.0068	537.3264	526.9098	528.2118	537.3265	532.5521	524.7396	520.3993	540.7985	528.6458
sd	41.62133	41.01048	38.69830	44.83060	43.25457	38.29424	46.80890	46.53343	37.92001	39.13140	44.12419	37.66384	37.80138
max	593.750	593.750	601.562	601.562	593.750	585.938	601.562	601.562	593.750	578.125	593.750	601.562	593.750
min	453.125	460.938	476.562	476.562	453.125	468.750	453.125	453.125	476.562	453.125	453.125	468.750	460.938
Peak latencies for −IA and incongruent													
chlabel	FC1	C3	CP1	P1	P3	Pz	Fz	FC2	Cz	C4	CP2	P2	P4
subjects	18	18	18	18	18	18	18	18	18	18	18	18	18
mean	518.6632	527.3437	516.9272	529.0799	534.7223	516.0590	508.2463	525.1736	518.6632	513.8887	528.6458	526.4757	526.0416
sd	44.12409	40.30640	42.05024	40.26689	38.29404	36.88535	39.24330	36.10655	44.93036	39.97602	46.49028	44.69003	46.72151
max	593.750	578.125	578.125	578.125	601.562	578.125	570.312	593.750	601.562	578.125	601.562	585.938	585.938
min	460.938	453.125	453.125	453.125	468.750	453.125	453.125	468.750	453.125	460.938	460.938	453.125	453.125
Peak latencies for +IA and incongruent													
chlabel	FC1	C3	CP1	P1	P3	Pz	Fz	FC2	Cz	C4	CP2	P2	P4
subjects	18	18	18	18	18	18	18	18	18	18	18	18	18
mean	526.4756	528.6458	535.5903	549.0452	531.2501	525.1737	526.4757	532.9861	533.8541	518.2291	529.9479	541.6667	544.7048
sd	46.73215	44.03134	47.42489	34.60873	41.16571	46.06796	46.80895	45.36120	43.37411	46.79823	37.05780	35.24535	41.23586
max	585.938	585.938	601.562	593.750	585.938	601.562	585.938	585.938	585.938	601.562	578.125	585.938	601.562
min	453.125	460.938	460.938	484.375	453.125	460.938	453.125	453.125	453.125	453.125	460.938	460.938	460.938

Statistical analysis:	$t_{\max}$ mass univariate permutation test, 2500 permutations in Matlab
Electrodes:	11 electrodes for the frontally expected metrical N400 (Fpz, FCz, Fz, AFz, Fp1, Fp2, FC1, FC2, F1, F2, AF3, AF4); 13 electrodes for the centro-parietally expected semantic N400 (Fz, Cz, FC1, FC2, P1, P2, C3, C4, Pz, P3, P4, CP1, CP2)
Time-windows:	$181-281$ for P2; $351-451$ for metrical N400; $450-600$ for semantic N400

C.5.1.5 Results



Behavioral: *Effect semantic congruency*

Participants were slower to respond to semantically incongruent sentences than to semantically congruent sentences.

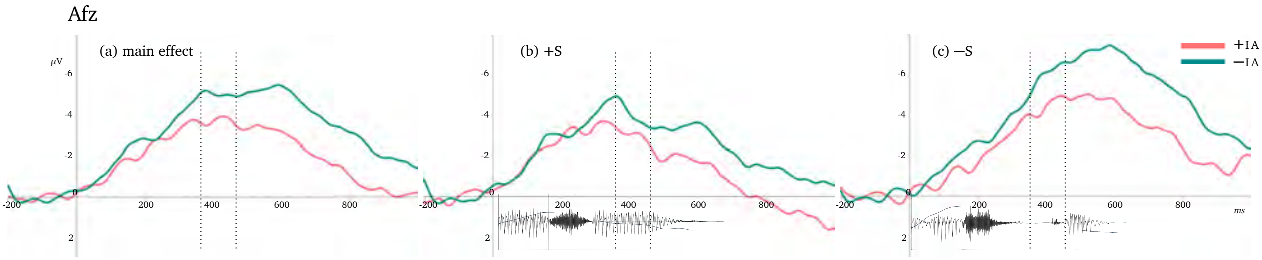
Effect ±IA

Participants were slower to respond to target-word −IA than target-words +IA. The plot indicates an interaction between IA and semantic congruency which is not significant.

P200: No differences.

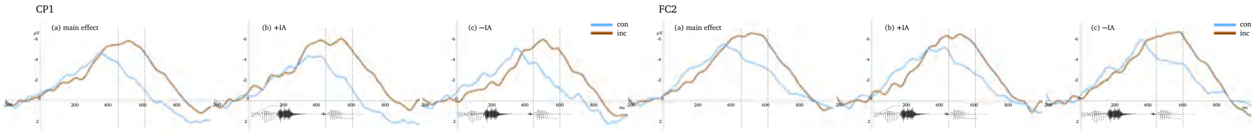
Metrical n400: *Effect of initial accent; main effect* (t_{max} $p = 0.044$ at Afz at 375ms).

Stimuli (irrespective of congruency) without initial accent resulted in a larger negativity compared to stimuli with initial accent in the anterior-frontal region . The effect of ±IA did not reach significance within the congruent or incongruent conditions (respectively; t_{max} $p = 0.18$ and $p = 0.16$). I actually would not have expected semantic congruency to affect metrical processing; whether the sentence makes sense or not, listeners should prefer words to be marked with IA. Also, IA extraction may be completed before the participants is able to judge semantic congruency. As such I'm not even sure if I should have tested within the semantic conditions and it suits me fine to only have a main effect.



Semantic n400: *Effect of semantic congruency; main effect* (t_{max} $p = 0.0 - 0.048$ at CP1 and FC2 between $492 - 593$).

The centro-parietal and fronto-central electrodes (CP1 and FC2) are significantly more negative for semantically incongruent sentences than for semantically congruent sentences at in the semantic N400 time-window. This effect is also significant within the condition without IA (t_{max} $p = 0.002 - 0.048$) and marginally significant within the condition with IA (t_{max} $p = 0.08$).



Interaction: *Interaction metrical ∗ semantic* (t_{max} $p = 0.0 - 0.048$ at Af4, Afz, CP1 and FC2 between 523 − 593).

There was an interaction effect between our two manipulations at centro-parietal and frontal electrodes which I can't yet interpret. I thought it could be a latency effect, such that conflict resolution starts later for incongruent word without initial accent. Because I have not yet found out how to test latency difference with the t_{max} analysis, I used a regression analysis. Dependent variable was peak latency, fixed effects were ±IA, semantic congruency and electrode cite (parietal, centro-parietal and central) and random effect was subject. It was not significant at $p = 0.11$.

C.5.1.6 Conclusion

There was a fronto-central main effect of IA, even when IA is embedded in a sentence and not phrase initial. Additionally we found a centro-parietal N400 effect of semantic congruency which interacted with IA such that semantic conflict resolution started later and after phonological repair when words were presented without IA.

C.5.2 Behavioral analyses

C.5.2.1 Descriptives

Table C.16: Overview reaction times per condition

Condition	mean	sd	max	min	Condition	mean	sd	max	min
mincon	939.09	202.87	1494.07	359.01	plus	951.75	226.89	1564.43	25.01
mininc	1011.12	209.89	1511.31	572.25	min	975.32	207.89	1511.31	359.01
pluscon	915.70	239.43	1561.71	25.01	con	1000.31	207.04	1564.43	572.92
plusinc	989.14	206.93	1564.43	573.43	inc	927.31	222.18	1561.71	25.10

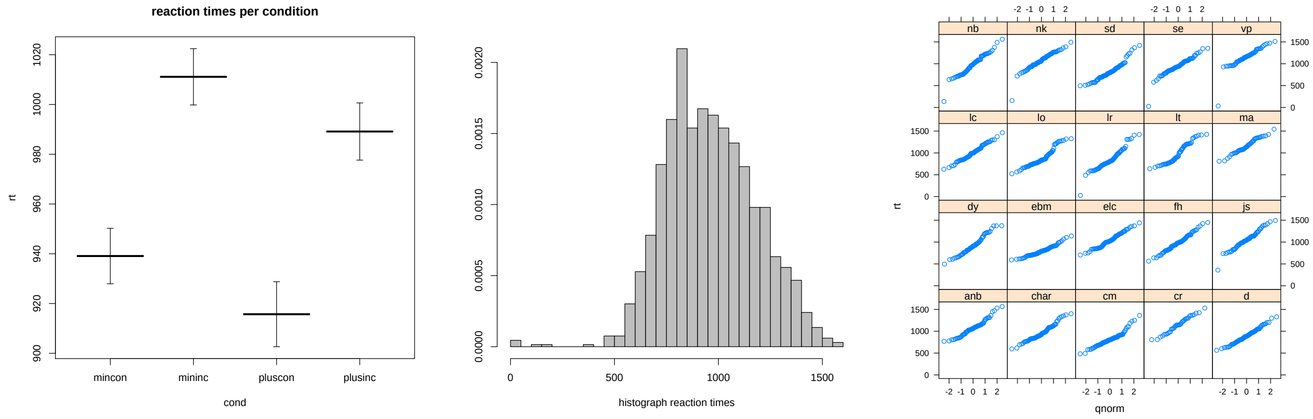


Figure C.32: Elaborate caption because there's quite a lot to say about these figures.

C.5.2.2 Regression models

Table C.17: Overview linear mixed models. The model fitting the data best takes semantic congruency as fixed factor and subjects and stimuli variability as random factors. Presence of initial accent significantly contributes to the prediction of reaction times when it is the only fixed effect and marginally contributes when entered together with semantic congruency.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	976.30***	977.21***	1016.63***	966.57***	1006.12***	1008.88***
	(25.07)	(26.89)	(28.18)	(27.24)	(28.52)	(28.91)
congruency			−78.46***		−78.27***	−83.64***
			(16.81)		(16.84)	(19.14)
±IA				21.00*	20.58*	15.20
				(9.37)	(9.31)	(13.03)
±IA:congruency						10.66
						(18.04)
AIC	17757.83	17567.17	17542.20	17557.93	17533.11	17527.13
BIC	17773.40	17603.50	17583.72	17599.45	17579.82	17579.03
Log Likelihood	−8875.92	−8776.58	−8763.10	−8770.96	−8757.55	−8753.57
Num. obs.	1326	1326	1326	1326	1326	1326
Num. groups: subj	20	20	20	20	20	20
Var: subj (Intercept)	11996.42	12415.76	12432.85	12339.64	12363.26	12377.19
Var: Residual	36671.85	25898.44	25946.90	25905.44	25948.81	25959.58
Num. groups: congr:stimuli		160	160	160	160	160
Num. groups: ±IA:stimuli		160	160	160	160	160
Num. groups: stimuli		80	80	80	80	80
Var: congr.stimuli (Intercept)		11031.88	7977.87	11057.78	8021.99	8016.48
Var: ±IA.stimuli (Intercept)		417.00	366.70	234.50	193.37	202.53
Var: stimuli (Intercept)		267.65	2153.78	591.00	336.86	1534.69

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table C.18: ANOVA: initial accent and semantic congruency as significant predictors for reaction times

ANOVA	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
N400—intercept	3	1.8e+04	1.8e+04	-8.9e+03	1.8e+04			
N400—random structure	7	1.8e+04	1.8e+04	-8.8e+03	1.8e+04	2.0e+02	4	$7.8e-42^{***}$
N400— ±IA	8	1.8e+04	1.8e+04	-8.8e+03	1.8e+04	4.9	1	$2.6e-02^{*}$
N400— ±S	8	1.8e+04	1.8e+04	-8.8e+03	1.8e+04	1.5e+01	0	0.0^{***}
N400—interaction	1e+01	1.8e+04	1.8e+04	-8.8e+03	1.8e+04	3.5e-01	1	$5.6e-01$

C.5.3 Non-parametric results

Separate repeated measures, two-tailed permutation tests based on the t-max statistic were conducted in Matlab for each of the time windows described above and each of the comparisons of interest, i.e.:

Time-windows:

- P2 → 181 − 281 ms (This is different from the previous studies for no reason. I misremembered. I could/should re-analyse with the correct time-window, it's unlikely to make a difference.)
- Metrical N400 → 351 − 451 ms
- Semantic N400 → 450 − 600 ms

Comparisons:

- main effect ±IA
- main effect semantic congruency
- ±IA ∗ semantic congruency (4 comparisons)

I expected the metrical N400 frontally and tested 12 electrodes: Fpz, FCz, Fz, AFz, Fp1, Fp2, FC1, FC2, F1, F2, AF3, AF4. The semantic N400 is expected centro-parietally, I tested 13 electrodes: Fz, Cz, FC1, FC2, P1, P2, C3, C4, Pz, P3, P4, CP1, CP2. The data was down-sampled to 128 Hz to reduce the number of comparisons. Also, 2500 permutations were used to estimate the distribution of the null hypothesis for the customary family-wise α level of 0.05.

C.5.3.1 P200

Effect initial accent; main effect

```
GND=tmaxGND(GND,9 , 'time_wind ',[181  281], 'include_chans ',
{ 'Fpz ' , 'Fz ' , 'F1 ' , 'FC1 ' , 'FC2 ' , 'Fp1 ' , 'Fp2 ' , 'F2 ' , 'AFz ' , 'AF3 ' , 'AF4 ' , 'FCz ' }) ;
```

Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 9 (minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	14
Nr of comparisons	168
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.9836 and 3.9836
Test-wise alpha level	0.000961.
Bonferroni test-wise alpha	0.000325.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.420400$

Effect semantic congruency; main effect

Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 10 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	14
Nr of comparisons	182
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0826 and 4.0826
Test-wise alpha level	0.000775.
Bonferroni test-wise alpha	0.000275.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.586000$

Effect initial accent; condition → congruent

Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 11 (con:minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	14
Nr of comparisons	168
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0172 and 4.0172
Test-wise alpha level	0.000893.
Bonferroni test-wise alpha	0.000325.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.715200$

Effect initial accent; condition → incongruent

Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 12 (inc: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	14
Nr of comparisons	168
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0625 and 4.0625
Test-wise alpha level	0.000810.
Bonferroni test-wise alpha	0.000325.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.125600$

Effect semantic congruency; condition → min initial accent

Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 13 (minai : inc - con) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	14
Nr of comparisons	182
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.2545 and 4.2545
Test-wise alpha level	0.000535.
Bonferroni test-wise alpha	0.000275.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.642400$

Effect semantic congruency; condition → with initial accent

Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 14 (plusai : inc - con) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	14
Nr of comparisons	182
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.1139 and 4.1139
Test-wise alpha level	0.000725.
Bonferroni test-wise alpha	0.000275.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed.
All $p - values >= 0.558400$

Interaction IA *and semantic congruency*
Attempting to use time boundaries of 181 to 281 ms for hypothesis test. Exact window boundaries are 1.796875e+02 to 2.812500e+02 ms (that's from time point 50 to 63). Testing null hypothesis that the grand average ERPs in Bin 15 (interaction) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	20
Nr of time points	14
Nr of comparisons	280
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.213 and 4.213
Test-wise alpha level	0.000585.
Bonferroni test-wise alpha	0.000179.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed.
All $p - values >= 0.239600$

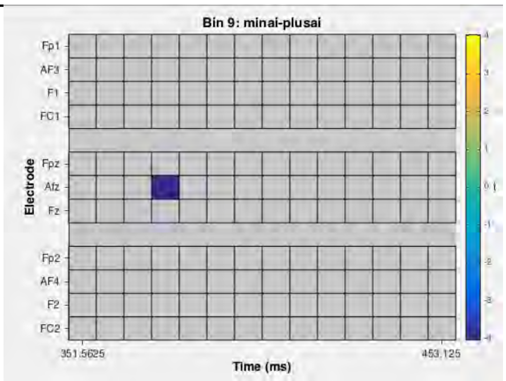
C.5.3.2 Metrical N400

Effect initial accent; main effect

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 9 (minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	14
Nr of comparisons	168
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.9078 and 3.9078
Test-wise alpha level	0.001132.
Bonferroni test-wise alpha	0.000325.

Significant differences from zero: 375 ms, electrode(s):
Afz
All significant corrected $p-values$ are between 0.044000 and 0.044000



Effect semantic congruency; main effect

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 10 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	14
Nr of comparisons	182
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0628 and 4.0628
Test-wise alpha level	0.000809.
Bonferroni test-wise alpha	0.000275.
ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.138400$	

Effect initial accent; condition → congruent

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 11 (con:minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	14
Nr of comparisons	168
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.859 and 3.859
Test-wise alpha level	0.001259.
Bonferroni test-wise alpha	0.000325.
ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.180800$	

Effect initial accent; condition → incongruent

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 12 (inc: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	14
Nr of comparisons	168
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0583 and 4.0583
Test-wise alpha level	0.000817.
Bonferroni test-wise alpha	0.000325.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.159200$

Effect semantic congruency; condition → min initial accent

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 13 (minai : inc - con) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	14
Nr of comparisons	182
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0631 and 4.0631
Test-wise alpha level	0.000809.
Bonferroni test-wise alpha	0.000275.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p - values >= 0.111600$

Effect semantic congruency; condition → with initial accent

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 14 (plusai : inc - con) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	14
Nr of comparisons	182
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.1018 and 4.1018
Test-wise alpha level	0.000744.
Bonferroni test-wise alpha	0.000275.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p-values >= 0.739200$

Interaction IA *and semantic congruency*

Attempting to use time boundaries of 351 to 451 ms for hypothesis test. Exact window boundaries are 3.515625e+02 to 4.531250e+02 ms (that's from time point 72 to 85). Testing null hypothesis that the grand average ERPs in Bin 15 (interaction) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	20
Nr of time points	14
Nr of comparisons	280
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.2396 and 4.2396
Test-wise alpha level	0.000552.
Bonferroni test-wise alpha	0.000179.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p-values >= 0.147200$

C.5.3.3 Semantic N400

Effect initial accent; main effect

Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 9 (minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	20
Nr of comparisons	220
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-3.8119 and 3.8119
Test-wise alpha level	0.001394.
Bonferroni test-wise alpha	0.000227.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p-values >= 0.269600$

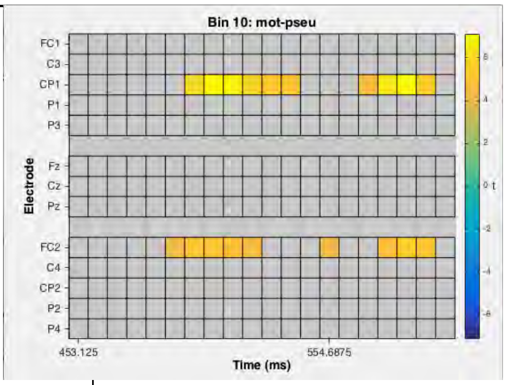
Effect semantic congruency; main effect

Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 10 (mot-pseu) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	20
Nr of comparisons	260
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.1627 and 4.1627
Test-wise alpha level	0.000652.
Bonferroni test-wise alpha	0.000192.

Significant differences from zero: 492-593 ms, electrode(s): CP1, FC2

All significant corrected $p-values$ are between 0.048000 and 0.000000



Effect initial accent; condition → congruent

Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 11 (con:minai-plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	12
Nr of time points	20
Nr of comparisons	220
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0575 and 4.0575
Test-wise alpha level	0.000819.
Bonferroni test-wise alpha	0.000227.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p-values$ >= 0.326000

Effect initial accent; condition → incongruent

Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 12 (inc: minai - plusai) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

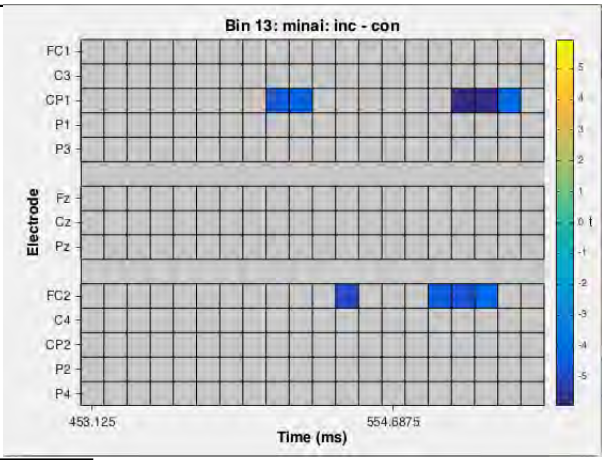
Nr of channels	12
Nr of time points	20
Nr of comparisons	220
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050400
Critical t-score(s)	-3.925 and 3.925
Test-wise alpha level	0.001091.
Bonferroni test-wise alpha	0.000227.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed. All $p-values$ >= 0.193600

Effect semantic congruency; condition → min initial accent

Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 13 (minai : inc - con) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	13
Nr of time points	20
Nr of comparisons	260
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.1038 and 4.1038
Test-wise alpha level	0.000741.
Bonferroni test-wise alpha	0.000192.



Significant differences from zero: 523-593 ms, electrode(s): CP1, FC2
All significant corrected $p-values$ are between 0.048000 and 0.002000

Effect semantic congruency; condition → with initial accent
Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 14 (plusai : inc - con) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

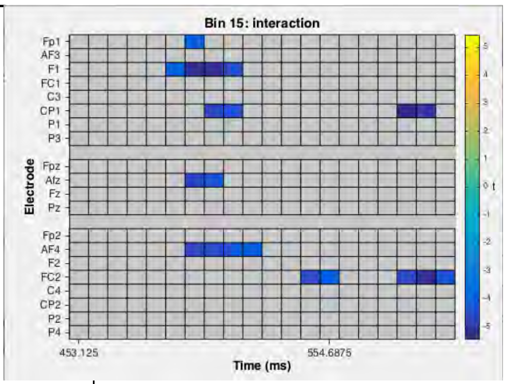
Nr of channels	13
Nr of time points	20
Nr of comparisons	260
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.1861 and 4.1861
Test-wise alpha level	0.000620.
Bonferroni test-wise alpha	0.000192.

ERPs are NOT significantly different from zero ($alpha = 0.05$) at any time point/window analyzed.
All $p-values >= 0.086000$

Interaction IA *and semantic congruency*

Attempting to use time boundaries of 450 to 600 ms for hypothesis test. Exact window boundaries are 4.531250e+02 to 6.015625e+02 ms (that's from time point 85 to 104). Testing null hypothesis that the grand average ERPs in Bin 15 (interaction) have a mean of 0.000000 microvolts. Alternative hypothesis is that the ERPs differ from 0.000000 (i.e., two-tailed test).

Nr of channels	20
Nr of time points	20
Nr of comparisons	400
Nr of participants	18
t-score degrees of freedom	17
Nr permutations	2500
Estimated actual family-wise alpha level	0.050000
Critical t-score(s)	-4.0893 and 4.0893
Test-wise alpha level	0.000764.
Bonferroni test-wise alpha	0.000125.

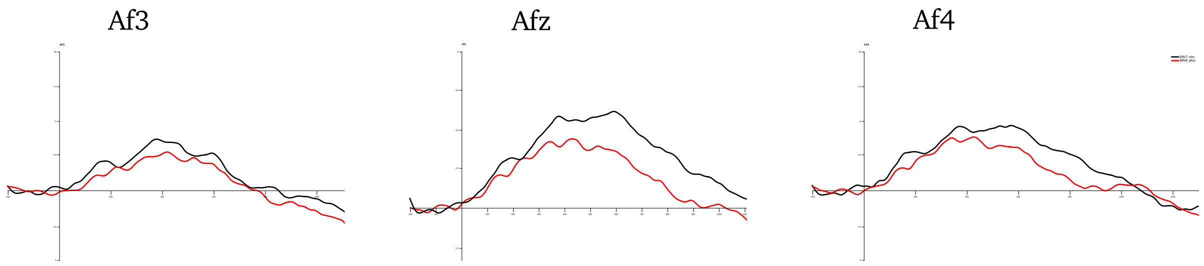


Significant differences from zero: 523-593 ms, electrode(s): CP1, FC2, Afz, AF4

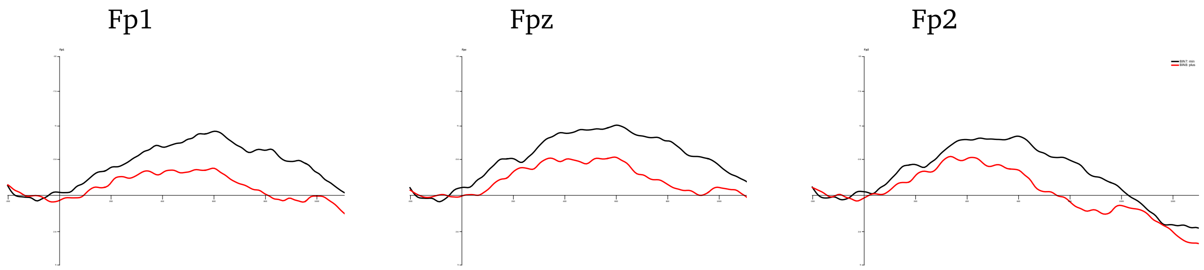
All significant corrected $p - values$ are between 0.046400 and 0.006800

C.5.4 All erp plots

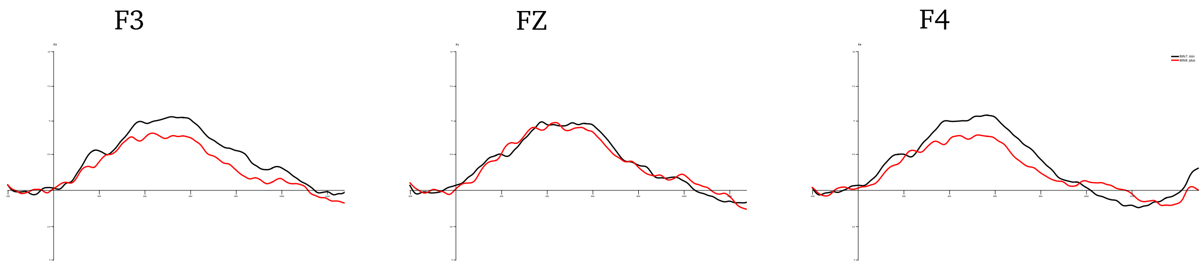
C.5.4.1 Effect initial accent; main effect



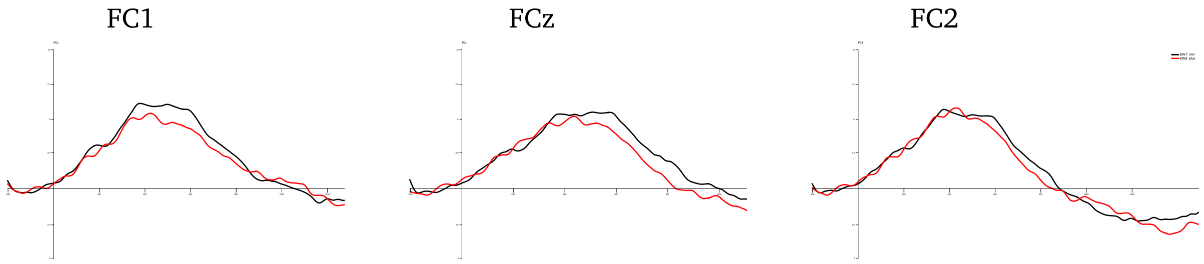
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

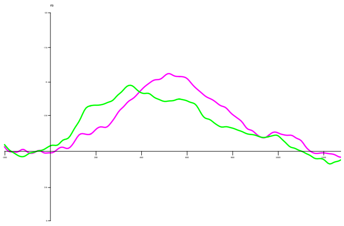


red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

Figure C.33: N400-IA — ERP's main effect initial accent: all plus ia sentences versus all min ia sentences.

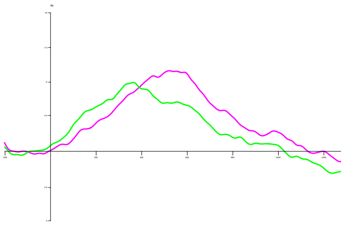
C.5.4.2 Effect semantic congruency; main effect

F3



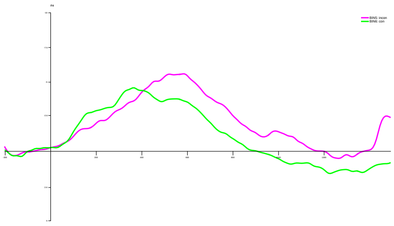
green:conguent magenta:incongruent

FZ



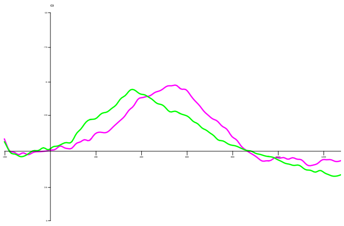
green:conguent magenta:incongruent

F4



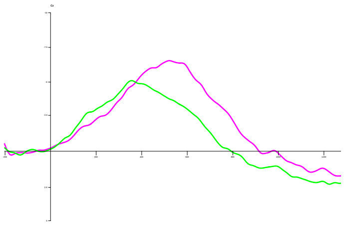
green:conguent magenta:incongruent

C3



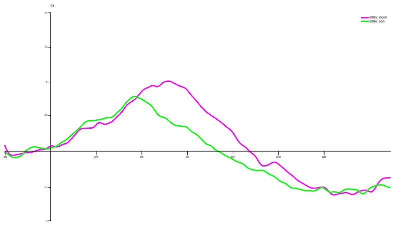
green:conguent magenta:incongruent

CZ



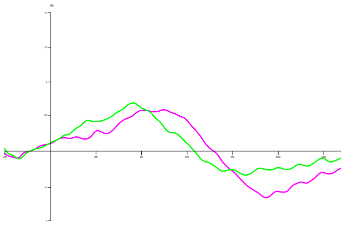
green:conguent magenta:incongruent

C4



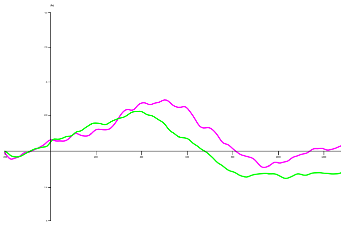
green:conguent magenta:incongruent

P3



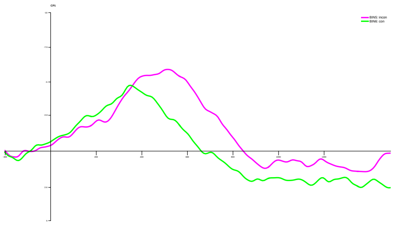
green:conguent magenta:incongruent

P4



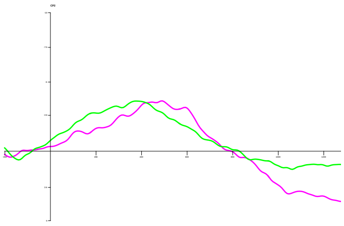
green:conguent magenta:incongruent

CP1



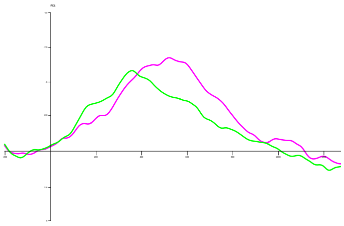
green:conguent magenta:incongruent

CP2



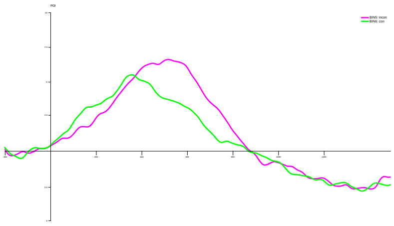
green:conguent magenta:incongruent

FC1



green:conguent magenta:incongruent

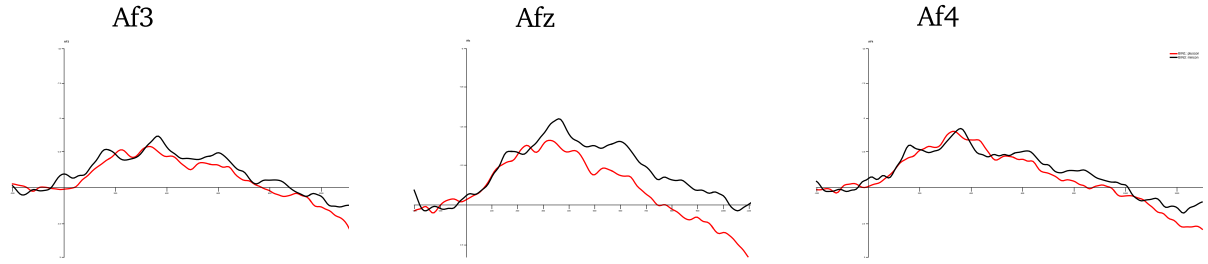
FC2



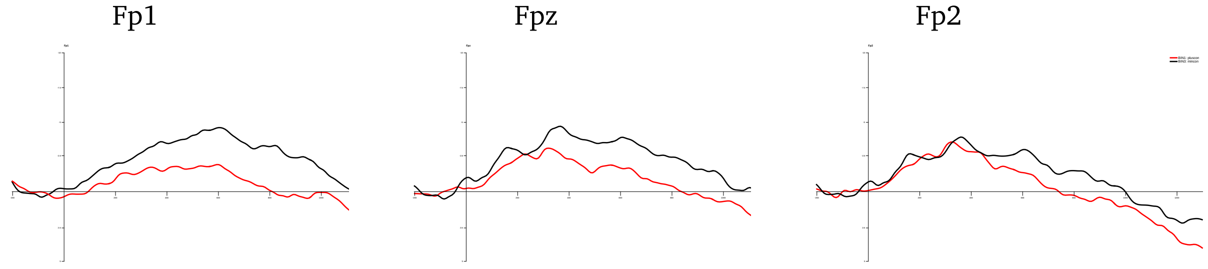
green:conguent magenta:incongruent

Figure C.34: N400-IA — ERP's main effect semantic congruency: so all congruent sentences versus all incongruent sentences.

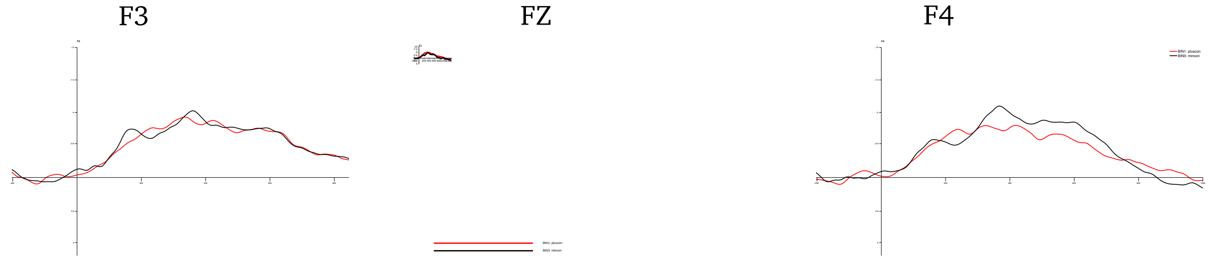
C.5.4.3 Effect initial accent; condition → semantically conguent



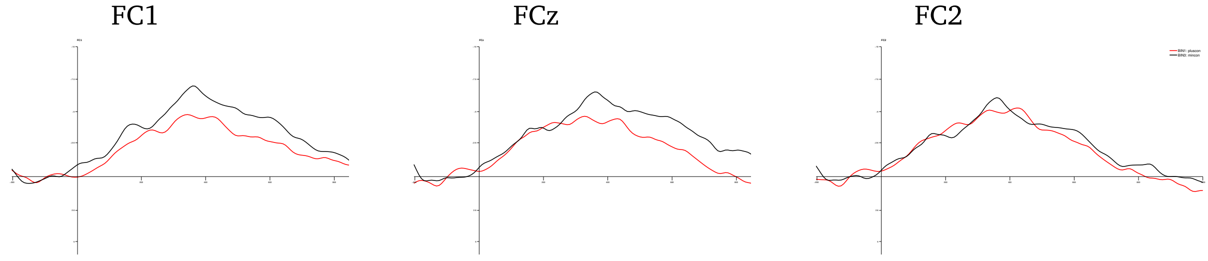
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



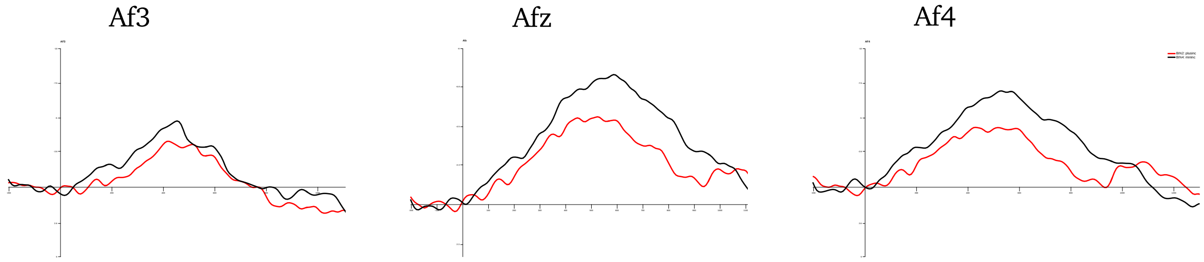
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



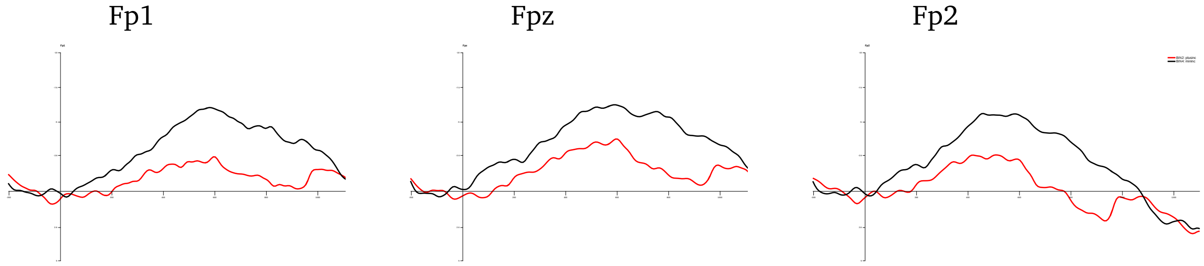
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

Figure C.35: N400-IA — ERP's effect initial accent but ONLY for semantically congruent sentences.

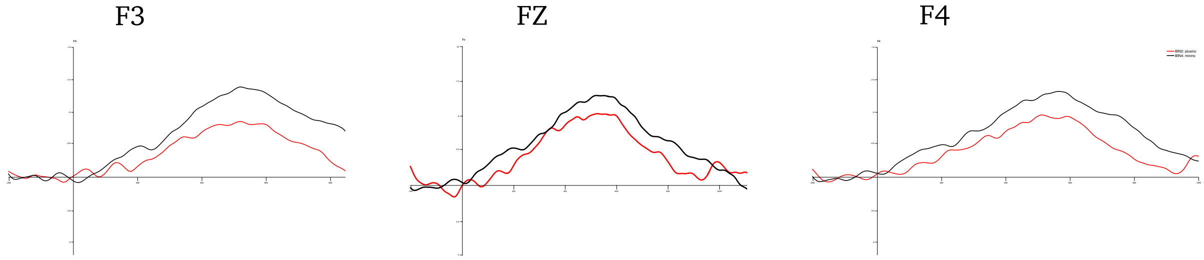
C.5.4.4 Effect initial accent; condition → semantically incongruent



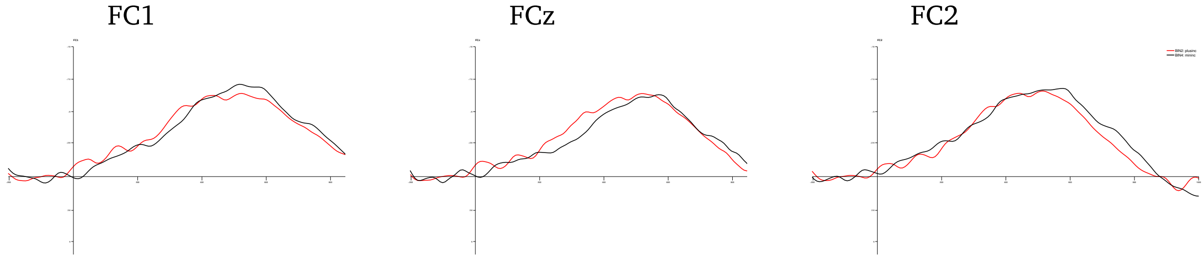
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



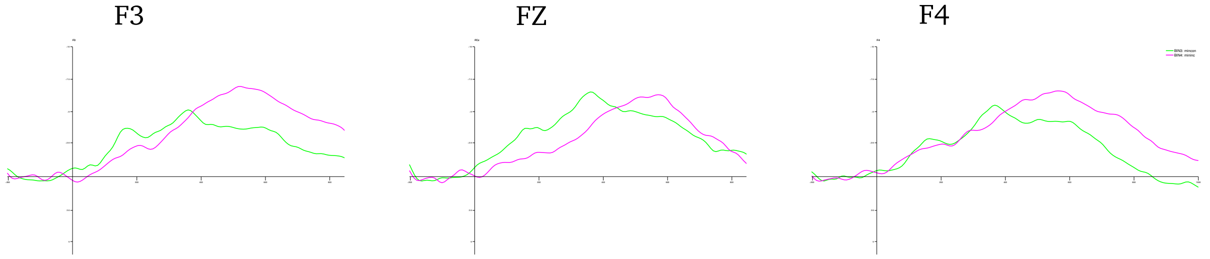
red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia



red:plus ia black:min ia red:plus ia black:min ia red:plus ia black:min ia

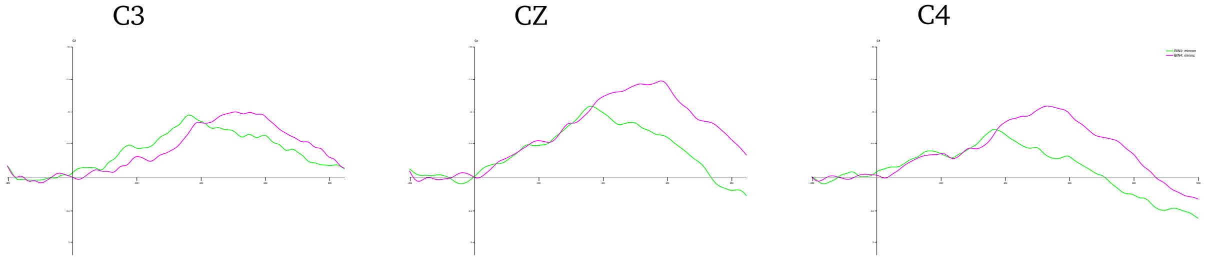
Figure C.36: N400-IA — ERP's effect initial accent but ONLY for semantically incongruent sentences.

C.5.4.5 Effect semantic congruency; condition → min initial accent



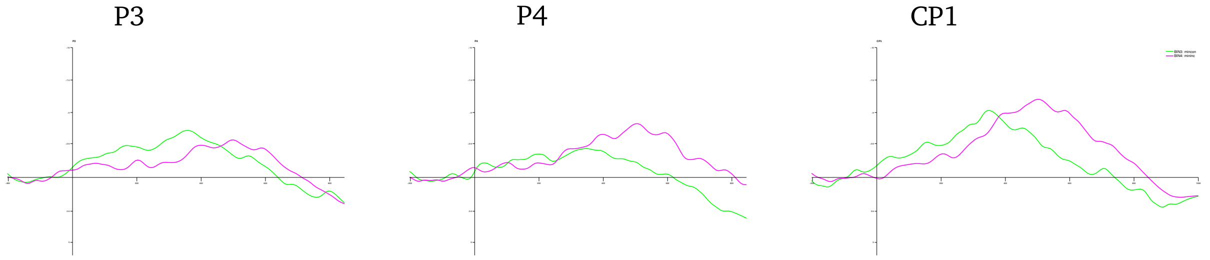
F3 FZ F4

green:congruent magenta:incongruent green:congruent magenta:incongruent green:congruent magenta:incongruent



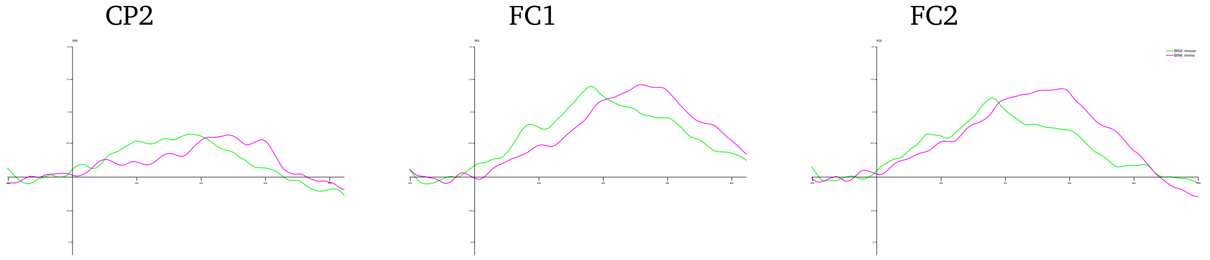
C3 CZ C4

green:congruent magenta:incongruent green:congruent magenta:incongruent green:congruent magenta:incongruent



P3 P4 CP1

green:congruent magenta:incongruent green:congruent magenta:incongruent green:congruent magenta:incongruent

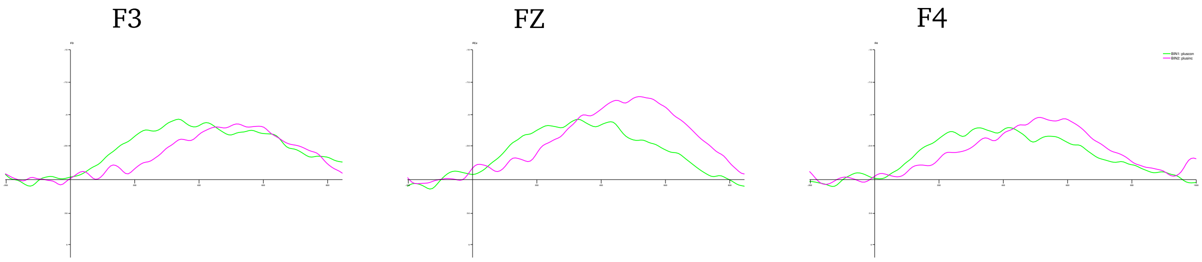


CP2 FC1 FC2

green:congruent magenta:incongruent green:congruent magenta:incongruent green:congruent magenta:incongruent

Figure C.37: N400-IA — ERP's effect semantic congruency but ONLY for items MINUS initial accent

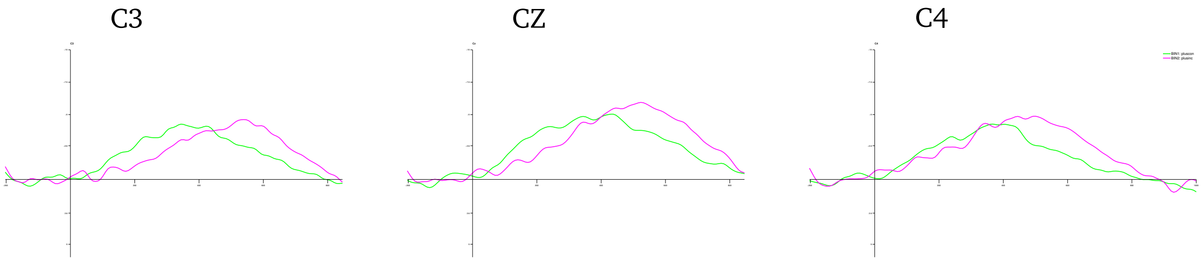
C.5.4.6 Effect semantic congruency; condition → with initial accent



green:congruent magenta:incongruent

green:congruent magenta:incongruent

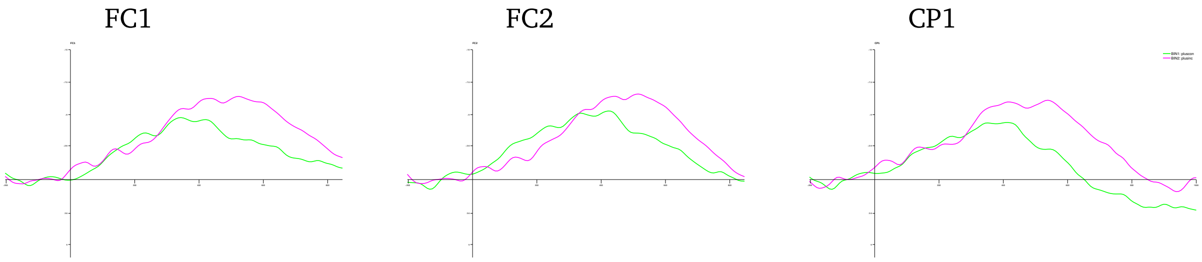
green:congruent magenta:incongruent



green:congruent magenta:incongruent

green:congruent magenta:incongruent

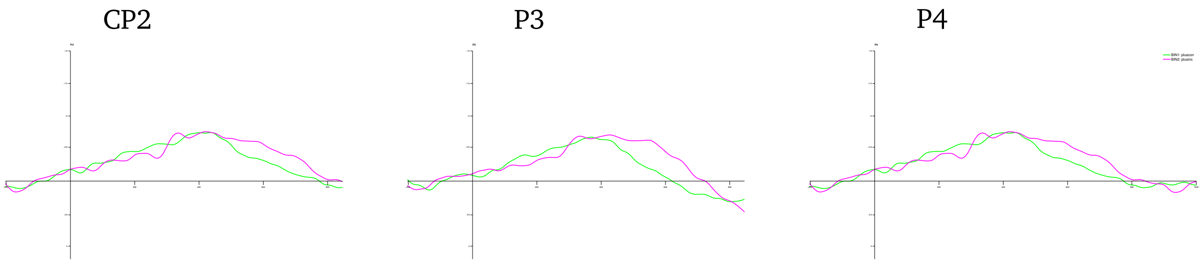
green:congruent magenta:incongruent



green:congruent magenta:incongruent

green:congruent magenta:incongruent

green:congruent magenta:incongruent



green:congruent magenta:incongruent

green:congruent magenta:incongruent

green:congruent magenta:incongruent

Figure C.38: N400-IA — ERP's effect semantic congruency but ONLY for items PLUS initial accent

D Conference papers

Investigating the phonological status of the Initial Accent in French: an Event-Related Potentials study

Noémie te Rietmolen[1], *Radouane El Yagoubi*[2], *Robert Espesser*[3],
Cynthia Magnen[4], *Corine Astésano*[1]

[1]U.R.I Octogone-Lordat (E.A. 4156), Université de Toulouse, UTM, Toulouse, France
[2]Laboratoire CLLE-LTC (UMR 5263), Université de Toulouse, UTM, Toulouse, France
[3]Laboratoire Parole & Langage (UMR 7309), Aix-Marseille Université, Aix-en-Provence, France
[4]MSHS-T (USR 3414), Université de Toulouse, UTM, Toulouse, France

`noemie.te-rietmolen@univ-tlse2.fr, yagoubi@univ-tlse2.fr, Robert.Espesser@lpl-aix.fr,`
`cynthia.magnen@univ-tlse2.fr, astesano@univ-tlse2.fr`

Abstract

This Event-Related Potentials (ERP) study investigates the use of prosodic information in the process of lexical access in French. In French, accentuation is said to be post-lexical, with a primary final accent (FA) and secondary initial accent (IA) marking the edges of the phrase. Results from previous studies, however, suggest IA may hold a demarcative function close to the level of the word. Still, the contribution of IA in word processing has not yet been empirically tested. In this study, participants listened to trisyllabic French nouns and pseudowords, with (+IA) or without (−IA) initial accent while completing a lexical decision task. We were mainly interested in modulations of the N325, a component assumed to reflect difficulties in the extraction of lexical stress patterns. ERP results show a larger N325 when stimuli were presented −IA, revealing both the automaticity of stress extraction and a preference for stress templates with initial accent.

Index Terms: stress, French, Initial Accent, lexical decision, Event-Related Potentials, N325

1. Introduction

The ability to understand spoken language is a fundamental and intriguing human skill. Considering speech is formed out of connected and co-articulated units with no spaces or other breaks, the manner in which we translate its signal into a sequence of words is far from obvious. One source that may help speech segmentation comes from the metrical structure of the signal. According to Metrical Segmentation Strategy (MSS), the segmentation of continuous speech is accomplished by relying on the dominant metrical pattern of the language [1]. In stress-based languages such as English and Dutch, where the vast majority of lexical words start with a strong syllable [2, 3], listeners are thought to exploit that high prosodic probability and initiate lexical access at each stressed syllable. But, while this may be a successful strategy in languages with lexical stress, segmenting on strong onsets is arguably much less efficient in languages in which the domain for metrical rules is not the lexical word.

French is often described as a syllable-based language with fairly homogeneous metrical weight on syllables. Consequently, it is held that the French metrical structure is defined by the syllable and that the syllable is used as the basic unit for segmenting speech [4, 5, 6]. This idea is further supported by the view that, in French, accentuation is post-lexical, demarcating boundaries not at the level of the word but at the level of groups of words. That is, the primary French accent, known as the final accent (FA), is fixed on the last syllable of the phrase, marking its right edge. When necessary (e.g. in case of long stretches of unaccented syllables), FA can be accompanied with a secondary initial accent (IA) that marks the left edge of the phrase [7]. So, French is considered a language without lexical stress, making accents unlikely candidates to cue lexical access.

In contrast with this view, Di Cristo's metrical model considers both FA and IA to be phonologically represented at the level of the prosodic word (i.e. close to the lexical word [8]) despite accentuation not being lexically distinctive in French [9]. According to this model, French accentuation thus provides not one, but two entries; at the left boundary and at the right boundary of the word. A number of studies investigating the use of French prosodic cues in word processing report results in line with Di Cristo's conjecture of (latent) stress templates underlying the representation of the prosodic word. Both the primary FA *and* the secondary IA have been found to guide French listeners in the segmentation of speech (for use of FA see [10, 11, 12]; for use of IA see [13, 14, 15]). These studies challenge the idea that French listeners adopt a syllable-based segmentation strategy (as proposed by [16]). They instead favor a strategy in which listeners rely on metrical stress patterns during speech comprehension. They, however, do not challenge the view that IA and FA demarcate phrase boundaries, and still consider accentuation to apply to the level of the Accentual Phrase (AP; [17]) and not to the level of the prosodic word. Assuming Di Cristo's view gives new perspectives on the speech segmentation strategy in French. Indeed, if French accentuation is actually a stress template encoded at the level close to the lexical word, IA and FA could readily notify listeners on when to initiate lexical access.

Here, we further investigate the representation of French accentuation and its contribution to word processing. More specifically, the status of the initial accent and its role in lexical access is examined. Because this accent is traditionally regarded as a secondary and optional accent, only complementary to the final accent, up until recently IA has received relatively little scientific attention. However, it been shown that, similar to FA, IA not only directs listeners in the segmentation of speech [13, 14, 15], but also that IA is a more reliable cue in the marking of lexical structure than FA [18]. In the study, IA was shown to mark lower levels of structure, close to the lexical word. A later study on the perception of prominences indicated that IA is perceived as stronger than FA, in a manner independent from the depth of prosodic structure. This points towards an association between IA and word demarcation [19].

Following up on these results, a recent event-related potentials (ERP) study was carried out that lends further support to the notion of a phonological representation of IA [20]. In the study, Aguilera et al. investigated the phonological status of IA, using an oddball paradigm in which the presence of IA was manipulated on trisyllabic words. When presenting the oddball without IA, a clear MisMatch Negativity component (MMN) emerged. But, when the oddball was presented with IA, the resultant MMN was significantly smaller. The authors took this to indicate that IA is represented at the phonological long-term representation of the word and part of the French preferred stress template.

In the present ERP study, we sought to build on the MMN study and manipulated the presence of IA in a lexical decision task. We were particularly interested in modulations of the N325, a component assumed to reflect difficulties in the extraction of lexical stress templates [21]. The component was first encountered in a study in which the authors presented Dutch participants with a stress discrimination task. In the task, sequences of four bisyllabic words were presented with stress on the first syllable (the dominant metrical pattern in Dutch) or on the second syllable. Results showed that the less frequent stress template elicited a larger frontal negativity (the N325) than did the dominant stress template. This led the authors to conclude that the N325 may reflect the extraction of metrical stress during lexical access. If IA is linked to the phonological representation of prosodic words and is, along with FA, the expected stress template in French, presenting words without IA should elicit a larger N325 than presenting words with IA. To further attest for the pre-lexicality of stress encoding in French and question whether IA is part of the preferred metrical template, we test this metrical pattern on words and pseudowords.

2. Methods

2.1. Speech stimuli

The stimuli consisted of 120 trisyllabic French nouns (e.g. *chocolat*) and 120 trisyllabic pseudowords (e.g. *chibute*). The stimuli were extracted from sentences spoken by a naïve native speaker of French. In the sentences, the target words (lexical word or pseudoword) were placed at the beginning of a major phrase to increase the probability of clear IA and FA marking [18]. Stimuli with the most natural IA (+IA) were selected by a panel of three experts and re-synthesized without IA (−IA) using a customized quadratic algorithm in PRAAT [22].

Using the same algorithm as [20], the $f0$ value of the first vowel (i.e. IA) was lowered near the $f0$ value of the preceding (unaccented) determinant, to deaccentuate the first syllable (i.e. remove IA; see Figure 1). The algorithm progressively modified the $f0$ values to reach the $f0$ value at the beginning of the last (accented) vowel. This quadratic transformation allowed for micro-prosodic variations to be maintained, thus keeping the natural sound of the stimuli. The +IA stimuli were forward and back transformed to equalize the speech quality between +IA and −IA stimuli. The duration of the target words was held constant in both stress conditions (+IA; −IA), since only the $f0$ parameter was manipulated (lexical words $m = 813$, $sd = 81$; pseudowords $m = 844$, $sd = 83$).

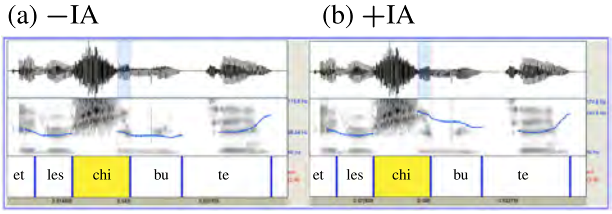


Figure 1: *Example of $f0$ resynthesis (a) −IA and (b) +IA on* '[et les] chibute'*, with quadratic interpolation from the $f0$ value of the preceding determinant to the $f0$ value at the beginning of the last stressed syllable for −IA targets.*

2.2. Participants

26 French native speakers, aged $19 - 31$ (mean age 25.4; 20 females), took part in the study. All subjects were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Due to excessive artifacts in the EEG signal, 3 participants were excluded from analyses.

2.3. Procedure

Each participant was comfortably seated in an electrically shielded and sound attenuated room. Stimuli were presented through headphones and participants were allowed to adjust the volume to their individual preferences.

Participants were instructed to judge as quickly and accurately as possible whether a word was a real word or a pseudoword by pressing the left or right button on a button-box (button assignment was counter-balanced across participants). To ensure participants understood the task requirements, the experiment began with a short practice phase. This phase consisted of 12 trials that were very similar to the experimental trials, but were not included in the analyses.

Each participant listened to all 240 stimuli. Using Latin square designs, the four conditions (word +IA, word −IA, pseudoword +IA, pseudoword −IA) were evenly distributed over blocks, and block order was balanced between participants. In order to better control for eye-related EEG activity, each trial started with a 400ms presentation of a white fixation cross at the center of a computer screen. The stimulus was presented immediately after the offset of the fixation cross. Participants were given a maximum of 2000 ms to give their answer. The intertrial interval (ITI) followed the participants' response and lasted until 2500 ms post stimulus onset. As a result, the duration of ITI varied, while trial duration was fixed at 2900 ms. Total duration of the experiment, including the set-up of the EEG electrodes, was approximately 2h.

2.4. EEG recording and preprocessing

The EEG data were recorded with 32 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958) at Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, F3, AF4, F4, C3, C4, P3, P4, PO3, PO4, P5, P6, O1, O2, F7, F8, T3, T4, T5, T6, FC5, FC6, CP1, CP2, CP5 and CP6. To detect blinks and eye-movements, 4 additional electrodes were placed around the eyes (HEOG: bipolar channel placed lateral to the outer corner of both eyes; VEOG: bipolar channel placed above and below the left eye). The EEG and EOG signals were amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 512 Hz.

The data were preprocessed using the EEGLAB package [23] in Matlab [24]. Each electrode was re-referenced offline to

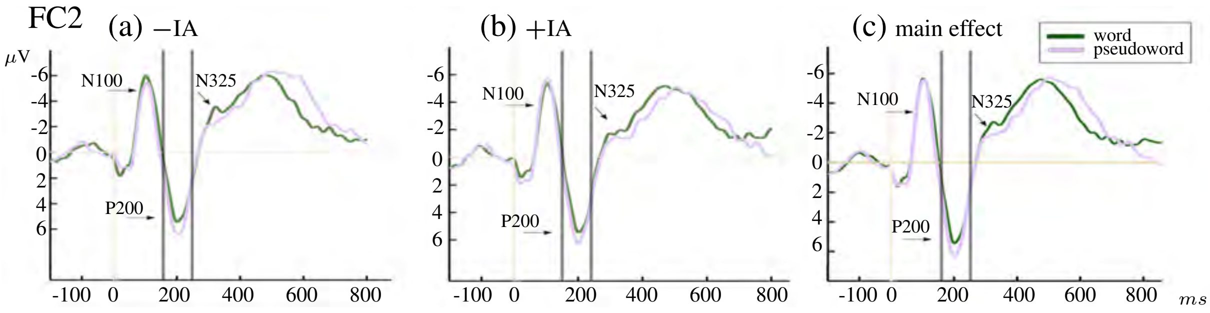


Figure 2: *Grand average P200 in the lexical condition (word, pseudoword), recorded at the FC2 (frontocentral) electrode for: (a) —IA, (b) +IA, (c) main effect. The vertical gray bars indicate the selected P200 time window* ($151 - 251$ *ms). For ease of presentation, ERP waveforms are cut off at 800 ms. Negativity is plotted as an upward deflection.*

the algebraic average of the left and right mastoids. The data were band-pass filtered between 0.01-30 Hz and epoched from -0.2 to 2 seconds surrounding the onset of the speech signal. Following a visual inspection, epochs containing EMG or other artifacts not related to eye-movements or blinks were manually removed. Independent Components Analysis (ICA) was performed on the remaining epochs in order to identify and subtract components containing oculomotor artifacts from the data. Finally, data were averaged within and across participants to obtain the grand-averages for each of the 4 conditions.

2.5. EEG analysis

With its high temporal resolution, EEG provides a rich database to determine the exact latency of an effect. However, testing at all data points independently quickly leads to a multiple comparison problem where the risk of making Type I errors increases considerably. As EEG measures are not independent, but in fact temporally and spatially correlated, we used a non-parametric $t_{\max}$ permutation test to analyze the data [25, 26].

In $t_{\max}$ permutation testing, the null distribution is estimated by repeatedly resampling the obtained data and calculating t-scores for each sample. The most extreme t-scores ($t_{\max}$) are selected for the null distribution. Finally, the t-scores of the observed data are computed and compared to the simulated $t_{\max}$ distribution, just as in parametric hypothesis testing.

As with each permutation the chance of obtaining a large $t_{\max}$ increases, the test automatically becomes more conservative when making more comparisons. Also, since the actual, obtained data is used to estimate the null distribution, the test does not assume test independence, allowing for stringent control of Type I error without considerable decrease in sensitivity. To further maximize power and reduce the number of comparisons, the data were down-sampled to 125 Hz and time-windows were estimated following the method used in [21]. We were mainly interested in modulations of the P200 ($151 - 251$ ms) and the N325 ($201 - 431$ ms), as these two components reflect auditory processes in the pre-lexical stage of word processing (acoustical processing and stress extraction, respectively [27] [21]).

Each comparison of interest was analyzed with a separate repeated measures, two-tailed t-tests, using the original data and 2500 random permutations to approximate the null distribution for the customary family-wise alpha (α) level of 0.05[1].

3. Results

3.1. Behavioral results

Behavioral data (error rates and reaction times) were analyzed with paired two-tailed t-tests in R [29]. Overall, performance on the lexical decision task revealed high accuracy ($< 5\%$ errors) with no differences between conditions. Reaction times showed a main effect of lexicality ($t = -16.85$, $p < 0.001$); words were responded to faster than pseudowords. Presence of IA had no effect on response latencies ($p = 0.7$, ns).

3.2. ERP results

In the P200 time-window (Figure 2), there was a main effect of lexicality (critical t-score: ± 3.5589, $p < 0.05$). Pseudowords elicited a larger P200 than words in the frontocentral region (FC2) peaking 182 ms after stimulus presentation. The difference between words and pseudowords was also significant within the condition without IA (critical t-score: ± 3.575, $p < 0.05$). Within the condition with IA this effect was not significant ($p = 0.4$, ns).

In the N325 time-window (Figure 3), there was a main effect of presence of IA (critical t-score: ± 3.6887, $p < 0.05$). Compared to stimuli +IA, stimuli −IA elicited a larger negativity in the frontocentral region (FC2 and Cz) from $318 - 358$ ms after stimulus presentation. The difference in ERP amplitude is small, but robust and comparable to the amplitude difference reported in Böcker et al ($1 - 2.5$ μV). The effect was also significant within the lexical words condition (critical t-score: ± 3.8546, $p < 0.05$); words −IA resulted in a larger negativity than words +IA. There was a similar trend in the pseudowords condition, although it did not reach significance ($p = 0.09$, ns). A visual inspection of the ERPs, however, suggests similar processes between words and pseudowords.

4. Discussion

In the present study, we examined the interplay between accentuation and lexical access in French. We were particularly interested in the status and possible roles of the initial accent. Our results show that IA is represented in the French preferred and expected stress template. As pre-lexical language-specific stress templates are suggested to serve as gateways to the mental lexicon, IA could thus play an important role in the process of speech segmentation in French.

We used a lexical decision task in which we manipulated the presence of IA. The manipulation modulated the resultant frontocentral N325; a larger N325 emerged when IA had been

---

[1] In fact we used more than twice the number of permutations Manly suggested for an alpha at 5%.[28]

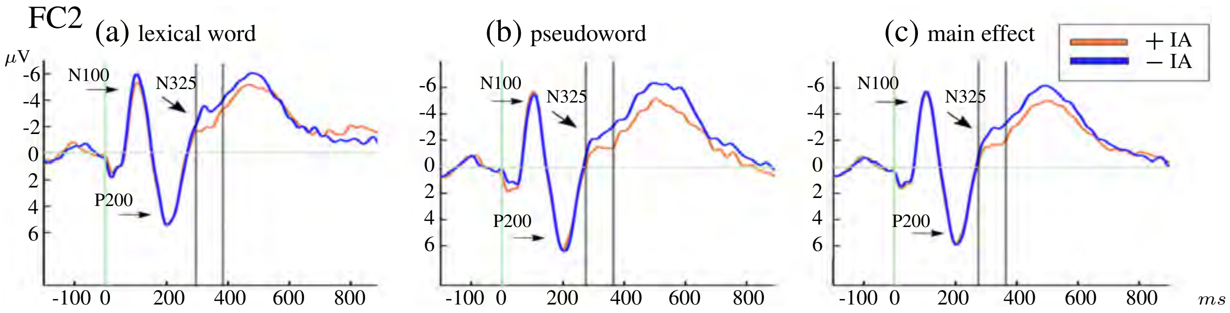


Figure 3: *Grand average N325 in the* ±*IA condition, recorded at the FC2 (frontocentral) electrode for: (a) lexical words, (b) pseudowords, (c) main effect.The vertical gray bars indicate the selected time window* $(201 - 431$ *ms). For ease of presentation, ERP waveforms are cut off at 800 ms. Negativity is plotted as an upward deflection.*

omitted. As the N325 is assumed to reflect difficulties in the extraction of lexical stress, this result indicates a stress processing cost when stimuli are presented without IA. Recall that Böcker et al. report similar findings after manipulating stress in Dutch, a language with lexical stress [21]. In their study, listeners were asked to discriminate between the Dutch dominant stress template and a less frequent stress template. Words presented without the dominant stress pattern elicited a more ample N325. In our study, words presented without IA resulted in the larger N325 suggesting that, even though stress is not lexically distinctive in French, IA is part of the French expected stress pattern (*cf* [20]). In addition, while in the study of Böcker et al. participants were asked to explicitly attend the metrical structure of the stimuli, in the present study, attention was diverted from the stress manipulation using a lexical decision task. Still finding a robust modulation of the N325, further demonstrates that word processing naturally engaged the pre-lexical extraction of IA. That is, we show that lexical access is facilitated when words are presented with the French preferred stress template, i.e. with the initial accent.

The amplitude modulation of the N325 was small (between $1 - 2.5\ \mu$V), but robust as revealed by our conservative non-parametric statistics (see *Methods section*). In fact, finding a relatively small difference in amplitude was expected and comparable to the amplitude difference in Böcker et al. [21]. Similar to Böcker et al., we did not manipulate the legality, but rather the probability of the presented stress templates. That is, while in French there is a preference for words marked with IA, words without IA are not illegal. Indeed, in continuous speech IA is not always realized and may be suppressed to serve for instance a more rhythmically balancing function. So, while French listeners may expect and prefer words to be marked with IA, words without IA do not exceedingly hamper word processing.

Our manipulation of IA did not modulate the P200, a component thought to reflect the bottom-up extraction of purely physical/acoustical parameters [27]. This indicates that our results reveal a more controlled process in which stress is extracted in a top-down fashion. We did find lexicality to affect the P200 when stimuli were presented without IA; pseudowords elicited a more ample P200 than did words when presented without IA. This effect is surprising since the latency range of the P200 precedes lexical processing [30]. Considering the location of the P200 (frontocentral; similar to the location of the N325) the effect appears to be the product of a temporal overlap between the P200 and the N325. The N325 was more negative for −IA stimuli than +IA stimuli and this difference was larger in the words condition than in the pseudowords condition. This means that in the −IA condition the overlap between the P200 and N325 will be more evident for words than for pseudowords, while in the +IA condition the overlap will be smaller (as +IA stimuli elicited a smaller N325). In fact, Böcker et al. report a similar overlap between the N325 and the P200 at the frontocentral electrodes. Finding an overlap between the N325 and the P200 implies that the process of stress extraction starts before our predefined N325 time-window $(201 - 431$ ms) and during the P200 time-window $(151 - 251$ ms). Such an early latency confirms that lexical access crucially involved an automatic, pre-lexical extraction of the French initial accent.

A visual inspection of the ERP components suggests ±IA also affected the later integration stages of word processing, as there seems to be an amplitude difference in the latency range typically associated with the N400 [31]. It is however unlikely that these late amplitude modulations really reflect difficulties in the post-lexical process of semantic integration, as word were presented in isolation. Moreover, the N400 is typically maximal over centroparietal sites [31, 32], while the reported ERPs in the current study have a frontocentral distribution. A more probable explanation is suggested by Böcker et al., who encountered a similar late frontocentral amplitude difference and interpret it to reflect N325 residue. A study to determine if ±IA also affects the later stages of speech processing, is currently in progress. In the study, ±IA words are embedded within a congruent or incongruent semantic context, and as such, the study is better adapted to give insight into whether IA also affects the later stages of speech processing.

5. Conclusions

In this study, we investigated the status of the French initial accent. Our ERP results demonstrate that IA is linked to the phonological representation of words. Words presented without initial accent elicited a more ample N325, a component that indexes difficulties in pre-lexical stress extraction. Moreover, as we diverted attention away from our stress manipulation with a lexical decision task, the extraction of IA seems to be an automatic step during the early stages of word processing. This indicates that the initial accent is part of the French preferred stress template and as such, contrary to popular belief, plays a valuable role in French speech comprehension.

6. Acknowledgments

This study is supported by the Agence Nationale de la Recherche grant ANR-12-BSH2-0001 (PI: Corine Astésano)

7. References

[1] A. Cutler and D. Norris, "The role of strong syllables in segmentation for lexical access." *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 14, no. 1, pp. 113–121, 1988.

[2] A. Cutler and D. M. Carter, "The predominance of strong initial syllables in the English vocabulary," *Comput. Speech Lang.*, vol. 2, no. 3-4, pp. 133–142, 1987.

[3] J. Vroomen and B. De Gelder, "Metrical segmentation and lexical inhibition in spoken word recognition." *Journal of Experimental Psychology: Human perception and performance*, vol. 21, no. 1, p. 98, 1995.

[4] J. Mehler and R. Hayes, "The role of syllables in speech processing: Infant and adult data [and discussion]," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 295, no. 1077, pp. 333–352, 1981.

[5] J. Mehler, J. Y. Dommergues, U. Frauenfelder, and J. Segui, "The syllable's role in speech segmentation," *Journal of verbal learning and verbal behavior*, vol. 20, no. 3, pp. 298–305, 1981.

[6] A. Cutler, J. Mehler, D. Norris, and J. Segui, "The syllable's differing role in the segmentation of french and english," *Journal of memory and language*, vol. 25, no. 4, pp. 385–400, 1986.

[7] M. Rossi, *L'intonation: le système du français: description et modélisation*. Editions Ophrys, 1999.

[8] E. Selkirk, "The prosodic structure of function words," *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, vol. 187, p. 214, 1996.

[9] A. Di Cristo, "Vers une modélisation de l'accentuation du français (seconde partie)," *Journal of French Language Studies*, vol. 10, no. 01, pp. 27–44, 2000.

[10] M.-H. Banel and N. Bacri, "On metrical patterns and lexical parsing in french," *Speech Communication*, vol. 15, no. 1, pp. 115–126, 1994.

[11] O. Bagou, C. Fougeron, and U. H. Frauenfelder, "Contribution of prosody to the segmentation and storage of" words" in the acquisition of a new mini-language," in *Speech Prosody 2002, International Conference*, 2002.

[12] A. Christophe, S. Peperkamp, C. Pallier, E. Block, and J. Mehler, "Phonological phrase boundaries constrain lexical access i. adult data," *Journal of Memory and Language*, vol. 51, no. 4, pp. 523–547, 2004.

[13] P. Welby, "The role of early fundamental frequency rises and elbows in french word segmentation," *Speech Communication*, vol. 49, no. 1, pp. 28–48, 2007.

[14] E. Spinelli, P. Welby, and A.-L. Schaegis, "Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of french elision," *Language and cognitive processes*, vol. 22, no. 6, pp. 828–859, 2007.

[15] E. Spinelli, N. Grimault, F. Meunier, and P. Welby, "An intonational cue to word segmentation in phonemically identical sequences," *Attention, Perception, & Psychophysics*, vol. 72, no. 3, pp. 775–787, 2010.

[16] A. Content, R. K. Kearns, and U. H. Frauenfelder, "Boundaries versus onsets in syllabic segmentation," *Journal of Memory and Language*, vol. 45, no. 2, pp. 177–199, 2001.

[17] S.-A. Jun and C. Fougeron, "A phonological model of french intonation," in *Intonation*. Springer, 2000, pp. 209–242.

[18] C. Astésano, E. G. Bard, and A. Turk, "Structural influences on initial accent placement in French." *Lang. Speech*, vol. 50, no. Pt 3, pp. 423–446, 2007.

[19] C. Astésano, R. Bertrand, R. Espesser, and N. Nguyen, "Perception des frontières et des proéminences en français," *JEP-TALN-RECITAL*, 2012.

[20] M. Aguilera, R. El Yagoubi, R. Espesser, and C. Astésano, "Event-Related Potential of Initial Accent Processing in French," *Speech Prosody 2014*, no. Umr 5263, pp. 3–7, 2014.

[21] K. B. Böcker, M. Bastiaansen, J. Vroomen, C. H. Brunia, and B. Gelder, "An erp correlate of metrical stress in spoken word recognition," *Psychophysiology*, vol. 36, no. 6, pp. 706–720, 1999.

[22] P. Boersma and D. Weenink, "{P} raat: doing phonetics by computer," 2010.

[23] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[24] T. MathWorks, "Image processing toolboxtm user's guide, version r2014a," *The MathWorks*, 2014.

[25] D. M. Groppe, T. P. Urbach, and M. Kutas, "Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review," *Psychophysiology*, vol. 48, no. 12, pp. 1711–1725, 2011.

[26] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.

[27] S. A. Hillyard and T. W. Picton, "Electrophysiology of cognition," *Comprehensive Physiology*, 1987.

[28] B. Manly, "Randomization, bootstrap, and monte carlo methods in biologychapman and hall," *New York*, 1997.

[29] R. C. Team, "The r project for statistical computing," *R Foundation for Statistical Computing web-site. www. R-project. org. Accessed June*, vol. 9, 2014.

[30] F. Grosjean, "Spoken word recognition processes and the gating paradigm," *Perception & Psychophysics*, vol. 28, no. 4, pp. 267–283, 1980.

[31] C. Brown and P. Hagoort, "The processing nature of the n400: Evidence from masked priming," *Journal of Cognitive Neuroscience*, vol. 5, no. 1, pp. 34–44, 1993.

[32] M. Kutas and K. D. Federmeier, "Electrophysiology reveals semantic memory use in language comprehension," *Trends in cognitive sciences*, vol. 4, no. 12, pp. 463–470, 2000.

The phonological status of the French Initial Accent and its role in semantic processing: an Event-Related Potentials study

Noémie te Rietmolen[1], Radouane El Yagoubi[2], Alain Ghio[3] and Corine Astésano[1]

[1]U.R.I Octogone-Lordat (E.A. 4156), Université de Toulouse, UTM, Toulouse, France
[2]Laboratoire CLLE-LTC (UMR 5263), Université de Toulouse, UTM, Toulouse, France
[3]Laboratoire Parole & Langage (UMR 7309), Aix-Marseille Université, Aix-en-Provence, France
noemie.te-rietmolen@univ-tlse2.fr, yagoubi@univ-tlse2.fr, alain.ghio@lpl-aix.fr, astesano@univ-tlse2.fr

Abstract

French accentuation is held to belong to the level of the phrase. Consequently French is considered 'a language without accent' with speakers that are 'deaf to stress'. Recent ERP-studies investigating the French initial accent (IA) however demonstrate listeners not only discriminate between different stress patterns, but also prefer words to be marked with IA early in the process of speech comprehension. Still, as words were presented in isolation, it remains unclear whether the preference applied to the lexical or to the phrasal level. In the current ERP-study, we address this ambiguity and manipulate IA on words embedded in a sentence. Furthermore, we orthogonally manipulate semantic congruity to investigate the interplay between accentuation and later speech processing stages. Preliminary results on 14 participants reveal a significant interaction effect: the centro-frontally located N400 was larger for words without IA, with a bigger effect for semantically incongruent sentences. This indicates that IA is encoded at a lexical level and facilitates semantic processing. Furthermore, as participants attended to the semantic content of the sentences, the finding underlines the automaticity of stress processing. In sum, we demonstrate accentuation plays an important role in French speech comprehension and call for the traditional view to be reconsidered.

Index Terms: speech perception, prosody, semantic processing, Event-Related Potentials, N400

1. Introduction

While in written form, language is structured by white spaces and punctuation marks, spoken language is organized through intonation, accentuation, and rhythm. Clearly, prosody plays an essential role in speech comprehension. Metrical structures, for instance, have long been considered crucial in the segmentation of speech. With no clear separation between words in the speech signal, the metrical segmentation strategy (MSS) proposes that listeners rely on their languages' metrical pattern to identify word boundaries [1, 2].

Indeed, in stress languages, such as English or Dutch, in which stress is part of the lexical entry, accents provide reliable cues to lexical boundaries. Even in French, a language often described to be syllable-based due to the fairly homogeneous metrical weight on syllables, prosodic structure has been found to guide speech segmentation [3, 4, 5, 6, 7, among others]. However, in these studies, segmentation was not considered lexical but presumed phrasal, i.e. listeners are assumed to adopt a *prosodic segmentation strategy* in which intonational and accentual patterns function to segment *prosodic groups* (level of AP [8]) from the speech signal [9]. This view stems from traditional descriptions of French as 'a boundary language' [10] or 'a language without accent' [11] according to which stress, because it is not lexically distinctive in French and its surface realization is acoustically merged with intonational boundaries, has no clear metrical value.

Di Cristo's metrical model of French, however, posits that lexical words are encoded with (latent) cognitive stress templates underlying their phonological representation [12]. These stress patterns comprise both a primary final accent (FA) on the last syllable of the word and a secondary and optional initial accent (IA) on the word's first syllable. That is, according to Di Cristo's model, words are marked with metrically strong syllables at both left and right lexical boundaries that can readily notify listeners on when to initiate lexical access. The model therefore provides a valuable theoretical context to speech segmentation in French.

In the current ERP-study, we investigate the representation of the French initial accent. Although the initial accent is thought of as an optional secondary accent in French, and mainly recognized for its rhythmic balancing function and role in emphasis placement, studies showing IA to also play an important role in the marking of lexical structure and speech segmentation are accumulating. Indeed, a perception study in which the accoustic parameters of IA had been manipulated, indicated listeners to have a strong phonological preference for IA [13]. In addition, IA has been found to be a more reliable cue to word boundaries than FA and to be perceived as more prominent at both phrasal and lexical levels [14, 15, 16]. These results prompted a recent paper to revisit the secondary and optional nature of IA and suggest IA carries a metrical strength that is equal to that of FA, both accents working together in the marking of the lexical word [17].

Recent neuroimaging studies corroborate this idea and underline the role of IA in French word processing. When presenting words with or without IA in an oddball study, Aguilera and colleagues obtained a larger MisMatch Negativity components (MMN) when the oddball had been presented without IA than when the oddball was presented with IA [18]. This not only shows that French listeners heard the accent, and are not deaf to the stress, but also that IA is encoded in long-term memory and part of the expected stress template. Following up on these results, we presented listeners with trisyllabic nouns and pseudowords with or without IA in a lexical decision task [19]. Omitting IA resulted in a processing cost during stress extraction as reflected by a more ample N325 [20] regardless of lexical condition. This demonstrates both the automaticity of stress extraction and an expectation for words to be marked with IA in the pre-lexical stage of speech processing.

However, in both ERP-studies, words were presented in isolation, with IA always in utterance initial position. Because words had been presented as independent utterances, they may have been processed as individual accentual phrases (AP). Hence, it can not be ruled out that the templates - and the processing cost when IA was omitted - applied to the phrase level instead of the level of the lexical word. In the current study we sought to elucidate this ambiguity and manipulated IA on words positioned within a sentence. Additionally, we manipulated the semantic congruity of the sentences, allowing us to investigate whether IA also affects later processing stages in speech comprehension.

Indeed, in a previous ERP study investigating the relationship between metrical structure and late speech processing in French, metrical violations were found to obstruct semantic processing [21, 22]. In the study, participants listened to sentences in which semantic and/or metrical congruity was manipulated. Semantic congruity was manipulated by presenting sentences in which the last word was incoherent with the semantic context of the sentence, while metrical congruity was manipulated by lengthening the medial syllable of the last word, an illegal stress pattern in French. The metrical violation resulted in an increased N400, even when the sentences were semantically congruent. As the N400 component is thought to reflect a discrepancy to lexico-semantic expectations [23, 24], these results indicate that accentual patterns affect the later stages of speech comprehension, during which access to meaning and semantic integration takes place.

However, as the processing cost resulted from presenting an illegal stress pattern, with metrical weight on the medial syllable, it remains unclear whether semantic processing also suffers when words are presented with metrical structures that deviate from the expected stress pattern. Or put more concretely, if IA is linked to the phonological representation of words and is, along with FA, the expected stress template in French, we anticipate that presenting words without IA also elicits a larger N400.

2. Methods

2.1. Speech stimuli

The same stimuli were used as in [21, 22] and consisted of French carrier sentences ending with either a semantically congruent (+S) or incongruent (−S) trisyllabic target noun (see Figure 1). Congruent and incongruent target words were matched in word frequency, acoustic and phonological characteristics, and word and syllable duration (a more detailed account on the construction of the sentences can be found in [21]).

80 carrier sentences with the most natural IA on the target noun in both semantic conditions were selected. As the fundamental frequency ($f0$) is the phonetic signature of IA [25], we selected only stimuli in the first syllable of the target noun was marked by a $f0$ rise of at least 10% compared to the preceding $f0$ value on the determinant [26, 14].

A customized quadratic algorithm [18] in PRAAT [27] was used to create the accent condition. To remove the natural IA on the target words (−IA condition), the $f0$ value of the first vowel (i.e. IA) was lowered near the $f0$ value of the preceding (unaccented) determinant. The algorithm progressively modified the $f0$ values to reach the $f0$ value at the beginning of the last (accented) vowel. This quadratic transformation allowed for micro-prosodic variations to be maintained, thus keeping the natural sound of the stimuli. The +IA stimuli were forward and back transformed to equalize the speech quality between +IA and −IA stimuli.

The resulting 320 stimuli over the four experimental conditions (+S+IA, −S+IA, +S−IA, and −S−IA) were divided over four lists, such that each participant was presented with 80 unique sentences.

2.2. Participants and procedure

16 French native speakers, aged $20 - 47$ (mean age 24.6), originally took part in the study. All subjects were right-handed, with normal hearing abilities and no reported history of neurological or language-related problems. Due to excessive artifacts in the EEG signal, two participants are excluded from analyses.

Each participant was comfortably seated in an electrically shielded and sound attenuated room and presented the stimuli through headphones.

Participants were instructed to judge as quickly and accurately as possible whether a sentence was semantically congruent or incongruent by pressing the left or right arrow key on a standard keyboard (arrow key assignment was counter-balanced across participants). To ensure participants understood the task requirements, the experiment began with a short practice phase, consisting of 10 trials that were similar to the experimental trials, but not included in the analyses.

Each participant listened to all 80 stimuli. Using Latin square designs, the four conditions (+S+IA, −S+IA, +S−IA, and −S−IA) were evenly distributed over two

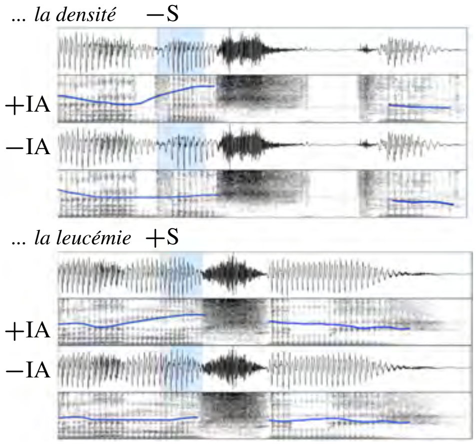
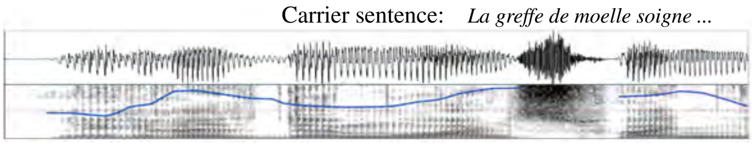


Figure 1: *Example of $f0$ resynthesis with (+IA) and without initial accent (−IA) on semantically incongruent (−S, top two) and semantically congruent (+S, bottom two) sentences with quadratic interpolation from the $f0$ value of the preceding determinant to the $f0$ value at the beginning of the last stressed syllable for +IA targets (visible in blue).*



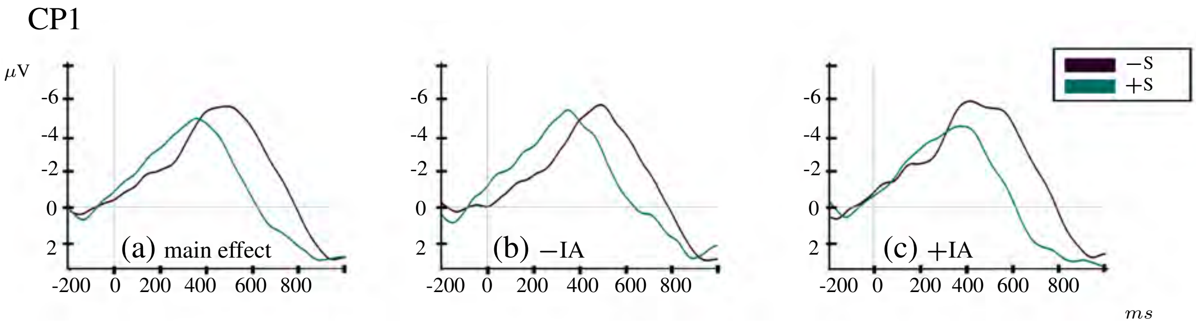


Figure 2: *Grand average N400 in the semantic congruity condition (*$-S$ *in purple,* $+S$ *in green), recorded at the CP1 (centro-parietal) electrode for: (a)* $-IA$*, (b)* $+IA$*, (c) main effect. For ease of presentation, ERP waveforms are filtered at* 10 *Hz.*

blocks, with block order balanced between participants. Total duration of the experiment, including the set-up of the EEG electrodes, was approximately $1,5$h.

2.3. EEG recording and preprocessing

EEG data were recorded with 64 Ag/AgCl-sintered electrodes mounted on an elastic cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 System; Jasper, 1958). The EEG signal was amplified by BioSemi amplifiers (ActiveTwo System) and digitized at 2048 Hz.

The data were preprocessed using the EEGLAB package [28] in Matlab [29]. Each electrode was re-referenced offline to the algebraic average of the left and right mastoids. The data were band-pass filtered between $0.01 - 30$ Hz and resampled at 256 Hz. Following a visual inspection, signal containing EMG or other artifacts not related to eye-movements or blinks was manually removed. Independent Components Analysis (ICA) was performed on the remaining data in order to identify and subtract components containing oculomotor artifacts. Finally, data were epoched from -0.2 to 1 seconds surrounding the onset of the target word and averaged within and across participants to obtain the grand-averages for each of the four conditions.

2.4. EEG analysis

The method of EEG is well known for its temporal precision and thus aptly suited to track down online processes. However, the high temporal resolution comes at the cost of many comparisons when ERP amplitude values for each individual electrode, at each recorded time-point, are tested independently, using standard parametric statistics (e.g. ANOVA or t-test). Because EEG measures are not independent, but instead temporally and spatially correlated, we use a non-parametric t_{max} permutation test to analyze the data [30, 31].

Since N400 amplitude modulations resulting from semantic incongruities are typically maximal in the centro-parietal region of the brain [24, 23], we selected the central and parietal electrodes to test for an effect on semantic processing. The effect of metrical expectancy violations on speech comprehension is typically found in temporal, central and frontal brain areas [20, 19]. Therefore, the effect of metrical expectancy was tested on the temporal and centro-frontal electrodes. To further reduce the number of comparisons and maximize statistical power, a time-window of $350 - 550$ ms surrounding the N400 was selected and data were down-sampled to 128 Hz.

3. Results

Behavioral data (error rates and reaction times) were analyzed with paired two-tailed t-tests in R [32]. Overall, performance on the semantic congruency task revealed high accuracy ($< 5\%$ errors) with no differences between conditions. Reaction times showed a main effect of semantic congruity ($t = -3.09$, $p < 0.05$); congruent sentences were responded to faster than incongruent sentences. Presence of IA had no effect on response latencies ($p = 0.8$, ns).

ERP data show a main effect of semantic congruity (critical t-score: ± 4.6322, $p < 0.05$; Figure 2a): semantically incongruent sentences elicited a larger N400 at 500 ms after the onset of the target word than semantically congruent sentences in the left centro-parietal region (CP1). This difference in N400 amplitude was also significant within the condition without IA

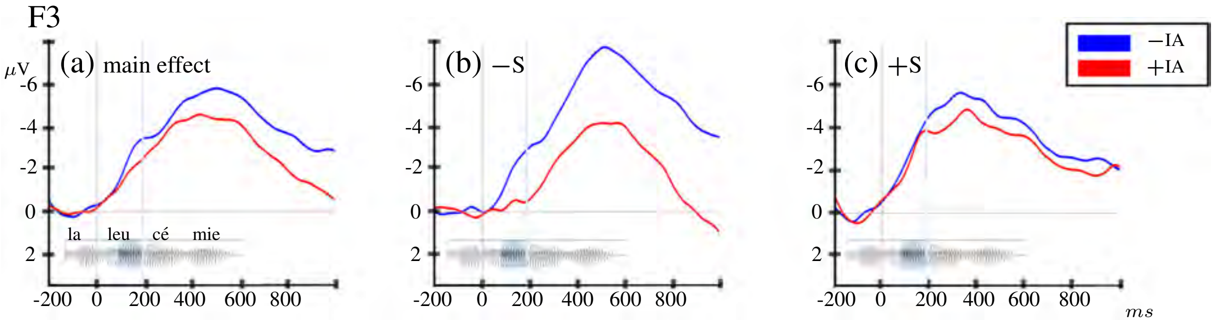


Figure 3: *Grand average N400 in the IA condition (*$-IA$ *in blue,* $+IA$ *in pink), recorded at the F3 (fronto-central) electrode for: (a) main effect, (b) incongruent sentences, (c) congruent sentences. For ease of presentation, ERP waveforms are filtered at* 10 *Hz.*

(critical t-score: ± 3.575, $p < 0.05$; Figure 2b) and marginally significant within the condition with IA (critical t-score: ± 3.575, $p = 0.056$; Figure 2c).

Furthermore, visual inspection suggested a difference in N400 onset latency between semantically congruent and incongruent sentences, but only in the $-$IA condition. Because this visual effect is important for the discussion of the additional semantic processing cost when words are presented $-$IA, we computed a t-test specifically focused on differences in peak latencies at CP1. Results show that the N400 onset was significantly delayed when sentences were incongruent compared to when they were congruent ($t = -3.09, p < 0.05$).

Finally, while the main effect of ± IA did not reach significance ($p = 0.24$, ns; Figure 3a), we did find a significant interaction effect between semantic congruity and ±IA in the left frontocentral regions (F3): the N400 was significantly bigger in $-$IA condition than in the $+$IA condition for semantically incongruent sentences 429 ms after target word onset (critical t-score: ± 4.1932, $p < 0.05$; Figure 3b). The difference between ±IA was considerably smaller, and not significant ($p = 0.26$, ns; Figure 3c), when sentences were semantically congruent.

4. Discussion

In the present study, we examined the phonological status of the French initial accent and its role in semantic processing. We were particularly interested in modulations of the N400 ERP component, a component typically observed subsequent to violations in lexico-semantic expectations [24]. Bellow, we will discuss each of our results in turn: in (4.1) we inspect whether metrical expectancy affected lexico-semantic processing, in (4.2) we discuss the effect of metrical expectation on speech processing and in (4.3) we revisit the role of IA as secondary boundary marker of the AP.

4.1. Semantic congruity effect

As expected, we found a main effect of semantic congruity in centro-parietal regions (*cf.* Figure 2): semantically incongruent sentences elicited a more ample N400 than did semantically congruent sentences. This effect was also significant within the $-$IA condition and marginally significant within the $+$IA condition. Interestingly, close inspection of the ERP waveforms revealed an additional delay in N400 onset latency when semantically incongruent words had been presented without initial accent. This indicates an interaction effect such that when words are presented without IA, semantic conflict resolution starts later. Furthermore, because there was no delay in semantic resolution processes when words were presented $+$IA, but only an amplitude difference typical to the semantic N400, listeners appeared to expect words to marked with IA, i.e. IA is phonologically natural.

4.2. Metrical expectancy effect

Our results in the fronto-central brain area partially confirm the phonological status of IA (*cf.* Figure 3): When sentences were semantically incongruent, words $-$IA elicited a larger N400 than did words $+$IA. Surprisingly, however, ± IA did not affect speech processing when sentences were semantically congruent. It is possible listeners only appealed to their preferred stress patterns when facing difficulties in the later stages of speech processing, however we consider it more probable the effect did not surface due to low statistical power.

The results presented here are preliminary, because, to this date, we collected data on only 14 participants, each answering to 20 sentences per condition. Moreover, we used a conservative permutation statistic that maximally avoids false discoveries, but may have instead given rise to an effect remaining undetected.

Another indication our statistics may have lacked in sensitivity is demonstrated by the marginally significant effect between semantically congruent and incongruent sentences when words were presented with initial accent. As previously stated, semantic congruency is abundantly found to influence N400 amplitudes, and in fact did so in our study when words were presented $-$IA. Arguably, the double processing cost when both semantic and metrical information was unexpected led to an effect that was big enough to be detected by our analysis, while the effect was less apparent when only was of the two conditions was unexpected. The N400 modulations resulting from our ± IA conditions confirm this theory. Whereas the effect was highly significant when sentences were semantically incongruent (and thus unexpected), when the sentences were congruent, the difference was much smaller. Still finding a significant interaction between stress patterns and lexico-semantic processing demonstrates that French accentuation is crucially involved in this later stage of speech comprehension.

4.3. The initial accent as a lexical boundary marker

The interactions reported above have another important implication. As discussed in the introduction, previous ERP studies investigating the phonological status of IA, while demonstrating a phonological expectancy for IA, had not been able to distinguish between lexical and phrasal processing. In the current study IA was not utterance initial but embedded in a sentence. Also, we manipulated the semantic congruency of the sentences, allowing us to investigate whether stress patterns affect lexico-semantic processes as reflected by the N400. We found metrical expectancy to modulate the N400 both in the centro-parial and in the centro-frontal brain regions. That is, when asking listeners to judge the semantic congruity of sentences that differed only in ± IA, lexico-semantic processing (as reflected by the N400) was still affected. This result not only indicates IA to play a valuable role in lexico-semantic processes, but also demonstrates that French speech comprehension *naturally* and *automatically* engages lexical stress processing.

5. Conclusions

In this study, we investigated the status of the French initial accent and its function in lexico-semantic processing. Our ERP results demonstrate that IA is linked to the representation of the lexical word, affecting every stage in the process of speech comprehension. Indeed, previous studies have shown a disruption in pre-lexical stress processing when IA had been omitted. As pre-lexical stress templates serve to access the mental lexicon, presenting words without IA in turn hinders lexical access and cascades up the process of speech comprehension to affect lexico-semantic processing. In sum, the study demonstrates that French listeners expect words to be marked with the initial accent, and actively, though automatically, make use of the accent throughout the process of speech comprehension.

6. Acknowledgment

This study is supported by the Agence Nationale de la Recherche grant ANR-12-BSH2-0001 (PI: Corine Astésano)

7. References

[1] A. Cutler and D. Norris, "The role of strong syllables in segmentation for lexical access," *J. Exp. Psychol. Hum. Percept. Perform.*, 1988.

[2] A. Cutler, *Exploiting prosodic probabilities in speech segmentation*. The MIT Press, 1990.

[3] P. Welby, "The role of early fundamental frequency rises and elbows in french word segmentation," *Speech Commun.*, vol. 49, no. 1, pp. 28–48, Jan. 2007.

[4] E. Spinelli, N. Grimault, F. Meunier, and P. Welby, "An intonational cue to word segmentation in phonemically identical sequences," *Atten. Percept. Psychophys.*, vol. 72, no. 3, pp. 775–787, Apr. 2010.

[5] M.-H. Banel and N. Bacri, "On metrical patterns and lexical parsing in french," *Speech Commun.*, vol. 15, no. 1, pp. 115–126, 1 Oct. 1994.

[6] O. Bagou, C. Fougeron, and U. H. Frauenfelder, "Contribution of prosody to the segmentation and storage of" words" in the acquisition of a new mini-language," in *Speech Prosody 2002, International Conference*, 2002.

[7] A. Christophe, S. Peperkamp, C. Pallier, E. Block, and J. Mehler, "Phonological phrase boundaries constrain lexical access i. adult data," *J. Mem. Lang.*, vol. 51, no. 4, pp. 523–547, Nov. 2004.

[8] S.-A. Jun and C. Fougeron, "A phonological model of french intonation," in *Intonation*, ser. Text, Speech and Language Technology, A. Botinis, Ed. Springer Netherlands, 2000, pp. 209–242.

[9] S. Wauquier-Gravelines, "Segmentation lexicale de la parole continue: la linéarité en question," *Recherches linguistiques de Vincennes*, no. 1, pp. 8–8, 1999.

[10] J. Vaissière, "Rhythm, accentuation and final lengthening in french," in *Music, Language, Speech and Brain*, J. Sundberg, L. Nord, and R. Carlson, Eds. London: Macmillan Education UK, 1991, pp. 108–120.

[11] M. Rossi, "Le français, langue sans accent?" *Studia Phonetica Montréal*, 1980.

[12] A. Di Cristo, "Vers une modélisation de l'accentuation du français (seconde partie)," *Journal of French language studies*, vol. 10, no. 01, pp. 27–44, 2000.

[13] L. Jankowski, C. Astésano, and A. Di Cristo, "The initial rhythmic accent in french: Acoustic data and perceptual investigation," in *Proceedings of the 14th International Congress of Phonetic Sciences*. 14th International Congress of Phonetic Sciences (ICPhS'99), San Francisco, 1999, pp. 257–260.

[14] C. Astésano, E. G. Bard, and A. Turk, "Structural influences on initial accent placement in french," *Lang. Speech*, vol. 50, no. Pt 3, pp. 423–446, 2007.

[15] C. Astésano, R. Bertrand, R. Espesser, and N. Nguyen, "Perception des frontières et des proéminences en français," in *Actes de la conférence conjointe JEP-TALN-RECITAL*, vol. 1, 2012, pp. 353–360.

[16] L. Garnier, L. Baqué, A. Dagnac, and C. Astésano, "Perceptual investigation of prosodic phrasing in french," in *Speech Prosody 2016*, 2016.

[17] C. Astesano and R. Bertrand, "Accentuation et niveaux de constituance en français : enjeux phonologiques et psycholinguistiques," *Langue française*, vol. 191(3), no. 11, 30 Sep. 2016.

[18] M. Aguilera, R. El Yagoubi, R. Espesser, and C. Astésano, "Event-related potential investigation of initial accent processing in french," pp. 383–387, 2014.

[19] N. te Rietmolen, R. El Yagoubi, R. Espesser, C. Magnen, and C. Astesano, "Investigating the phonological status of the initial accent in french: an Event-Related potentials study," in *Speech Prosody 2016*, 2016.

[20] K. B. Böcker, M. C. Bastiaansen, J. Vroomen, C. H. Brunia, and B. de Gelder, "An ERP correlate of metrical stress in spoken word recognition," *Psychophysiology*, vol. 36, no. 6, pp. 706–720, Nov. 1999.

[21] C. Magne, C. Astésano, M. Aramaki, S. Ystad, R. Kronland-Martinet, and M. Besson, "Influence of syllabic lengthening on semantic processing in spoken french: Behavioral and electrophysiological evidence," *Cereb. Cortex*, vol. 17, no. 11, pp. 2659–2668, 1 Nov. 2007.

[22] C. Astésano, C. Magne, R. El Yagoubi, and M. Besson, "Influence du rythme sur le traitement sémantique en français: approches comportementale et électrophysiologique," in *JEP 2004, Fes, Maroc*. JEP 2004, Maroc, 2004.

[23] M. Kutas and K. D. Federmeier, "Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP)," *Annu. Rev. Psychol.*, vol. 62, pp. 621–647, 2011.

[24] C. Brown and P. Hagoort, "The processing nature of the n400: evidence from masked priming," *J. Cogn. Neurosci.*, vol. 5, no. 1, pp. 34–44, 1993.

[25] C. Astésano, *Rythme et accentuation en français: invariance et variabilité stylistique*. L'Harmattan, 2001.

[26] D. Robert Ladd, *Intonational Phonology*. Cambridge University Press, 4 Dec. 2008.

[27] P. Boersma and D. Weenink, "Praat software. university of amsterdam," *Im Internet: http://www. fon. hum. uva. nl/praat*, 2016.

[28] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 15 Mar. 2004.

[29] I. Mathworks, "MATLAB: R2014a," *Mathworks Inc, Natick*, 2014.

[30] D. M. Groppe, T. P. Urbach, and M. Kutas, "Mass univariate analysis of event-related brain potentials/fields i: a critical tutorial review," *Psychophysiology*, vol. 48, no. 12, pp. 1711–1725, Dec. 2011.

[31] S. Luck, "The mass univariate approach and permutation statistics," in *ERP analysis*, 2014.

[32] R. C. Team, "The R project for statistical computing," *Available at www. R-project. org/. Accessed October*, vol. 31, p. 2014, 2014.