UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department: Dipartimento di Psicologia dello Sviluppo e della Socializzazione

Ph.D. COURSE IN: Psychological Sciences

SERIES: XXXVIII

Adapting Predictions: How Speaker Variability Shapes Prediction during Language Comprehension

Coordinator: Prof. Lucia Regolin

Supervisor: Prof. Francesca Peressotti

Co-Supervisor: Prof. Francesco Vespignani

Ph.D. student: Marco Sala

Contents

Abstract

English

Humans regularly interact with speakers who vary in accent, voice, speech rate, and articulation. Despite this variability, listeners typically comprehend speech with remarkable ease. One explanation for this efficiency is that comprehension involves not only the bottom-up integration of incoming information, but also top-down predictions about upcoming input that are guided by contextual information and internal knowledge. Nevertheless, it remains unclear how detailed these predictions are and how speaker variability influences their generation.

In Studies 1 and 2, we investigated whether comprehenders anticipate the phonological form of a predictable word by using behavioral and electroencephalography (EEG) methods. To do so, we capitalized on the fact that foreign-accented speakers typically make systematic phonological errors. In both studies, participants read sentence fragments followed by a final word spoken either by a native- or foreign-accented speaker. The spoken word could be predictable or not based on the context of the sentence. Crucially, the speaker's identity was either cued or not by an image of the face of the speaker. In Study 1, participants performed a lexical decision task on the spoken target. In Study 2, they judged whether the final word matched their expectations on some proportion of trials and with no time pressure, while we measured Event-Related Potentials (ERPs). We observed that cueing the speaker's identity was associated with both faster lexical decision times and a smaller negative amplitude (300-500 ms after word onset) for predictable words but not for unpredictable words. These results indicate that prediction relies on flexible and finely tuned processes capable of accommodating interindividual phonological variability, suggesting that lexical information is pre-activated at the phonological level.

Study 3 aims to investigate how listeners cope with acoustic differences between native speakers and whether speaker variability, in turn, influences prediction during naturalistic speech processing. In this ongoing study, we collected EEG data while participants listened to narrative stories under two experimental conditions: in the *Single-speaker* condition, one speaker narrated the entire story; in the *Multi-speaker* condition, different speakers narrated different sections of the story. We plan to apply Temporal Response Function (TRF) analysis to model how the brain encodes stimulus features related to both speech perception and predictive processing across listening conditions.

Overall, this work underscores the flexibility of the human brain in processing linguistic

input. Our findings suggest that prediction operates as a flexible mechanism accommodating speaker variability, and point towards future investigations of its function in naturalistic speech.

Italiano

Gli esseri umani interagiscono quotidianamente con parlanti che variano per accento, timbro vocale, velocità dell'eloquio e stile articolatorio. Nonostante questa variabilità, le persone riescono a comprendere il linguaggio con notevole facilità. Una possibile spiegazione è che la comprensione del linguaggio non dipende solo dall'integrazione passiva delle informazioni in arrivo, ma coinvolge anche meccanismi di predizione guidati dal contesto e dalla conoscenza del mondo. Tuttavia, non è ancora chiaro quanto siano dettagliate queste predizioni né in che modo la variabilità tra parlanti ne influenzi la generazione.

Negli Studi 1 e 2, abbiamo investigato se è possibile anticipare la forma fonologica di una parola predicibile utilizzando misure comportamentali e l'elettroencefalogramma (EEG). A tal fine, abbiamo capitalizzato sul fatto che i parlanti con accento straniero tendono a commettere errori fonologici sistematici. In entrambi gli studi, i partecipanti leggevano delle frasi la cui ultima parola veniva pronunciata da un parlante con accento nativo o straniero. Questa parola poteva essere predicibile o meno in base al contesto della frase. La presentazione degli stimoli era accompagnata da uno stimolo visivo neutro o da un'immagine del volto del parlante, che permetteva di anticipare l'identità del parlante. Nello Studio 1, i partecipanti svolgevano un compito di decisione lessicale sulla parola presentata uditivamente. Nello Studio 2, dovevano giudicare se si aspettavano la parola finale in una proporzione dei trial e senza pressione temporale, mentre venivano misurati i Potenziali Evento-Correlati (ERP). I risultati hanno mostrato che la presentazione dell'immagine del volto del parlante è associata sia a tempi di decisione lessicale più rapidi sia a una minore negatività nella finestra temporale dell'N400 (tra 300 e 500 ms dall'inizio della parola) per le parole predicibili ma non per le parole non predicibili. Questi risultati indicano che i processi predittivi durante la comprensione del linguaggio considerano la variabilità fonologica tra parlanti, suggerendo che l'informazione lessicale viene pre-attivata a livello fonologico.

Lo Studio 3 ha lo scopo di esplorare come le persone gestiscono le differenze acustiche tra parlanti nativi e se la variabilità tra parlanti influenza i processi predittivi durante l'elaborazione del linguaggio in contesti naturalistici. In questo studio, tuttora in corso, abbiamo registrato l'attività EEG dei partecipanti mentre ascoltavano delle storie. Le storie potevano essere narrate interamente da un unico parlante oppure da parlanti diversi, ciascuno dei quali narrava una sezione differente della storia. In entrambe le condizioni sperimentali, utilizzeremo

la Temporal Response Function (TRF) per esaminare come il cervello elabora proprietà degli stimoli legate alla percezione di categorie fonemiche e ai processi predittivi.

Nel complesso, questo lavoro mette in luce la flessibilità del cervello umano nell'elaborazione del linguaggio. I nostri risultati suggeriscono che i processi predittivi supportano la comprensione del linguaggio nonostante le differenze di pronuncia tra parlanti, aprendo la strada a ulteriori sviluppi sul ruolo della predizione nell'elaborazione del linguaggio in contesti naturalistici.

Introduction

Cognitive science has long been dominated by bottom-up frameworks, where the brain was conceptualized as a passive system that progressively transforms sensorial information into coherent mental representations. According to this view, information processing occurs in a hierarchical, step-by-step manner, where each stage of processing builds upon the previous one, progressively integrating lower-level features into higher-order representations (Fodor, 1983; Marr, 1982). Over the past few decades, cognitive science has undergone a profound paradigm shift, moving away from models that equate cognitive processes with computer-like operations, and emphasizing the predictive and inferential nature of cognition (Hutchinson & Barrett, 2019). The development of 'predictive coding' and Bayesian brain theories within the domain of visual perception represented the cornerstone of this change of perspective. This approach defines the brain as an active inference machine that continuously generates predictions about incoming stimuli (Clark, 2013; Friston, 2005; Hohwy, 2013; Rao & Ballard, 1999; Vilares & Kording, 2011). Perception is conceptualized as a form of hypothesis testing in uncertain conditions, where internal models built on prior experiences are used to make inferences about what is not (yet) present in the environment. The brain engages in a continuous cycle of comparing predicted and actual sensory input, using prediction errors to progressively refine its internal models (Clark, 2016) and optimize information processing (Bubic et al., 2010). The 'predictive brain' framework rapidly extended beyond the domain of visual perception, influencing theories of motor control (Adams et al., 2013), social cognition (Tamir & Thornton, 2018) and even consciousness (Hohwy & Seth, 2020).

The notion of prediction always had a controversial status in psycholinguistic literature, often leading researchers to take opposite stances, despite being implicitly present in classical theories of language comprehension. Influential parsing models, such as left-corner parsers (Crocker, 1999) and the garden-path model (Frazier & Fodor, 1978; Frazier & Rayner, 1982; Rayner et al., 1983), suggest that comprehenders anticipate syntactic nodes before encountering the corresponding lexical items. Nonetheless, debates about the possibility of predicting specific words, at least in highly constraining contexts, have been central to psycholinguistic theory for decades. Some authors argued that predicting specific words would be an unnecessary waste of resources due to the large number of possible sentence continuations (Forster, 1981; Jackendoff, 2002; see also Van Petten & Luka, 2012). Despite this initial skepticism, the widespread popularity of the 'predictive brain' framework over the past two decades has reignited interest in the role of prediction in language comprehension.

Psycholinguistic research has accumulated substantial empirical evidence supporting the idea that language comprehension involves the context-based pre-activation of upcoming words (Altmann & Mirković, 2009; Dell & Chang, 2014; Kutas et al., 2011; Pickering & Gambi, 2018; Pickering & Garrod, 2007, 2013). Nevertheless, it remains unclear how detailed these predictions are, especially considering that we rarely pronounce words in exactly the same way, and this variability is even greater across different speakers (i.e., lack of invariance; Liberman et al., 1967). For instance, in everyday life, we might interact with speakers whose speech doesn't align with native phonological and phonetic norms, such as non-native speakers and aphasic patients. Understanding the extent to which comprehenders can predict sublexical information has significant implications for theories of language processing. Prominent models of language comprehension propose that highly constraining contexts enable the pre-activation of word meaning, grammar and phonological form before the onset of forthcoming words (Huettig et al., 2022; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Pickering & Garrod, 2013). Prediction of phonological information has been proposed to play a relevant role in language learning, with some proposals suggesting that prediction errors may help children to unlearn overgeneralizations (Gambi et al., 2018; Rabagliati et al., 2016). However, the evidence supporting prediction of phonological information remains largely inconsistent, even across studies using the same experimental paradigm (Ito, 2024).

Moreover, research in the field of speech perception has shown that listeners deal with the variability of the speech signal by relying on adaptation mechanisms (Weatherholtz & Jaeger, 2016). Understanding the extent to which comprehenders can predict linguistic input requires taking into account concurrent cognitive processes that may influence or interact with predictive mechanisms. Exploring how speaker variability influences predictive processing is particularly relevant in the context of an increasingly multicultural society, where individuals are frequently exposed to diverse linguistic backgrounds.

The work presented here employed behavioral and electroencephalography (EEG) paradigms to answer these (related) questions: Do we predict phonological information? What is the impact of speaker variability on predictive processing? Specifically, Studies 1 and 2 investigated whether comprehenders anticipate the specific phonological form of a predictable word when it is spoken by native versus foreign-accented speakers. We implemented an experimental paradigm that controlled for sentence-context processing demands across accent conditions, while using face stimuli to cue the speaker's identity and, consequently, their phonological features. In Study 1, participants performed a lexical decision task, whereas in Study 2, we examined the Event-Related Potentials (ERPs). Study 3 is currently ongoing and

aims to investigate how listeners deal with speaker variability in native speech, and whether this, in turn, influences prediction during language comprehension. In this study, we manipulated participants' exposure to speakers by presenting narrative stories produced either by a single speaker or by multiple native speakers. We plan to apply Temporal Response Function (TRF) analysis to model how the brain encodes stimulus features related to both speech perception and predictive processing across listening conditions.

The thesis is structured as follows:

- Chapter 1 introduces the concept of prediction in language comprehension, exploring why the brain engages in predictive processing and presenting the main theoretical accounts of linguistic prediction.
- Chapter 2 reviews the empirical evidence for prediction of phonological information and examines how the human brain deals with speaker variability.
- Chapter 3 presents Study 1.
- Chapter 4 presents Study 2.
- Chapter 5 presents Study 3.
- The final section summarizes the main findings and their relevance to current debates in the literature.

Chapter 1 – Prediction in language comprehension

1.1. What is prediction in language comprehension?

As mentioned in the Introduction, the brain does not operate as a passive processor of information; rather, it functions as an active inference machine, constantly generating predictions about incoming stimuli (Clark, 2013; Friston, 2005; Hohwy, 2013; Rao & Ballard, 1999; Vilares & Kording, 2011). Psycholinguistic research has shown that the brain integrates available information to infer what will come next in a conversation or while reading (Kuperberg & Jaeger, 2016; Kutas et al., 2011). But what do we mean by prediction in language comprehension? What sources of information guide these anticipations? What kinds of linguistic representations are predicted? The following sections will address these questions.

1.1.1. Sources of information used to predict

One source of information used to predict is the linguistic input itself, which contains various types of cues that can guide prediction. First of all, the sentence structure constrains the syntactic categories that can follow. For instance, in English, the occurrence of a correlative conjunction such as *either* signals that a coordinate phrase will follow. Staub & Clifton (2006) leveraged the syntactic constraint imposed by *either* to demonstrate that comprehenders use the structural cues within a sentence to anticipate upcoming input. In their study, the eye movements of readers were recorded while they were reading sentences in which two noun phrases or two independent clauses were connected by the word *or* (NP-coordination and S-coordination, respectively). The word *either* may or may not be present earlier in the sentence. The authors observed that the presence of *either* was associated with faster reading times for the material following *or*, regardless of sentence type. Moreover, the absence of *either* led readers to misinterpret the S-coordination structure as an NP-coordination structure. These results suggest that the presence of *either* allows predicting the arrival of a coordination structure, facilitating the processing of this structure and helping readers to not misanalyze S-coordination structures.

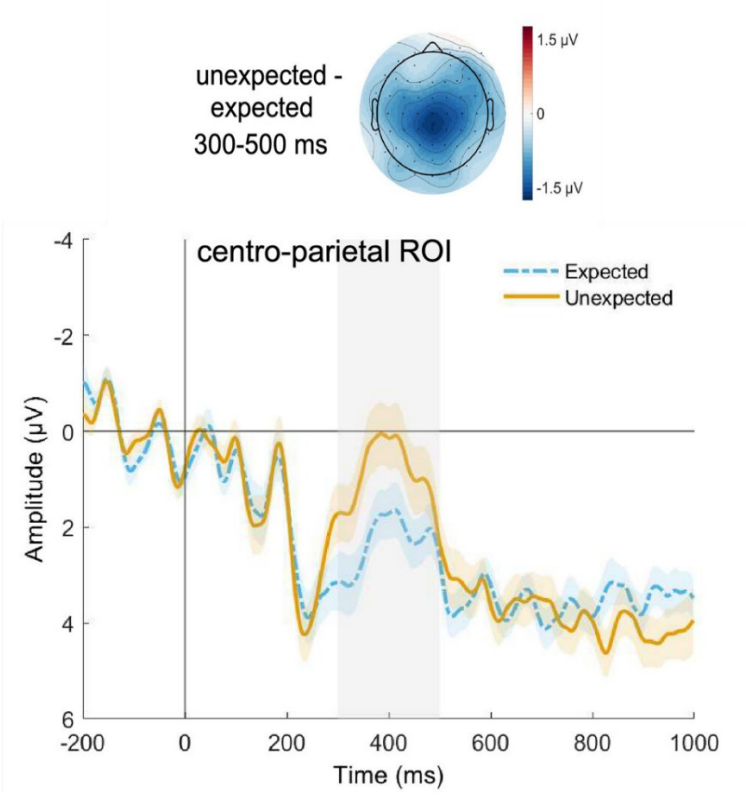
The processing of upcoming words is also influenced by the semantic information presented so far. For example, in sentences like (1a) the lexical-semantic properties of the context (i.e., “farmer” and “milks”) narrow the range of possible continuations, creating a strong expectation for the last word of the sentence (“cow”). In contrast, in sentence (1b), a broader range of continuations is compatible with the lexical-semantic properties of the context, leading to less precise predictions.

(1a) The farmer milks a cow

(1b) The child draws a cow

In experimental settings, the predictability of a target word is often operationalized in terms of the word's *cloze probability* using a cloze task (Taylor, 1953). In this task, a group of participants is asked to guess the next word from a sentence frame. The word's cloze probability is defined as the proportion of individuals who provide that word, assuming that higher cloze probability reflects greater word predictability. Behavioral research using paradigms such as lexical decision (Fischler & Bloom, 1979; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985) and naming tasks (Forster, 1981; McClelland & O'Regan, 1981; Stanovich & West, 1983; Traxler & Foss, 2000) showed that response times are faster for semantically predictable words compared to unpredictable words. Eye-tracking research showed that when a word is predictable from the context, readers are more likely to skip it or spend less time fixating on it (Balota et al., 1985; Ehrlich & Rayner, 1981; Smith & Levy, 2013; see Staub, 2015, for a review). In the Event-Related Potentials (ERP) literature, most studies investigating prediction in language comprehension focused on the N400 component. The N400 is a negative-going, centro-parietally distributed component of the ERP, peaking around 400 ms after word onset and strongly associated with lexical-semantic processing (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984). The N400 is usually observed in response to open-class words, with reduced amplitude for high frequency words and semantically primed words (Kutas & Van Petten, 1988). Within sentences, the N400 is larger for semantically anomalous words than for congruent ones (Kutas & Hillyard, 1980). The N400 is elicited even by semantically congruent words, and its amplitude shows a strong negative correlation with cloze probability (Kutas & Hillyard, 1984), such that more predictable words elicit smaller N400 responses (Fig. 1.1).

Figure 1.1. The N400 is larger for unpredictable yet semantically congruent sentence completions compared to predictable completions (negative voltage plotted upwards). Figure adapted from Hodapp & Rabovsky (2021), licensed under CC BY-NC 4.0.



Further research has shown that prediction does not rely only on the sentence context, but also takes into account broader discourse-level information. Nieuwland & Van Berkum (2006) showed that given a suitable discourse context (e.g., a story about an amorous peanut), predicates that violate animacy (e.g., "the peanut was in love") are processed more easily than canonical predicates (e.g., "the peanut was salted"). Converging evidence for the use of sentential and discourse information for prediction comes from studies using the visual world paradigm (Altmann & Kamide, 1999; Kaiser & Trueswell, 2004; Kamide et al., 2003). In this paradigm, participants listen to an utterance while looking at a display showing semi-realistic scenes or pictures, while their eye movements are recorded. In studies of language prediction, predictability effects emerge as anticipatory eye movements toward a predictable object. Altmann & Kamide (1999) presented participants with semi-realistic visual scenes depicting, for instance, a boy, a cake, and some toys while they listened to sentences such as "The boy will move the cake" or "The boy will eat the cake". Their results showed that eye movements towards the cake (the only edible object that could be eaten) started significantly earlier in the "eat" condition than in the "move" condition. Visual world paradigm studies provided evidence

suggesting that listeners can use many types of linguistic information for prediction, including case-marking (Kamide et al., 2003) and prosody (Weber et al., 2006).

The sources of information used to predict extend beyond the linguistic input, including visual information (Huettig et al., 2022). This relates to a limitation of the visual world paradigm, which implies presenting speech alongside visual input. Huettig et al. (2011) suggested that language-mediated eye movements observed in the visual world paradigm may reflect continuously updated mental representations derived from both the linguistic and the visual input. Supporting empirical evidence indicates that listeners can use visual information to disambiguate sentence structures (Knoeferle et al., 2005; Tanenhaus et al., 1995). Furthermore, listeners exploit multimodal cues to facilitate the processing of upcoming speech, as the information conveyed by visual speech provides cues about the timing of the incoming acoustic signal and its content (Peelle & Sommers, 2015). The integration of auditory and visual speech cues helps listeners to achieve robust speech perception, especially in the presence of noise or signal degradation (Gosselin & Gagné, 2011; Tye-Murray et al., 2007).

1.1.2. Levels of representation

In the previous section, we focused on the sources of information used for prediction, but what is the actual content of these predictions? What types of representations does the brain anticipate during language processing?

The idea that comprehenders can predict syntactic categories was central to early parsing models, such as left-corner parsers (Crocker, 1999) and the garden-path model (Frazier & Fodor, 1978; Frazier & Rayner, 1982; Rayner et al., 1983), and it continues to be a prominent feature of contemporary information-theoretic models (Futrell et al., 2020; Hale, 2001, 2016). Syntactic prediction is supported by empirical evidence showing that the parser can generate syntactic nodes before encountering the corresponding lexical input (Ferreira & Qiu, 2021, for a review). A well-known example of syntactic prediction involves the phenomenon of garden-pathing during language comprehension (Frazier & Fodor, 1978; Frazier & Rayner, 1982; Rayner et al., 1983), which reflects the brain's tendency to interpret sentences incrementally, formulating hypotheses about how words will combine before encountering lexical items that clarify the sentence structure. In this context, the main challenge for the parser is ambiguity: the possibility of applying multiple syntactic structures to the input. According to the garden-path model, the parser favors the simpler grammatical structure, a strategy known as the *Minimal Attachment principle*. A classic example of garden-pathing is the sentence *"The horse raced past the barn fell".* Upon reading "*raced*", the parser assumes that it is the main verb of

the sentence, leading to the interpretation that *"The horse raced past the barn"* is a complete clause. However, this interpretation becomes untenable upon encountering *"fell"*, which requires a subject. The correct reading reveals that *"raced past the barn"* is a reduced relative clause, "*that was raced past the barn*", modifying *"the horse"*, while *"fell"* serves as the main verb. The garden-path effect is reflected in slower per-word reading times (Ferreira & Clifton, 1986; Garnsey et al., 1997; MacDonald et al., 1992; Spivey-Knowlton et al., 1993) and reduced comprehension accuracy (Ferreira et al., 2001; Ferreira & Qiu, 2021). Crucially, the strength of the garden path effect varies continuously and is strongly influenced by the predictability of the intended parse given the preceding context (Kuperberg & Jaeger, 2016). Additionally, subsequent research has provided evidence for the prediction of specific syntactic categories, including coordinate clauses (Staub & Clifton, 2006), direct objects (Arai & Keller, 2013), reduced relative clauses (Arai & Keller, 2013), and null forms (Lau et al., 2006).

A large body of empirical evidence suggests that prediction in language comprehension encompasses the semantic features of upcoming words. One of the earliest pieces of evidence for semantic prediction comes from ERP studies using the *related anomaly paradigm*, introduced by Martha Kutas (Kutas et al., 1984). This paradigm contrasts different types of sentence completions: high-cloze congruent words, contextually anomalous words, and contextually anomalous words that are semantically related to the expected (congruent) completion. Studies using this paradigm have observed that related anomalies elicit a larger N400 than congruent endings, but substantially smaller N400 than unrelated anomalies (Federmeier et al., 2002, 2007; Federmeier & Kutas, 1999a, 1999b; Kutas et al., 1984). These findings suggest that the sentence context enables the pre-activation of semantic features of upcoming words, as reflected in the reduced N400 amplitude for semantically related anomalies (Kutas et al., 1984). Further support for semantic prediction comes from visual world paradigm studies, which have shown that verbs that constrain the semantic content of likely continuations lead to an increase in anticipatory looks towards semantically plausible objects (Altmann & Kamide, 1999, 2007). There is also evidence that semantic pre-activation can extend to perceptual representations. Rommers et al. (2013) explored whether visual features are pre-activated during language comprehension using both the visual world paradigm and ERPs. In the visual world experiment, participants listening to highly constraining sentences (e.g., "*In 1969 Neil Armstrong was the first man to set foot on the…*") tended to fixate the target picture (e.g., "*moon*") as well as a shape competitor (e.g., "*tomato*") more often than unrelated control objects (e.g., "*rice*"). This finding was supported by ERP results, which showed a reduced N400 response to sentence-final words for shape competitors compared to unrelated words.

Prediction in language comprehension extends not only to semantic features but also to specific lexical items. Thornhill & Van Petten (2012) conducted an ERP study to investigate whether it is possible to identify different electrophysiological correlates for semantic and lexical predictions. Participants read sentence frames that could end with a high-cloze word, a similar-meaning alternative, or a semantically unrelated but still congruent word. In line with previous studies (Federmeier et al., 2002, 2007; Federmeier & Kutas, 1999b, 1999a; Kutas et al., 1984), they observed that a smaller N400 was elicited by words semantically related to a more predictable completion compared to unrelated words. They also observed an anterior positivity elicited by unpredictable words, regardless of their semantic relationship with the high-cloze word. The authors interpreted these findings as suggesting that while the N400 is sensitive to the conceptual overlap between the eliciting word and prior context, the anterior positivity is sensitive to disconfirmed lexical predictions. Other authors focused on ERP modulations elicited by a prenominal article to provide evidence for lexical prediction. A frequent criticism of studies interpreting N400 effects elicited by nouns as evidence for prediction is that these effects could reflect easier integration of predictable words within the sentence context compared to unpredictable words (Kutas & Federmeier, 2011; Nieuwland et al., 2020). Investigating ERP responses elicited by prenominal articles that do not differ in meaning allows researchers to rule out the influence of integration processes. Several studies exploited the grammatical relationship between the prenominal article and the predicted noun to provide evidence for lexical prediction. For instance, research in languages that encode grammatical gender using morphemes (e.g., feminine vs masculine) has shown that adjectives and articles mismatching in morphological gender with a predicted word elicit different ERP responses (Ito et al., 2020; Van Berkum et al., 2005; Wicha et al., 2003; but see also Nieuwland et al., 2020).

Whether people can predict phonological information remains a topic of considerable debate. Several studies, employing different experimental paradigms and methods, have investigated whether sublexical features of predictable words can be anticipated. DeLong et al. (2005) provided the first empirical evidence of phonological prediction. The authors leveraged the English phonological rule in which the indefinite article is realized as "*a"* before consonant-initial words and "*an"* before vowel-initial words (e.g., "*a kite*" and "*an airplane*"). Participants read sentences with varying levels of contextual constraint that led to expectations of either a consonant- or vowel-initial word. For instance, given the sentence *'The day was breezy so the boy went outside to fly…'*, the most likely continuation was *'a kite'*. However, the sentence could also continue with a plausible, though less likely, alternative such as *'an airplane'*. The

authors investigated the modulation of the N400 amplitude elicited by both the target word and the preceding article. Since prenominal articles do not differ in meaning, this allowed them to disentangle prediction effects from the influence of integration processes. Results showed that N400 amplitude decreased as a function of increasing *cloze probability*, both for the target noun and, critically, for the preceding article. The authors interpreted this result as evidence that participants predicted the phonological form of the upcoming word, leading to increased negativity when the article mismatched the expected form. Despite the impact of the DeLong et al. (2005) findings, subsequent research has yielded conflicting results, using both the same and different paradigms (Ito, 2024). A more detailed discussion of the empirical evidence supporting phonological prediction is provided in Chapter 2.

1.1.3. All-or-nothing or probabilistic prediction

The notion of prediction during language comprehension has considerably evolved over time, alongside the empirical evidence for linguistic prediction. In its earlier formulations, prediction was often seen as a categorical, all-or-nothing mechanism: the system either fully commits to a specific anticipated outcome or no prediction occurs at all. According to this view, if the predicted structure, word, or meaning turns out to be incorrect, the system must completely revise its initial interpretation. References to an all-or-nothing view of prediction can be found in early theories of garden-path processing, which posited that sentence comprehension involves committing to a single structural interpretation of the sentence, usually the simplest one. The same stance was later adopted by early theories of lexical-semantic prediction, where prediction was defined as a strategic and cognitively demanding process (Becker, 1980, 1985; Neely et al., 1989; Posner & Snyder, 1975). This approach to prediction contributed to the initial skepticism among linguists and psycholinguists regarding the utility of anticipating linguistic input, given the variability of possible sentence continuations (Jackendoff, 2002).

In recent years, the notion of prediction has evolved from a rigid, all-or-nothing phenomenon limited to highly constraining contexts to a more dynamic, probabilistic process that operates continuously during language comprehension (Heilbron et al., 2022). Contextual information has a graded influence on processing forthcoming linguistic information, and the strength of predictions reflects the estimated probability of a given word, meaning, or structure. Compelling evidence for probabilistic prediction comes from studies that, instead of relying on cloze probability as a measure of predictability, used text corpora to calculate objective measures of word probability given the preceding context. A widely used information-theoretic measure of predictability is *word surprisal*, which reflects how unexpected a word is in a given

context. This measure is computed as the negative logarithm of the word's conditional probability, with less predictable words yielding higher surprisal values (Hale, 2001; Levy, 2008). Several studies observed that surprisal is a good predictor of measures of processing difficulty such as reading times (Boston et al., 2008; Demberg et al., 2013; Demberg & Keller, 2008; Frank & Bod, 2011; McDonald & Shillcock, 2003; Smith & Levy, 2013) and N400 amplitude (Frank, 2024; Frank et al., 2015; Li & Futrell, 2023; Lindborg et al., 2023; Michaelov et al., 2024). These studies offer robust empirical support for probabilistic models of prediction, underscoring the role of graded expectations in language comprehension.

1.2. Why do we predict?

In the previous sections, we reviewed extensive empirical evidence showing that the brain leverages different types of information to anticipate linguistic input at different levels of representation. However, what is the function of linguistic prediction? In this section, we will explore different proposals on the role of prediction in language comprehension.

1.2.1. Prediction for learning

One prominent proposal in the literature is that prediction plays a crucial role in learning (Kuperberg, 2016), including language acquisition (Gambi et al., 2018; Rabagliati et al., 2016). This perspective addresses earlier criticism that regarded prediction as an inefficient strategy, given the infinite number of possible continuations of a sentence (Jackendoff, 2002). The shift towards probabilistic models of prediction has led psycholinguists to reframe prediction errors, seeing them not as detrimental to comprehension, but rather as opportunities for learning. According to this view, prediction errors refine linguistic knowledge by aligning internal representations with the statistical regularities of the language (Dell & Chang, 2014). This aligns with comprehensive accounts of human adaptive behavior, such as Bayesian brain (Doya et al., 2007) and predictive coding (Clark, 2013). During language comprehension, the brain makes probabilistic hypotheses about the upcoming linguistic input, and possible discrepancies are used to revise the prior probability distribution at a given level of representation, reducing prediction errors and implementing a form of implicit learning. Some authors have proposed that the N400 can be considered an electrophysiological correlate of belief updating, reflecting the brain's response to prediction errors and revision of prior expectations during language comprehension (Kuperberg, 2016; Nour Eddine et al., 2024). This contrasts with interpretations of the N400 as reflecting integration processes, where the component is assumed to index the degree of (mis)match between the context and the current word and/or the ease with which the

current word can be integrated with prior contextual information (Kutas & Federmeier, 2011; Petten, 1993; van Berkum et al., 1999; Van Petten & Luka, 2012). Although prediction and integration have often been viewed as competing accounts of the N400, some researchers have argued that the two views are not mutually exclusive. The N400 response might arise from the combined activity of multiple processes, involving both prediction and integration (Baggio, 2012, 2018; Baggio & Hagoort, 2011; Kutas & Federmeier, 2011; Newman et al., 2012; Pylkkänen & Marantz, 2003). Emerging evidence supports the view that the N400 reflects distinct yet parallel processes, which are associated to partially dissociable subcomponents (Nieuwland et al., 2020).

One issue that is difficult to untangle is whether prediction drives learning or vice versa, especially when it comes to language acquisition. According to prediction-based accounts of learning, predictions should be highly detailed for driving learning, as this allows abstract representations to be grounded in a format that can be directly compared with incoming sensory input. In particular, the prediction of phonological forms has been proposed to play a relevant role in language learning (Gambi et al., 2018). For example, children may unlearn overgeneralizations by predicting "mouses" (as the plural form of mouse) and update their mental representations when hearing "mice" (i.e., when encountering a prediction error signal). Gambi et al. (2018) used the visual world paradigm to test whether 2- and 5-year-old children are able to predict semantic and phonological information. They observed that 2-year-old children can generate expectations about meaning based on a determiner (e.g., "Can you see one...ball/two...ice creams?"). However, not even at 5 years old, children are able to predict the phonological form of upcoming words based on a determiner (e.g., "Can you see a...ball/an...ice cream?"). In a subsequent study, Gambi et al. (2021) investigated the relationship between linguistic knowledge and revision skills (i.e., the ability to make use of prediction error) in 2- and 5-year-old children. The authors observed that the development of revision skills is shaped by language knowledge, rather than vice versa. They also observed that prediction skills and processing speed are significant predictors of linguistic knowledge. Taken together, these studies suggest that prediction may facilitate learning by enhancing processing speed, although it is not a necessary mechanism for learning to occur.

1.2.2. Prediction for improving processing speed and efficiency

Another key function of anticipatory language processing might be related to the constraints imposed by the linguistic input on comprehension. The speech signal is highly transient, unfolding at a rate of approximately 10 to 15 phonemes per second – equivalent to roughly 5 to

6 syllables per second or 150 words per minute (Studdert-Kennedy, 1986) – with acoustic information being lost after 50 ms (Remez et al., 2010). The same applies to sign language perception, where American Sign Language (ASL) syllables last about a quarter of a second (Wilbur & Nolkn, 1986), and visual information fades within 60-70 ms (Pashler, 1988). Moreover, the human brain has a limited ability to retain sequences of auditory (Warren et al., 1969) and visual input (Wilson & Emmorey, 2006). Christiansen & Chater (2016) proposed an integrated framework in which the organization of the language processing system is strictly related to the general principles of perception and memory. Perceptual and memory constraints give rise to the *Now-or-Never bottleneck*: linguistic information needs to be rapidly and efficiently processed; otherwise, it will be lost or overwritten by new input. For this reason, our language system rapidly encodes the input into "chunks," which are immediately "passed" to a higher level of linguistic representation. This "Chunk-and-Pass" mechanism optimizes information processing by encoding larger and larger stretches of input starting from lower levels of representation and avoiding interferences between local units of information at the sensory level. Crucially, prediction provides an opportunity to begin the chunking process as early as possible, allowing higher-level representations to constrain the processing of the input at lower levels. Chunk-and-Pass processing implies that it is not possible to go back once a chunk is created, since the backtracking of information would interfere with processing the incoming sensorial input. When sufficient contextual cues are available, the system can anticipate upcoming input to efficiently resolve local ambiguities and thereby speeding up processing while minimizing potential interferences. Several studies have shown that prediction can aid speech perception (Arnal & Giraud, 2012; de Lange et al., 2018). Humans can draw on prior experience to anticipate how a speaker is likely to pronounce a given word, thereby facilitating the mapping between the speech signal and linguistic categories (Kleinschmidt & Jaeger, 2015). Furthermore, research has shown that words with a missing phoneme occurring in highly constraining contexts are processed similarly to intact words, as predictive mechanisms allow the restoration of the omitted information (Groppe et al., 2010; Samuel, 1996; Sivonen et al., 2006). Finally, predictive processing has been shown to increase the intelligibility of degraded speech (Hakonen et al., 2017). These findings suggest that prediction enhances processing speed and efficiency. Moreover, it may be helpful in the perception of degraded or ambiguous speech stimuli.

1.2.3. Prediction for coordinating the dialogue

Linguistic prediction may play a role in coordinating dialogue and facilitating mutual

understanding between speakers. Conversation is widely regarded as the core context of language use, representing the primary mode of language acquisition for children and, in some cultures, the sole medium of linguistic interaction (Clark & Wilkes-Gibbs, 1986). Successful communication requires coordinating both the content and timing of what is meant and understood. Furthermore, speakers use language to perform communicative acts – such as asserting, requesting, or promising – with the expectation that their interlocutors will recognize and appropriately respond to their intentions (Austin, 1962; Grice, 1957, 1968; Schiffer, 1972; Searle, 1969). Even though conversation requires performing several tasks simultaneously, humans manage turn-taking in dialogue with striking efficiency. For instance, in Dutch, it has been observed that about half of turn-taking role transitions are exchanged within -250 and + 250 ms from the end of the current turn (de Ruiter et al., 2006). This finely tuned coordination across speakers has been observed across a wide range of languages and cultures (Stivers et al., 2009). Pickering & Garrod (2007) observed that in conversation, an utterance often constrains the set of appropriate responses. They argue that such constraints help align the mental states of the interlocutors, facilitating smooth and coordinated exchanges (see also Pickering & Garrod, 2004). Anticipating what a conversational partner will say might enable interlocutors to align their mental states. Furthermore, prediction and imitation in conversation seem to be tightly coupled, making conversation effortless even though it requires performing multiple tasks simultaneously (Pickering & Garrod, 2007). Evidence for this perspective comes from studies showing that conversational partners align their behavior during interaction, for example, by synchronizing their breathing (Pardo, 2006), syllable speech rate (Wilson & Wilson, 2005), and adopting similar pronunciation patterns (Pardo, 2006). Furthermore, there is evidence suggesting that people can estimate the timing of turn-endings based on the content of their predictions. Magyari & de Ruiter (2012) conducted a study in which participants were asked to guess the final words of a conversation turn. To estimate how easily turn endings could be predicted, they used data from a previous study in which participants had to press a button exactly when a turn ended. They observed that participants were more accurate in completing sentences when the turn ending was easier to anticipate.

Some researchers have questioned whether prediction is truly essential for dialogue coordination, observing that much of the supporting evidence comes from experimental settings in which completing a conversational partner's utterance may be unusually straightforward (Huettig, 2015). Moreover, people make predictions even when not engaged in conversation, for example, while reading texts or listening to stories. Although it remains unclear whether prediction evolved primarily to facilitate dialogue, available empirical evidence suggests that it

supports turn-taking and enables more efficient responses during conversation.

1.3. Theoretical models of linguistic prediction

Despite the growing body of research supporting the involvement of predictive processes in language comprehension, there is still considerable debate concerning the cognitive mechanisms underlying linguistic prediction. The following sections aim to provide an overview of the most influential theoretical accounts of prediction in language comprehension.

1.3.1. Prediction as passive spreading of activation

Passive spreading of activation across representations has been proposed as a core mechanism in influential models of language comprehension (e.g., Huettig et al., 2022). This perspective draws from early models of semantic processing, which sought to explain *semantic priming* effects (Balota & Lorch, 1986; Meyer & Schvaneveldt, 1971; Neely, 1976, 1977; Tweedy et al., 1977). Meyer & Schvaneveldt (1971) provided the first empirical evidence of semantic priming using a lexical decision task, showing that participants responded more quickly when the target word was preceded by a semantically related prime (e.g., *coffee–cup*) compared to an unrelated pair (e.g., *dog–cup*). Collins & Loftus (1975) proposed that concepts are represented as nodes within a semantic network. The activation of a concept spreads to related nodes following a gradient of decreasing activation strength as a function of semantic distance. Lexical representations are stored in a separate network that is organized according to phonemic similarity and connected to the semantic network. According to this view, passive spreading of activation from a related prime facilitates word recognition by making it easier to reach the activation threshold necessary for identifying the target. Within sentences, a semantically rich context (e.g., "The farmer milks a…") will rapidly activate a set of associated concepts (e.g., "cow"). Given a sufficiently constraining context, activation will accumulate for a specific lexical item, resulting in pre-activation even to the level of phonological (Huettig et al., 2022) or orthographic (Kim & Lai, 2012; Molinaro et al., 2013) features. Moreover, the activation of a lexical item (e.g., "book") can spread to phonologically similar items (e.g., "hook") due to connections within the lexical network, a mechanism known as phonemic priming (Hillinger, 1980). This would all occur automatically, as a function of the relational links that the comprehender has formed in lexical-semantic memory.

In their seminal study, Kutas & Hillyard (1984) observed that the amplitude of the N400 component is inversely related to the participant's expectancy for the eliciting word, as measured by cloze probability. Crucially, in highly constraining contexts, they also examined

whether the N400 amplitude elicited by congruent low-cloze words was sensitive to the semantic relationship between the eliciting word and the expected best completion. They found that the N400 amplitude varied according to the degree of semantic relatedness: larger N400s were elicited by words that were unrelated to the best completion. These findings provided the basis for interpreting predictability effects as the result of passive spreading of activation between conceptual representations. More recently, Huettig et al. (2022) proposed a theory of prediction based on the Parallel Architecture framework (PA: Jackendoff, 2002; Jackendoff & Audring, 2020), where passive spreading of activation plays a central role in the prediction of upcoming words. The PA defines phonology, syntax, and semantics as systems that operate in parallel while also mutually influencing each other. This cognitive architecture includes an extended lexicon encompassing not only individual words, but also idiomatic expressions, collocations, meaningful constructions, and grammatical rules (defined as schemas). The authors distinguish between two different mechanisms through which linguistic representations can be pre-activated: *between-item prediction* and *within-item prediction*. Between-item prediction occurs when the activation of a lexical item spreads to similar or semantically related words. Instead, within-item prediction occurs when incoming input activates lexical items whose beginning (or incipit) matches the input encountered so far. The incipit of a word can automatically pre-activate multiple possible lexical candidates based on a pattern completion mechanism (Falandays et al., 2021). Therefore, the generation of predictions can be considered a natural byproduct of language processing, as it is part of the lexical access process (similarly to Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980). This view is compatible with current probabilistic models of language processing, since prediction is defined as the relative activation strength of a linguistic representation compared to that of competing candidates.

1.3.2. Integrated models of linguistic prediction

Integrated models of linguistic prediction acknowledge the role of passive spreading of activation, while also emphasizing the pro-active aspect of internal prediction generation. In this context, prediction-by-production accounts of linguistic prediction have received particular attention. This theoretical perspective is based on empirical evidence showing several commonalities in the processes and representations involved in language comprehension and language production (AbdulSabur et al., 2014; Dell & Chang, 2014; Gambi & Pickering, 2017; Gastaldon et al., 2020, 2023, 2024; Okada & Hickok, 2006; Pickering & Gambi, 2018; Pickering & Garrod, 2014; Silbert et al., 2014). For instance, Martin et al. (2018) conducted an

experiment where Spanish-speaking participants read highly constraining sentence contexts followed by either expected or unexpected words differing in grammatical gender (e.g., *El rey llevaba en la cabeza una corona/un sombrero*; "The king wore on his head a crown/a hat"). Unexpected words elicited a larger N400 compared to predictable words, and articles mismatching in grammatical gender with the expected word also elicited an N400 predictability effect, which was considered a marker of lexical prediction. Crucially, the authors found that the N400 predictability effect elicited by pre-nominal articles was eliminated when participants simultaneously performed an articulatory suppression task that taxed the production system, suggesting that production processes contribute to prediction. Prediction-by-production models converge on the idea that prediction during comprehension relies on processes traditionally attributed to language production (Dell & Chang, 2014; Federmeier, 2007; Huettig, 2015; Pickering & Gambi, 2018; Pickering & Garrod, 2007, 2013; Pickering & Strijkers, 2024), albeit they do not entirely agree on the exact nature of the mechanisms and representations involved.

Pickering & Garrod (2013) proposed an account of language processing based on the theory of forward modeling in motor control (Wolpert, 1997; Wolpert & Flanagan, 2001), which posits that executing an action implies the generation of an efference copy of the action command. The efference copy is used to create a forward model, which predicts the sensory consequences of the action sequence (predicted percept). The output of the motor command can be compared with the predicted percept, enabling the correction of the action command (or the forward model) if they do not match. Forward models can also be used to predict the outcome of a perceived action (Wolpert et al., 2003). In action perception, the perceiver can covertly imitate the actor to generate a forward action model, enabling the prediction of the sensory consequences of the action and the comparison of the expected percept with the actual percept. Pickering & Garrod (2013) framed language production and language comprehension as forms of action and action perception, respectively. In language production, forward models are used to predict the sensory consequences of a production command and to monitor the outcome by comparing the predicted percept with the actual percept. The authors assumed a traditional serial account of language production (e.g., Levelt, 1989). Therefore, forward production models are generated from higher to lower levels of the linguistic hierarchy: first from semantics to syntax, and then from syntax to phonology. In language comprehension, comprehenders can use covert imitation and forward modelling to generate predictions at multiple levels of representation, a process referred to as *prediction-by-simulation*. Since predictions rely on forward production models, the system is expected to generate predictions for semantic content before syntax, and for syntax before phonological form. Under this

framework, forward models enable the rapid generation of predictions without requiring the full activation of production representations. Although *prediction-by-simulation* is considered the primary mechanism for prediction, comprehenders can also anticipate the linguistic input using associative mechanisms (*prediction-by-association*), which rely on prior experience in language comprehension and do not involve the production system. *Prediction-by-simulation* will be preferred when comprehenders perceive themselves as similar to the speaker, since simulations are more likely to be accurate. Conversely, when the speaker is dissimilar to the comprehender (e.g., when the comprehender is a native adult speaker and the producer is a non-native speaker or a child) simulating the speaker's production becomes more difficult. As a result, prediction-by-simulation may play a reduced role in predicting upcoming speech.

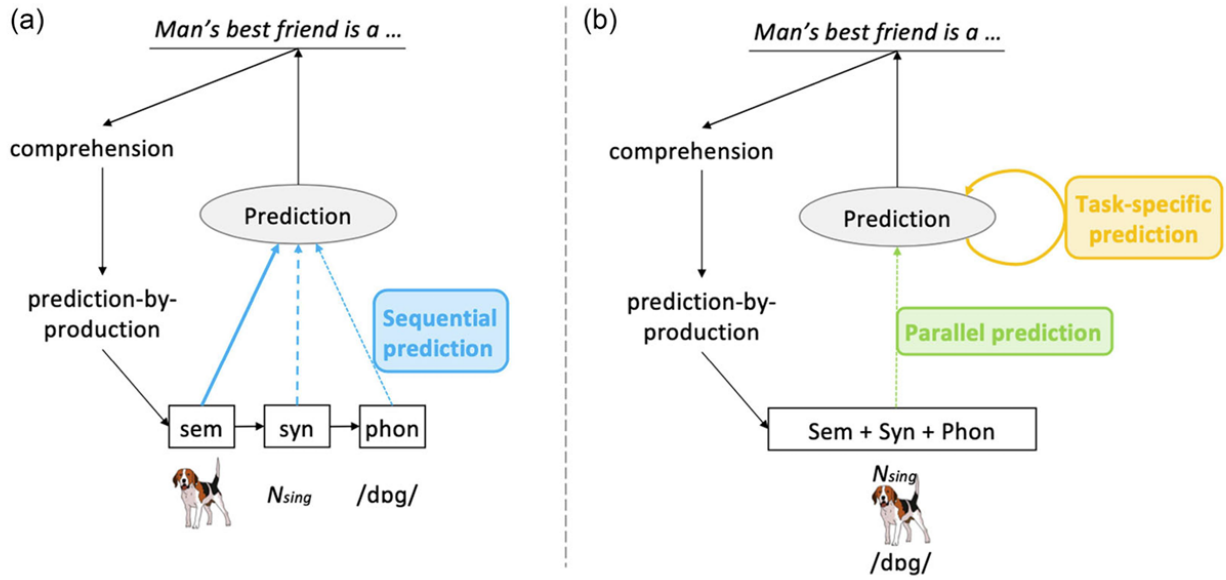
Pickering & Gambi (2018) more clearly distinguished between production and associative mechanisms in linguistic prediction. *Prediction-by-Production* relies on the covert imitation of the linguistic context to derive the underlying intention of the utterance. The system also considers the non-linguistic context, which includes shared background knowledge and shared visual (or other extra-linguistic) information. By doing so, the comprehender can take into account any differences between himself and the speaker. The derived intention is then run through the production system to produce the next part of the utterance. Similarly to Pickering & Garrod (2013), predictions follow a top-down progression from higher to lower levels of the linguistic hierarchy (i.e., from semantics to phonology). However, this process requires cognitive resources, making it optional. Depending on the circumstances, comprehenders might predict semantic and syntactic features without engaging in phonological prediction, as they do not necessarily need to go through all the stages of word production. In contrast, *Prediction-by-Association* is based on the passive spreading of activation between linguistic representations, such as in semantic and phonemic priming. This mechanism is fast and automatic, but it cannot account for the constraints imposed by the linguistic context, leading to predictions that are not always accurate.

Huettig (2015) argued that an adequate explanation of predictive language processing must involve a broader set of mechanisms, referred to as PACS (Production, Association, Combinatorial, and Simulation-based prediction). Under this framework, predictions are generated using different mechanisms that continuously interact with each other. Comprehenders sometimes rely on their production system to anticipate upcoming utterances. The system uses fully-specified production representations to predict the linguistic input, rather than relying on the impoverished representations of a forward model as proposed by Pickering & Garrod (2013). Prediction can also occur through associative mechanisms, which are not

limited to semantic knowledge but also involve phonological, orthographic and even non-linguistic information (Arias-Trejo & Plunkett, 2009, 2013; Federmeier & Kutas, 2001; Ganis et al., 1996; Mani & Plunkett, 2011). Combinatorial mechanisms can contribute to predictive language processing by taking into account various linguistic constraints and constructing complex meanings. Finally, comprehenders can use event simulation to pre-activate linguistic representations, similarly to when mental imagery is used to simulate future events based on past experiences (Moulton & Kosslyn, 2009).

More recently, Pickering & Strijkers (2024) proposed a parallel prediction-by-production model that contrasts with prediction models assuming that production representations are accessed sequentially (i.e., from semantics to phonology) (Pickering & Gambi, 2018; Pickering & Garrod, 2013). Figure 1.2 compares sequential and parallel prediction-by-production models.

Figure 1.2. Prediction-by-production in a sequential (a) versus parallel (b) prediction model. In the sequential model, comprehenders predict meaning first, then syntax, and then phonology, with decreasing strength. In the parallel model, all levels (semantics, syntax, phonology) are activated at once, and then selectively enhanced based on task relevance. Figure from Pickering & Strijkers (2024), licensed under CC BY 4.0.



The proposal by Pickering & Strijkers (2024) is based on empirical evidence showing a similar time-course for the activation of semantic, lexical, and phonological representations in language production (Fairs et al., 2021; Feng et al., 2021; Miozzo et al., 2015; Riès et al., 2017; Strijkers

et al., 2010; Strijkers & Costa, 2016). According to the authors, word learning involves integrating meaning, grammar, and phonology into a single representation. Therefore, different linguistic components of a word can be retrieved simultaneously during language production (Kerr et al., 2023; Strijkers, 2016; Strijkers & Costa, 2016). Pickering & Strijkers (2024) extended this idea to language comprehension, proposing that comprehenders can simultaneously activate all linguistic components of a predicted word. As the different components of a word are available at the same time, comprehenders may flexibly prioritize the prediction of the most relevant information according to task demands.

The literature also includes prediction models suggesting that comprehenders can actively predict the linguistic input without necessarily relying on production mechanisms. Lau et al. (2013) proposed that the pre-activation of upcoming words may result either from *passive spreading of activation* or from *predictive commitment*. The authors define *predictive commitment* as the process by which the internal representation of the context, held within working memory, is updated to include upcoming linguistic information before its onset. Kuperberg & Jaeger proposed a hierarchical generative model aimed at explaining how linguistic information can be pre-activated at multiple levels of representation (Kuperberg, 2016; Kuperberg & Jaeger, 2016). According to this model, prediction relies on internal generative models, defined as sets of hierarchically organized internal representations. Internal representations are continuously shaped by both linguistic and non-linguistic information, and encompass a range of information types, from higher-level semantic representations to sub-phonemic features (Connine et al., 1991; Szostak & Pitt, 2013). According to this view, comprehenders aim to maximize the likelihood of accurately recognizing the incoming language input. To achieve this, internal generative models probabilistically pre-activate information from higher to lower levels of representation. Comprehenders can adapt to diverse linguistic environments by constructing distinct generative models that reflect different statistical regularities (Kleinschmidt & Jaeger, 2015). This framework thus emphasizes prediction as an adaptive mechanism that supports comprehension by integrating prior knowledge with current input, enabling comprehenders to account for the variability of linguistic environments in a goal-directed and probabilistic manner.

Chapter 2 – Phonological prediction and speech variability

2.1. Phonological prediction: a debated mechanism

Research consistently showed that people predict information during language comprehension. Prominent models of language comprehension propose that such predictions can include sublexical information (Altmann & Mirković, 2009; Huettig et al., 2022; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Pickering & Garrod, 2013; Pickering & Strijkers, 2024). Furthermore, prediction-based accounts of language learning propose that prediction of phonological forms might play a key role in language acquisition (Gambi et al., 2018; Rabagliati et al., 2016). However, the extent to which comprehenders can predict phonological forms in highly constraining contexts remains a matter of debate, with current models of language comprehension remaining rather underspecified in this regard (Ito, 2024). Understanding whether and how we predict phonological information has important implications for theories of language processing and learning. Phonological prediction has been mostly investigated using EEG and eye-tracking paradigms, though with mixed results. In the following sections, I will review studies that used these approaches to investigate prediction of phonological forms.

2.1.1. Evidence from EEG studies

In the Event-Related Potentials (ERP) literature, most studies focused on the N400 component to investigate whether highly constraining sentence contexts allow prediction of the phonological form of the upcoming word. The study by DeLong et al. (2005), already introduced in Chapter 1, was considered the first crucial evidence in support of phonological prediction. The authors observed that N400 amplitude decreased as a function of increasing *cloze probability* of the pre-target article, suggesting that participants were predicting the phonological form of the upcoming word. Although these findings significantly shaped theories of prediction, subsequent research has provided mixed evidence for the modulation of the N400 amplitude on the pre-target article (Ito et al., 2017, 2020; Martin et al., 2013, 2018; Nicenboim et al., 2020; Nieuwland et al., 2018). In additional unpublished datasets, the authors did not find a relationship between article-elicited N400 and cloze probability (DeLong, 2009). Furthermore, an alternative analysis of the original dataset did not yield statistically significant results (DeLong, 2009). Martin et al. (2013) found a modulation of the N400 on the pre-target article, similar to that reported by DeLong et al. (2005), in native speakers, but not in bilinguals using their second language. They concluded that bilinguals engage less in prediction as a

consequence of the slower and less accurate processing in the second language (see also Ito et al., 2017). However, this study differed in many methodological aspects from DeLong et al. (2005). Nieuwland et al. (2018) conducted a large-scale, multi-lab direct replication of DeLong et al. (2005). While they replicated the N400 modulation on the target noun as a function of its cloze probability, they failed to observe a significant N400 predictability effect on the pre-target article. The authors also performed additional exploratory analyses, suggesting that the article-elicited predictability effect might not be entirely absent but is likely much smaller than originally reported. In line with this, a more recent meta-analysis of studies that adopted pre-target article manipulations presented evidence in favor of a small predictability effect (Nicenboim et al., 2020). One relevant limitation of the *"a"/"an"* paradigm is the assumption that English indeterminate articles can disconfirm the prediction of an upcoming noun. In the sentence "*The day was breezy so the boy went outside to fly…*", the most likely continuation might be "*a kite*," but a less probable alternative like "*an airplane*" is also plausible. Crucially, the article only provides information about the initial phoneme of the following word, and may not directly refer to the predicted noun itself: for instance, "*an old kite*" would still be consistent with the article "*an*." Given this limitation and the difficulty in replicating pre-target predictability effects, researchers employed different paradigms to investigate phonological prediction.

Laszlo & Federmeier (2009) implemented a paradigm in which participants read sentence frames that could end with a word, pseudoword, or illegal string. Words could be expected or unexpected, whereas the other stimuli were, by definition, unexpected. Unexpected items could be either phonologically/orthographically related to the predictable word or not. The authors observed that unexpected items elicited a larger N400 compared to predictable words. Crucially, the N400 predictability effect was reduced for form-related items compared to unrelated items. The authors interpreted their findings as suggesting that prediction involves the pre-activation of form-related information, facilitating the processing of phonological/orthographic neighbors of expected words. Ito et al. (2016) implemented a similar paradigm, in which participants read sentence frames (e.g., "*The student is going to the library to borrow a…*") ending with either a predictable word (e.g., *book*), a form-related word (e.g., *hook*), a semantically related word (e.g., *page*), or an unrelated word (e.g., *sofa*). They observed that semantically related and form-related words elicited a reduced N400 predictability effect compared to unrelated words. However, the reduced N400 elicited by form-related words was observed only for high-cloze sentences and slow presentation rates (700 ms per word), suggesting that form pre-activation is more limited than meaning pre-activation. The authors

argued that their findings are in line with prediction-by-production accounts of language comprehension. Most of the prediction-by-production models are based on serial accounts of language production (e.g., Levelt, 1989), where production representations are accessed sequentially (i.e., from semantics to phonology). Since prediction in language comprehension relies on processes and representations that are shared with language production, semantic prediction will precede the anticipation of phonological information. Phonological prediction will occur only when sufficient cognitive resources and time are made available by the specific task and paradigm (Pickering & Gambi, 2018). DeLong et al. (2019, 2021) conducted a conceptual replication of Ito et al. (2016) and observed a reduced N400 effect for form-related words at both 500 and 700 ms per word presentation rates. Their findings may indicate that phonological prediction based on production mechanisms is more rapid than previously thought, or that comprehenders rely on mechanisms that are less constrained by time and/or processing resources to pre-activate phonological information.

Phonological prediction has also been associated with ERP (and ERF, Event-Related Fields, for MEG) components that precede the N400. Connolly & Phillips (1994) conducted a study in which participants read sentence frames ending either with a predictable word, a semantically congruent word that did not share the initial phoneme with the expected word, a semantically anomalous word that shared the initial phoneme with the expected word, or a semantically anomalous word that did not share the initial phoneme with the expected word. They observed that phoneme mismatch conditions elicited enhanced negativity compared to predictable words between 150-350 ms after word onset, while semantic mismatch conditions elicited a later enhanced negativity interpreted as a modulation of the N400 amplitude. The authors labeled the early negativity elicited by phoneme-mismatching conditions as Phonological Mismatch Negativity (PMN), suggesting that this component reflects pre-lexical phonological processing and is therefore sensitive to speech sounds that disconfirm phonological predictions. Since the same effect is observed for words and nonwords (Connolly et al., 2001), the PMN has been interpreted as an ERP component distinct from the N400. However, following research has not consistently confirmed the independence of these two components (Lewendon et al., 2020; Newman et al., 2003; Newman & Connolly, 2009). Nieuwland (2019) observed that some studies reported N400 effects beginning around 150-200 ms after word onset (Bölte & Coenen, 2002; O'Rourke & Holcomb, 2002), suggesting that early phoneme mismatch effects may reflect a modulation of the N400 component. Furthermore, early phoneme mismatch effects do not always emerge (Diaz & Swaab, 2007; Poulton & Nieuwland, 2022; Van Berkum et al., 2008), and the topographic distribution of the

PMN does not seem to be consistent across studies. In conclusion, even though some studies report modulations in ERP/ERF components preceding the N400 (Dikker et al., 2009, 2010; Dikker & Pylkkanen, 2011; Kim & Lai, 2012), the replicability of these effects appears weak and inconsistent (for a review, see Nieuwland, 2019).

2.1.2. Evidence from eye-tracking studies

Eye-tracking studies used the visual world paradigm to investigate phonological prediction. Participants usually listen to sentences containing a target word that is predictable from the sentence context while they look at a display showing semi-realistic scenes or pictures. Predictive processing is measured as anticipatory looks toward an object representing the target word. Phonological prediction has been investigated by measuring whether a phonological competitor of the predictable target word (e.g., they share the initial phonemes) attracts more anticipatory looks than phonologically or semantically unrelated distractors (*predictive phonological competitor effect*). Anticipatory looks toward the phonological competitor are taken as evidence of pre-activation of the phonological form of the target word. Ito et al. (2018) used the visual world paradigm to investigate whether native and non-native speakers of English predict the phonological form of the target in highly constraining sentences (e.g., "*mouth*", in *"In order to have a closer look, the dentist asked the man to open his mouth a little wider*"). Participants listened to the sentences while viewing a scene containing either the target (*mouth*), the phonological competitor (*mouse*), or an unrelated object (*socks*). In addition to one of these objects, the scene contained unrelated distractors that were identical across conditions. Participants were asked to judge whether each sentence mentioned any of the objects in the display. The results showed that both native and non-native speakers displayed more anticipatory looks toward the target object compared to the unrelated distractors, and this occurred earlier in native speakers. Importantly, only native speakers showed anticipatory looks toward the phonological competitor. The authors interpreted the results suggesting that while both groups implement semantic predictions, only native speakers have enough cognitive resources available to generate phonological predictions. This interpretation is in line with prediction-by-production models based on serial accounts of language production, where the prediction of semantic information precedes the pre-activation of phonological information. Recently, Pickering & Strijkers (2024) proposed an interpretation of the results reported by Ito et al. (2018) based on a parallel prediction-by-production account of language comprehension. According to this view, the meaning, grammar, and phonology of a predictable word can be pre-activated simultaneously. Since different components of the word are simultaneously

available, comprehenders may flexibly prioritize the prediction of the most relevant information in the current context. The authors suggested that native speakers are more likely to engage in parallel predictions, encompassing both semantic and phonological features, due to stronger memory representations. In contrast, non-native speakers, who could rely on weaker memory representations, would primarily focus on semantic predictions, which are directly relevant to the task.

Following research investigating the *predictive phonological competitor effect* obtained inconsistent results across studies (Angulo-Chavira et al., 2023; Ito, 2019; Ito & Husband, 2017; Ito & Sakai, 2021; Kukona, 2020; X. Li et al., 2022; X. Li & Qu, 2024; Zhao et al., 2024). These studies differed in many aspects of the experimental paradigm and materials, which might have contributed to the inconsistency of the results (for a review, see Ito, 2024). Nonetheless, a recent meta-analysis reported a small but reliable *predictive phonological competitor effect* (Ito, 2024).

2.2. Speech variability and prediction

Leading models of language comprehension propose that individuals are capable of predicting sublexical information during processing (Altmann & Mirković, 2009; Huettig et al., 2022; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Pickering & Garrod, 2013; Pickering & Strijkers, 2024). In this respect, it is important to consider that the speech signal presents substantial variability: we rarely pronounce words in the same way and this variability is even greater across speakers. Research in the field of speech perception has shown that listeners easily understand speech despite its variability (Bradlow & Bent, 2008; Clarke & Garrett, 2004; Clopper & Bradlow, 2008; Maye et al., 2008; Weatherholtz & Jaeger, 2016). However, how do we deal with speech variability? Is the prediction system able to generate speaker-specific predictions? In the following sections, I will focus on how speech perception research addressed the challenges posed by the acoustic variability of speech and present studies examining the relationship between predictive processing and speaker variability.

2.2.1. How do we deal with acoustic variability in speech perception?

Speech perception is broadly assumed to involve the mapping of a continuous and transient acoustic signal into abstract linguistic categories. Although speakers typically understand one another with apparent ease, this process is far from straightforward. The physical properties of speech sounds vary even between the productions of a single speaker, due to both articulatory (Ladefoged, 1980; Lindblom, 1990; Marin et al., 2010; Öhman, 1966; Stevens, 1972) and

psychological factors (Protopapas & Lieberman, 1997). Furthermore, variations in speech production among speakers arise from anatomical differences (Fitch & Giedd, 1999; Peterson & Barney, 1951), as well as from factors such as age (Lee et al., 1999) and gender (Perry et al., 2001). For instance, perceived pitch depends on the vibration rate of the vocal cords, which is known as fundamental frequency (F0). Although F0 varies across productions of the same speaker, average F0 is around 100–120 Hz for adult males and 200–220 Hz for adult females (Simpson, 2009). Furthermore, in everyday life, it's common to interact with speakers whose speech doesn't align with native speech norms. For example, non-native speakers often deviate from both phonological and phonetic native norms due to their reduced familiarity with the language (Best et al., 2001; Clopper et al., 2005; Flege, 1988). Similarly, individuals with aphasia may commit phonological errors in language production due to impairments in phonemic processing or motor speech planning (Ardila, 2010; Blumstein, 1973; Croot et al., 2012; Miceli et al., 1980; Ogar et al., 2005). The lack of invariance of the speech signal represents a significant challenge for models of speech perception, since it prevents a one-to-one mapping between the physical properties of the acoustic input and linguistic categories (Liberman et al., 1967). Research in the field of speech perception focused on investigating how listeners recognize phonemes despite the variability of the speech signal. Earlier approaches to acoustic variability in speech perception assumed that perceptual constancy is achieved through the extraction of category-specific invariants from the speech signal (Cole & Scott, 1974; Fant, 1960; Kewley-Port, 1983). Several studies focused on the identification of acoustic invariants, providing valuable insights regarding the acoustic cues that distinguish the place of articulation of stop consonants (Halle et al., 1957; Stevens & Blumstein, 1977, 1981). Nonetheless, the *acoustic invariance* approach to speech perception has yet to identify category-specific cues for the remaining sounds of the language (Weatherholtz & Jaeger, 2016).

An alternative account of invariance posits that invariant cues originate not from the physical properties of the speech signal, but from the motor processes involved in speech production. According to the Motor theory, speech perception and production rely on shared cognitive and neural mechanisms (Liberman, 1982; Liberman & Mattingly, 1985). Perceptual constancy is achieved by identifying the *invariant motor commands* underlying the production of speech sounds. According to this view, perceiving an utterance requires identifying the speaker's intended articulatory gestures (i.e., the abstract motor plans used to produce speech sounds), which are not directly manifested in the acoustic signal or in the observable articulatory movements (Liberman & Mattingly, 1985). A major challenge for the Motor Theory is explaining how listeners recover intended articulatory gestures from acoustic and visual

information. Subsequent formulations have proposed that listeners rely on observed articulatory gestures to recognize phonemes (Best, 1995; Fowler, 1986; Galantucci et al., 2006; Vroomen & Baart, 2012). However, a critical limitation of this theory is represented by the growing body of empirical evidence showing that brain damage can result in selective impairments in either speech perception or production (Berndt & Caramazza, 1980; Naeser et al., 1989; Nickels, 2014). These dissociations suggest that the neural processes underlying speech perception and production are, at least partially, distinct.

A different perspective on speech perception suggests that perceptual constancy derives from a process of normalization of the speech signal (Barreda, 2012; Irino & Patterson, 2002; Johnson & Sjerps, 2021; Nearey, 1989; Pisoni, 1992; Strange, 1989; Zahorian & Jagharghi, 1993). Normalization involves a substantial reduction of information and transformation of the signal into a discrete symbolic representation (Pisoni, 1992). This involves removing irrelevant variability or "noise" from the speech signal so that different instances of the same word are made equivalent. According to this view, listeners normalize speech variability by interpreting certain aspects of the speech signal in relation to other aspects. For instance, the first two formant frequencies (F1 and F2) are among the primary cue dimensions used to distinguish vowels (Forgie & Forgie, 1959; Sakayori et al., 2002). However, vowels produced by adult male speakers typically exhibit lower formant frequencies than those produced by female speakers, due to differences in vocal tract length (Chiba & Kajiyama, 1941; but see also Gunnar, 1966). Listeners might account for differences in speech production by normalizing category-relevant speech cues (e.g., F1 and F2) based on acoustic correlates of anatomic differences, such as the fundamental frequency of speech (F0). The *normalization approach* to speech perception has been criticized since speech variability arises not only from biological factors but also from cultural norms and social identity (Johnson, 2006; Labov, 2022; Smyth et al., 2003). Furthermore, several studies showed that listeners encode and retain surface details of speech in episodic memory, which have been typically considered as noise in perceptual systems (e.g., Church & Schacter, 1994; Goldinger, 1996, 1998; Hawkins, 2003; Johnson, 1997; Schacter & Church, 1992).

More recent approaches to speech perception emphasize the brain's ability to dynamically adapt to changes in sensory stimuli (Gutnisky & Dragoi, 2008; Maule & Franklin, 2020; Sharpee et al., 2006; Stocker & Simoncelli, 2006). Under this framework, speech variability among speakers is considered a meaningful and informative cue that listeners can exploit to facilitate perception (Pisoni, 1997). This approach describes phonemes as overlapping distributions across multiple acoustic dimensions, framing speech perception as a

probabilistic process (Kleinschmidt, 2019; Kleinschmidt et al., 2018; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). Moreover, variability in speech production can result in different cue distributions across speakers. According to this view, speech perception involves tracking the variability of the speech signal across different contexts to construct distributional models of acoustic cues. Listeners exploit their distributional (statistical) knowledge of language to map the acoustic cues into phonemes, making speech perception an inference process under uncertainty about the appropriate model for the current speaker. Since cue distributions do not vary arbitrarily across context (e.g., speakers with the same accent will share similar cue distributions), listeners can generalize speaker-specific experiences over groups of speakers. This approach to speech perception is consistent with empirical evidence showing that listeners' expectations about a speaker's social and linguistic background can influence how speech is perceived (Drager, 2011; Hay et al., 2006; Staum Casasanto, 2008; Walker & Hay, 2011). Moreover, this perspective aligns with probabilistic models of language processing (Huettig et al., 2022; Kuperberg & Jaeger, 2016; Levy, 2008; Nour Eddine et al., 2024), which propose that comprehenders integrate prior knowledge and contextual information to maximize the probability of accurately recognizing the linguistic input.

2.2.2. Speaker variability and prediction

Psycholinguistic research has recently begun to explore the extent to which predictions are sensitive to differences between speakers. Brothers et al. (2019) investigated whether comprehenders can flexibly implement predictions during language comprehension by manipulating the reliability of the speaker. They measured ERPs while participants listened to sentences for comprehension, comparing responses to predictable and unpredictable words. The sentences were produced either by a reliable speaker, who tended to complete sentences with highly predictable words, or by an unreliable speaker, who tended to complete the sentences with unpredictable but still plausible words. The results showed that predictable words elicited a smaller N400 compared to unpredictable words. Crucially, the N400 predictability effect was larger when the sentences were produced by the reliable speaker, suggesting that comprehenders rely more on predictive processing if their predictions are likely to be correct. While these findings suggest that semantic predictions are sensitive to inter-individual differences, the extent to which the comprehension system flexibly predicts phonological information is less well understood.

Given the lack of invariance in the speech signal (Liberman et al., 1967), it is unclear whether prediction targets specific phonological forms or remains confined to an abstract

representation that neglects phonological variability of word realizations across, for example, different speakers of a language. Brunellière & Soto-Faraco (2013) leveraged the rule of vowel reduction in Catalan, which differs by regional accent, to investigate whether listeners predict detailed phonological forms of expected words. In Eastern Catalan, vowels such as /a/ and /e/ undergo reduction to schwa (/ə/) when they occur in unstressed syllables. This leads to vowel neutralization (Alarcos, 1953), which makes different words sound more similar. In contrast, Western Catalan does not apply vowel reduction, maintaining distinct vowels in unstressed syllables. Consequently, a word like *third* is produced as /tərsé/ in Eastern Catalan and /tersé/ in Western Catalan, providing a natural experimental manipulation to test whether predictions are tuned to the specific phonology of the speaker's dialect. Brunellière & Soto-Faraco (2013) conducted two experiments: in the first experiment, Eastern Catalan speakers were presented with semantically constraining sentence frames produced in their native regional accent, while in the second experiment, the sentence frame was produced in the non-native regional accent (Western Catalan). In both studies, the sentence final word was either fully expected (the most expected word pronounced with the same accent as the sentence context), phonologically unexpected (the most expected word pronounced in the other regional accent), or fully unexpected (a semantically incongruent word pronounced with the same accent as the sentence context). Results showed that, when the sentence frame was produced in the native accent, phonologically unexpected words elicited a larger negative response (∼250 ms after word onset) compared to fully expected words. Instead, when the sentence frame was produced in the non-native accent, no phonological mismatch effects were observed. The authors suggested that comprehenders implement precise phonological predictions when the sentence context is produced by a native speaker. In contrast, phonological predictions are less specified in the non-native context due to less precise priors for prediction or a lower frequency of occurrence of the phonological variants in the mental lexicon (Connine et al., 2008). While this study provides initial evidence that listeners can generate detailed phonological predictions, it does not clearly distinguish between prediction of non-native phonological forms and the cognitive processes associated with processing an unfamiliar accent in the sentence context. Several studies showed that processing speech from a non-native speaker is associated with an increased cognitive load (Adank et al., 2009; Cristia et al., 2012; Floccia et al., 2006, 2009; Munro & Derwing, 1995), thereby reducing the cognitive resources necessary for predicting semantic (Romero-Rivas et al., 2016) and grammatical information (Schiller et al., 2020). This line of research suggests that the presence of an unfamiliar accent in the sentence context may constrain the system's ability to generate precise phonological predictions. Thus, the extent to which comprehenders

can flexibly implement phonological predictions remains an open question. Addressing this research question may inform models of language comprehension that posit the pre-activation of sublexical information, as well as current theories of speech perception, which assume that listeners draw on their knowledge of how linguistic units (words, syllables, or phonemes) are realized by different speakers.

Chapter 3 – Study 1: I know how you'll say it: evidence of speaker-specific speech prediction

3.1. Introduction

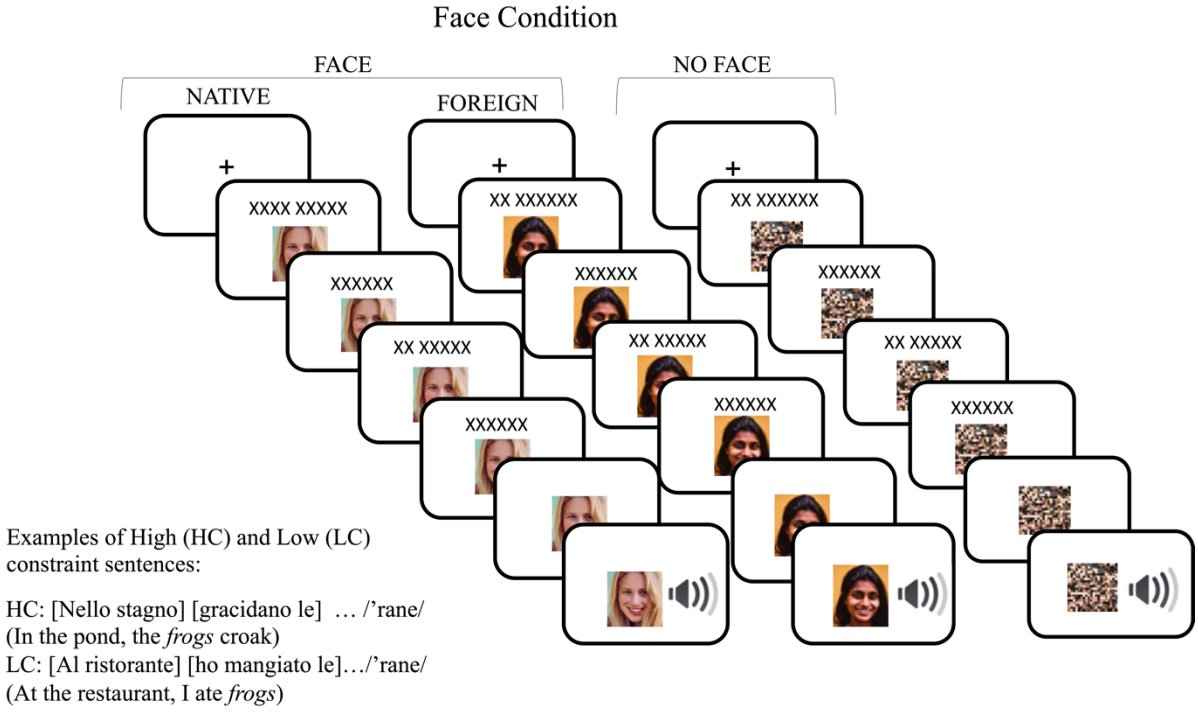
Imagine listening to a sentence that strongly hints at a particular word. You might anticipate its meaning or grammatical role, but can you also predict how it will sound? Current theories of language comprehension assume that highly constraining sentence contexts allow the prediction of the meaning, grammar and phonology of upcoming words (Huettig et al., 2022; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Pickering & Garrod, 2013; Pickering & Strijkers, 2024). A large body of empirical evidence supports the idea that comprehenders can predict semantic (Altmann & Kamide, 1999, 2007; Chambers et al., 2002; Federmeier & Kutas, 1999a; Kamide et al., 2003; Metusalem et al., 2012; Paczynski & Kuperberg, 2012) and syntactic features (Crocker, 2000; Levy, 2008; Lewis, 2000; Staub & Clifton, 2006; Traxler, 2014; Traxler et al., 1998; van Gompel et al., 2005). However, the extent to which it is possible to anticipate the phonological form of a predictable word remains unclear. Although a wide range of experimental paradigms have been employed to investigate this issue, the empirical evidence for phonological prediction remains mixed and controversial, even across studies using the same manipulation. Notably, most of the research on phonological prediction has overlooked speech variability: we rarely pronounce words in exactly the same way, and this variability is even more pronounced across different speakers (Liberman et al., 1967). Speaker variability is particularly evident in non-native speech, which often deviates from native phonetic and phonological norms (Best et al., 2001; Clopper et al., 2005; Flege, 1988). Given the lack of invariance in the speech signal, it is unclear whether prediction targets specific phonological forms or remains confined to more abstract representations that neglect speech variability across speakers.

3.1.1. The present study

This chapter presents a study published in *Psychonomic Bulletin & Review* (Sala et al., 2024), which investigated whether speaker identity (native vs. foreign) can be used to implement specific phonological predictions. To address this question, we employed an experimental paradigm that leverages the fact that non-native speakers frequently produce phonological errors. Participants were first familiarized with a native and a foreign-accented speaker: an image of the speaker's face was presented along with a 1-minute audio clip in which each speaker introduced themselves. In the experimental task, participants read sentence frames that

were missing a last word to be grammatical. The final word was produced by either the native- or foreign-accented speaker. The foreign-accented speaker made consistent phonological errors on the first phoneme of the spoken word. The spoken final word could be semantically predictable or not, depending on the sentence context. Crucially, speaker identity could be either cued or not, as the written sentences were presented in association with the speaker's face or with a neutral visual stimulus (see Figure 3.1). Participants were asked to perform an auditory lexical decision task on the spoken target. In the foreign-accented condition, they were explicitly instructed to accept mispronounced words as correct. We decided to present the sentence context in written form rather than auditorily, as this methodological choice allows us to test for phonological prediction effects while controlling for sentence-context processing demands across accent conditions (Adank et al., 2009; Cristia et al., 2012; Floccia et al., 2006, 2009; Munro & Derwing, 1995). We hypothesized that the image of the speaker's face should facilitate the recognition of the target word by cueing the speaker-specific phonology. Specifically, we expect that cueing the speaker's identity should be associated with faster lexical decision times compared to when the speaker's identity is not cued. Crucially, this facilitation should occur primarily when the target word is predictable from the sentence context, reflecting context-driven phonological predictions. We also aim to explore whether the native and the foreign-accent conditions differ in the generation of phonological predictions, as they require predicting a standard versus a specific deviant phonology, respectively.

Figure 3.1. Schematic overview of the experimental paradigm and procedure. Each trial consisted of a sequence of frames displayed for 800 ms each. The total number of frames varied depending on sentence length. In each frame, a segment of the sentence appeared on the screen alongside a visual cue, which could either be the face of the native or foreign-accented speaker, or a neutral control stimulus. The face cued the accent of the upcoming spoken target (in the example, RANE/frogs). Sentence frames were either Highly constraining (HC) or Low constraining (LC) toward the spoken target. Figure from Sala et al. (2024), licensed under CC BY 4.0.



3.2. Methods

3.2.1. Participants

Fifty-four adults (40 females; M_{age} = 23.80 ± 2.97 y.o.) were initially recruited for the study. Two individuals were excluded from the final sample: one had previously participated in a cloze task related to the present study, and the other was not a native Italian speaker. The final sample included 52 participants (40 females; M_{age} = 23.67 ± 2.96 y.o.), all native Italian speakers with no reported history of neurological, psychiatric, or language-related disorders. Because precise estimates of the effect sizes of our interest were not available, sample size was determined based on methodological recommendations for regression analyses, which suggest including at least five to ten observations per predictor to ensure acceptable coefficient estimates (Hair et al., 2010; Tabachnick & Fidell, 1989). Notably, in linear mixed-effects models, the total number of

observations includes both individual participants and the repeated measures nested within each participant (Bates et al., 2015). The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki. All participants provided their informed consent before participating in the experiment. Permission to conduct the study was given by the Ethics Committee for Psychological Research of the University of Padova.

3.2.2. Materials

All stimuli are publicly accessible in the project's Open Science Framework (OSF) repository (https://osf.io/n42x9/). Target stimuli included 64 words (M_{length} = 6.23 ± 1.97 phonemes) and 64 non-words (M_{length} = 6.45 ± 1.98 phonemes) of comparable length, each beginning with one of three phonemes: /r/, /p/ and /k/. These phonemes did not occur in any other position within the items. To ensure clear discrimination between mispronounced words and non-words in the foreign-accent condition, non-words had no phonological neighbors, making them easily identifiable as non-lexical items. Each spoken word was preceded by a written sentence frame varying in semantic constraint, being either Highly Constraining (HC: example in 1a) or Low Constraining (LC: example 1b) toward the target word.

1a) Nello stagno gracidano le **rane**
In-the pond croak the **frogs**
'In the pond the **frogs** croak'
1b) Al ristorante ho mangiato le **rane**
At-the restaurant have eaten the **frogs**
'At the restaurant I ate **frogs**'

To assess the degree of contextual constraint, we conducted an online cloze task with 22 participants, none of whom took part in the main experiment. Participants were asked to complete sentence frames with the first word that came to mind. Sentence constraint was calculated as the proportion of responses involving the most frequent continuation (*HC*: $M_{constraint}$ = 0.94 ± 0.07; *LC*: $M_{constraint}$ = 0.18 ± 0.08) (Staub, 2015). HC sentence frames were always followed by the most frequent continuation (*HC*: $M_{cloze\ probability}$ = 0.94 ± 0.07), while LC sentence frames were followed by a semantically plausible continuation (*LC*: $M_{cloze\ probability}$ = 0.02 ± 0.05). Sentence frames were similar in length across conditions (*HC*: M_{length} = 9.14 ± 2.12 words, range = 4-14 words; *LC*: M_{length} = 8.86 ± 1.88 words, range = 4-13 words; p = .429). Non-words were also preceded by sentence frames comparable in both length (*HC*:

M_{length} = 9.31 ± 2.05 words, range = 5–15 words; *LC*: M_{length} = 8.83 ± 1.90 words, range = 5–15 words; p = .167) and constraint (*HC*: $M_{constraint}$ = 0.90 ± 0.11; *LC*: $M_{constraint}$ = 0.22 ± 0.08) to those used for real-word targets.

Speech stimuli for both target words and non-words were generated using the Microsoft Azure text-to-speech service, which offers prebuilt neural voices in various languages. We used two distinct voices: an Italian voice for the native-accented speaker ("Fabiola") and an Indian voice for the foreign-accented speaker ("Neerja"). The voices were selected to make participants associate a given face stimulus with either a native or a foreign voice. We created a novel foreign accent to control for participants' prior exposure to specific foreign accents and to introduce systematic phonological variability across speakers. The novel foreign accent was created by manipulating either the place or manner of articulation of three target phonemes (/k/, /r/, and /p/), which were produced by the foreign-accented speaker as /ɢ/, /l/ and /b/, respectively. The speech stimuli for words and non-words were generated starting from IPA transcriptions in Italian. The foreign-accented stimuli were created by systematically manipulating target phonemes in the IPA transcriptions, ensuring both voices received matched phonetic inputs except for the phonological manipulation. For example, the target word "*caldo*" ("heat") was pronounced as /kˈaldo/ by the native speaker, whereas the foreign-accented speaker produced it as /ɢˈaldo/. The foreign-accented speaker consistently mispronounced the first phoneme of each target stimulus; mispronounced words did not correspond to any existing word in the Italian lexicon.

For the familiarization phase, two different discourses, each approximately one minute in length, were prepared and associated with different speakers. The familiarization stimuli were generated using either the native-accented or the foreign-accented voice. The phonological manipulation was applied consistently in both the experimental stimuli and the familiarization speech of the foreign-accented speaker.

3.2.3. Procedure and design

Participants wore headphones and sat comfortably in a quiet room during testing. Stimuli were presented using Psychopy (Peirce et al., 2019). In the familiarization phase, participants were introduced to the native and the foreign-accented speaker through brief audiovisual presentations. For each speaker, we presented a one-minute speech in which the speaker introduced herself, associated with a static image representing the speaker's face. The computer screen displayed either an Indian-looking female face (for the foreign-accented speaker) or an Italian-looking female face (for the native speaker). The familiarization speeches were divided

into two 30-second sections, alternating between speakers.

In the experimental task, participants were asked to silently read the sentence frames presented on-screen and to determine whether the following spoken target was a real word or not. Spoken targets were produced by either the native or foreign-accented speaker introduced during familiarization. The speaker's face (10 × 10 cm) appeared 2500 ms before sentence onset, positioned on the lower part of the screen (4.5 cm below center), and remained visible throughout the whole trial. In half of the trials, the face image was replaced by a control stimulus, created by scrambling together the faces of the two speakers. This approach ensured that the control stimulus was identical across accent conditions and provided no information that could allow participants to predict the speaker's accent. Sentence onset was preceded by a 50-ms fixation cross positioned 4.5 cm above the center of the screen. Sentence frames were presented incrementally, phrase-by-phrase, with each phrase displayed on screen for 800 ms, followed by a 150 ms interval. After the final phrase, the spoken target (word or non-word) was presented following an 800 ms interval. Participants were asked to decide as quickly and accurately as possible whether the spoken target was a real word or not by pressing one of two keys ('C', 'M'), using their index fingers. They were told to use their dominant hand for words. When the speaker had a foreign accent, they were asked to accept mispronounced words as real words. Response times were measured from target onset, and up to 2000 ms after target offset. A forced-choice comprehension question (yes/no) was presented after the lexical decision in 10% of trials (20% of trials in which the target was a word), to ensure that participants paid attention to the semantic content of the sentence. For comprehension questions, participants used the same response keys as during lexical decision, with "Yes" responses made using the dominant hand. A total of 12 practice trials were presented before the start of the experimental task. The experimental task consisted of 256 trials: 128 word trials and 128 non-word trials. Each target was presented twice, once in a High and once in a Low constraint sentence frame. To avoid close repetitions, materials were divided into two blocks, with each target appearing only once per block and being spoken in a different accent across blocks. In each block, 64 spoken targets (32 words, 32 non-words) were produced by the native speaker and 64 by the foreign speaker. The speaker's identity was either cued or not by the speaker's face, resulting in 16 observations per condition. Block order was rotated across participants, and trial order was randomized within blocks. Four experimental lists were created using a Latin square design to ensure that each stimulus occurred in all conditions across participants. Participants were evenly assigned to the lists. The full experimental session lasted about 45 minutes.

3.2.4. Statistical analyses

Data were analyzed using the open-source R software (R Core Team, 2023). The complete dataset and analysis script are publicly available in the OSF repository of the project (https://osf.io/n42x9/). Before conducting the statistical analyses, we verified that all participants presented an accuracy level above 80% in both lexical decision and comprehension questions. No participants were excluded from the analysis. Only trials ending with a word were included in the analysis (non-word results are reported in the online Supplementary materials of the original paper). Responses faster than 150 ms and slower than 2500 ms were excluded from the analysis (0.007% of total observations). Response accuracy was modelled using generalized linear mixed-effects models with a binomial distribution and logit link. Reaction times (RTs) for correct responses were log-transformed and analyzed via linear mixed-effects models assuming a Gaussian distribution. Mixed-effects models were estimated using the lme4 package (Bates et al., 2015). To identify the best-fitting model for our data, we used a hierarchical model comparison approach (Heinze et al., 2018). Model selection was based on AIC values (Akaike Information Criterion; Akaike, 1974), with delta AIC and AIC weights as measures of comparative model fit. The AIC and AIC weight provide information on the relative evidence of models (i.e., likelihood and parsimony), and the model with the lowest AIC and the highest AIC weight is to be preferred (Wagenmakers & Farrell, 2004). For the analyses of both accuracy and response times, model comparison included a baseline model with random intercepts for Participants and Items, accounting for individual differences and item-specific variability (Baayen et al., 2008). Due to convergence failures, random slopes were excluded from the analyses. The order of the predictors in model comparison prioritized main effects over interactions, focusing first on well-established factors in the literature, specifically Accent and Constraint (Faust & Kravetz, 1998; Federmeier, 2007; Floccia et al., 2006, 2009; Munro & Derwing, 1995). Predictors were included sequentially in the following order: (i) Accent (Native vs. Foreign); (ii) Constraint (HC vs. LC); (iii) Face (Face vs. No Face); (iv) Constraint*Accent; (v) Constraint*Face; (vi) Constraint*Accent*Face. Observations identified as outliers using the "outlierTest" function of the *car* package were excluded (0.0006 % of model observations) (Fox & Weisberg, 2019). Main effects were estimated using sum coding as contrast coding (Brehm & Alday, 2022). Post-hoc comparisons were conducted using the "contrast" function of the *emmeans* package (Lenth et al., 2023), applying Bonferroni correction to p-values (Bonferroni, 1936).

3.3. Results

3.3.1. Response accuracy

Table 3.1 presents the descriptive statistics for response accuracy in each experimental condition.

Table 3.1. Mean accuracy and standard deviation for each experimental condition.

Accent	*High Constraint*		*Low Constraint*	
	No Face	*Face*	*No Face*	*Face*
Native	1 ± 0.01	1 ± 0.01	0.98 ± 0.04	0.98 ± 0.03
Foreign	0.95 ± 0.07	0.97 ± 0.04	0.79 ± 0.11	0.82 ± 0.10

Table 3.2 shows the model comparison for predicting response accuracy. The best-fitting model (lower delta AIC and higher AIC weight) for accuracy is Model 3:

Accuracy ~ Accent + Constraint + Face + (1Participant) + (1Item)

Table 3.2. Model comparison for predicting response accuracy. Deviance = residual deviance; dAIC = difference between AIC of each model and the model with lower AIC; AICw = AIC weight.

Models	**Deviance**	**dAIC**	**AICw**
M0. Accuracy ~ (1|Participant) + (1|Item)	2718.208	465.31	0.0
M1. Accuracy ~ Accent + (1|Participant) + (1|Item)	2291.899	41.0	0.0
M2. Accuracy ~ Accent + Constraint + (1|Participant) + (1|Item)	2251.291	2.40	0.15
M3. Accuracy ~ Accent + Constraint + Face + (1|Participant) + (1|Item)	**2246.896**	**0.0**	**0.49**
M4. Accuracy ~ Accent + Constraint + Face + Constraint*Accent + (1|Participant) + (1|Item)	2246.482	1.59	0.22
M5. Accuracy ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + (1|Participant) + (1|Item)	2245.961	3.07	0.11
M6. Accuracy ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + Constraint*Accent*Face + (1|Participant) + (1|Item)	2244.698	5.80	0.03

The results of the best-fitting model for accuracy are reported in Table 3.3. The Accent effect indicates reduced accuracy in the foreign compared to the native-accent condition. The Constraint effect indicates greater accuracy for predictable compared to unpredictable words. Finally, the Face effect indicates that cueing the speaker's identity is associated with greater accuracy compared to when the speaker's identity is not cued.

Table 3.3. Results of the best-fitting model for response accuracy.

	Estimate	***CI (95%)***	***Std. Error***	***z-value***	***p-value***
Intercept	4.471	[4.061 4.881]	0.209	21.356	< .001
Accent [Foreign]	-1.466	[-1.653 -1.279]	0.095	-15.363	< .001
Constraint [HC]	1.012	[0.722 1.302]	0.148	6.841	< .001
Face [Face]	0.124	[0.008 0.240]	0.059	2.092	.036

3.3.2. Response times

Table 3.4 presents the descriptive statistics for response times (ms) in each experimental condition.

Table 3.4. Mean response times and standard deviation for each experimental condition.

Accent	***High Constraint***		***Low Constraint***	
	No Face	***Face***	***No Face***	***Face***
Native	808.37 ± 220.42	767.20 ± 238.47	963.81 ± 213.86	958.62 ± 221.74
Foreign	980.90 ± 238.40	933.94 ± 245.21	1266.67 ± 291.45	1258.43 ± 297.76

Table 3.5 shows the model comparison for predicting response times. The best fitting model (lower delta AIC and higher AIC weight) for response times is Model 5:

*LogRTs ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + (1|Participant) + (1|Item)*

Table 3.5. Model comparison for predicting response times. Deviance = residual deviance; dAIC = difference between AIC of each model and the model with lower AIC; AICw = AIC weight.

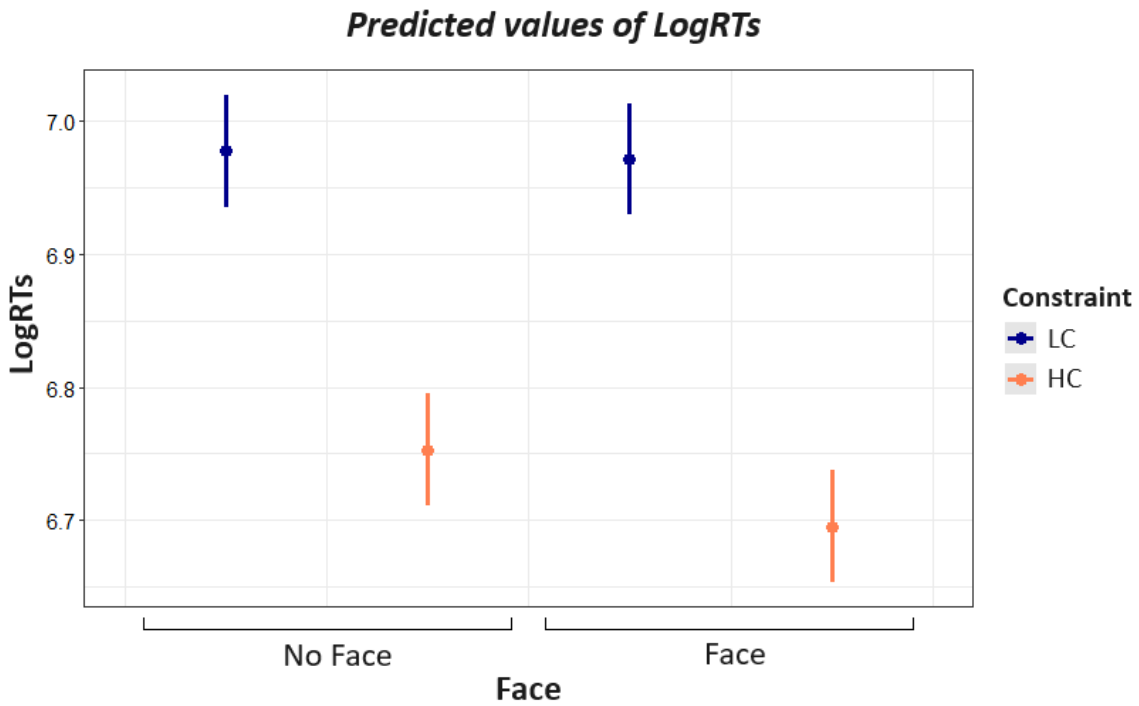
Models	Deviance	dAIC	AICw
M0. LogRTs ~ (1|Participant) + (1|Item)	-519.555	2322.63	0.0
M1. LogRTs ~ Accent + (1|Participant) + (1|Item)	-2563.356	280.83	0.0
M2. LogRTs ~ Accent + Constraint + (1|Participant) + (1|Item)	-2717.974	128.21	0.0
M3. LogRTs ~ Accent + Constraint + Face + (1|Participant) + (1|Item)	2767.483	80.70	0.0
M4. LogRTs ~ Accent + Constraint + Face + Constraint*Accent + (1|Participant) + (1|Item)	-2822.234	27.95	0.0
M5. LogRTs ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + (1|Participant) + (1|Item)	**-2852.183**	**0.0**	**0.85**
M6. LogRTs ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + Constraint*Accent*Face + (1|Participant) + (1|Item)	-2852.650	3.53	0.15

The results of the best-fitting model for response times are reported in Table 3.6. As all main effects are implied in interactions, we focus on the latter. The interaction between Constraint and Accent indicates that the Constraint effect, namely faster response times for predictable words compared to unpredictable words, is larger for the foreign compared to the native-accent condition. Crucially, the interaction between Constraint and Face indicates that cueing the speaker's identity is associated with a larger Constraint effect compared to when the speaker's identity is not cued (Figure 3.2).

Table 3.6. Results of the best-fitting model for LogRTs.

	Estimate	*CI (95%)*	*Std. Error*	*t-value*	*p-value*
Intercept	6.856	[6.817 6.896]	0.020	343.904	< .001
Accent [Foreign]	0.120	[0.115 0.125]	0.002	51.004	< .001
Constraint [HC]	-0.126	[-0.140 -0.112]	0.007	-17.533	< .001
Face [Face]	-0.016	[-0.021 -0.011]	0.002	-6.864	< .001
Constraint[HC] Accent[Foreign]*	-0.017	[-0.022 -0.013]	0.002	-7.427	< .001
Constraint[HC] Face[Face]*	-0.013	[-0.017 -0.008]	0.002	-5.522	< .001

Figure 3.2. Model estimates for the interaction between Constraint and Face. The error bars indicate 95% confidence intervals.



Post-hoc comparisons show that cueing the speaker's identity is associated with faster response times for predictable words (b = -0.029, z-ratio = -8.996, $p < .001$) but not for unpredictable words (b = -0.003, z-ratio = -0.926, $p = .709$). The Constraint*Accent*Face interaction did not improve model fit; thus, we found no evidence that the face cueing effect for predictable words is modulated by the speaker's accent.

3.4. Discussion

The present study aimed to investigate whether comprehenders can anticipate the phonological form of a highly predictable word. Specifically, we sought to determine whether speaker identity (native vs. foreign) can be used to implement speaker-specific phonological predictions. To address this question, we employed an experimental paradigm that leverages on the fact that non-native speakers frequently produce phonological errors. Participants read sentence frames in which the last word was produced either by a native- or a foreign-accented speaker. The sentence-final word could be predictable or not based on sentence context. Crucially, during trial presentation, speaker identity could be cued or not by an image of the speaker's face, thus manipulating the availability of information about the phonological features of the spoken target before its presentation. Participants were asked to categorize the spoken target as a word or non-word. In the foreign-accented condition, they were explicitly instructed to accept mispronounced words as correct. Results showed that cueing speaker identity was associated with faster lexical decision times for predictable but not for unpredictable words. We found no evidence suggesting that the face cueing effect is modulated by the speaker's accent (native or foreign).

3.4.1. Face cueing effect and phonological prediction

The results of Study 1 provide compelling evidence for the involvement of phonological representations in prediction during language comprehension, at least in highly informative and constraining contexts. The speaker's face provided a cue to the speaker-specific phonology, suggesting that comprehenders used the available information to anticipate the phonological form of the upcoming word. Since cueing the speaker identity led to faster response times only for predictable words, this effect is unlikely to stem solely from the activation of speaker-specific representations (Creel et al., 2008; Creel & Bregman, 2011; Goldinger, 1996; Nygaard & Pisoni, 1998; Palmeri et al., 1993; Remez et al., 1997). Rather, it seems to also reflect the contribution of predictive mechanisms driven by sentence-level constraints. Previous research showed that predictions during language comprehension are sensitive to inter-individual differences. For instance, Brothers et al. (2019) showed that the extent to which comprehenders predict semantic information is influenced by the speaker's reliability. Our findings extend the flexibility of the prediction system to the phonological domain, showing that comprehenders can flexibly anticipate specific phonological forms as well.

Brunellière & Soto-Faraco (2013) provided evidence suggesting that listeners can implement speaker-specific phonological predictions for native speakers but not for non-native

speakers. The authors attributed the results to less reliable priors for prediction or a lower frequency of occurrence of phonological variants in the mental lexicon during comprehension of unfamiliar accented speech (Connine et al., 2008). In our study, we observed that cueing the speaker's identity facilitates the recognition of predictable words, and this effect does not seem to be influenced by the speaker's accent. Crucially, our experimental paradigm involved the presentation of sentence contexts in the written form, rather than auditorily as in Brunellière & Soto-Faraco (2013). Several studies showed that processing the speech from a non-native speaker is associated with an increased cognitive load (Adank et al., 2009; Cristia et al., 2012; Floccia et al., 2006, 2009; Munro & Derwing, 1995; Porretta et al., 2020; Schiller et al., 2020), and the availability of cognitive resources impacts language predictive processing (Ding et al., 2023; Huettig & Janse, 2016). By presenting the sentence context in the written form, we have been able to isolate the effects of phonological prediction from the additional cognitive load associated with processing the sentence context with an unfamiliar (foreign) phonology. Therefore, our findings are not incompatible with the results reported by Brunellière & Soto-Faraco (2013). Rather, they indicate that the system can generate very precise phonological predictions when it is not faced with the additional challenge of processing non-native speech in the sentence context.

3.4.2. Implications for models of language prediction

A widely accepted assumption in psycholinguistics is that the activation of a lexical item spreads to semantically similar or related items (Anderson, 1983; Collins & Loftus, 1975; Huettig et al., 2022; Hutchison, 2003; McRae et al., 1997). When a sentence context is sufficiently constraining, activation may accumulate on a specific lexical item, leading to its pre-activation even at the level of phonological (Huettig et al., 2022) or orthographic (Kim & Lai, 2012; Molinaro et al., 2013) features. Huettig et al. (2022) suggested that pre-activation can be constrained by both linguistic and non-linguistic information. For instance, if we see a man standing on the edge of a rooftop and simultaneously hear someone say "Don't...", our cognitive system is likely to pre-activate the word "jump". This prediction is based not only on the unfolding linguistic input, but also on the visual context and our real-world knowledge. In our experimental paradigm, the speaker's face provided a cue to the phonological features of the upcoming word before its presentation. Crucially, spreading of activation accounts assume that the pre-activation of low-level features is a downstream consequence of the (pre)activation of lexical-semantic representations. Although these models acknowledge that non-linguistic (semantic) information can influence the pre-activation of phonological representations, they

fail to explain how comprehenders flexibly implement predictions based on a phonologically informative non-linguistic cue. Moreover, passive spreading of activation accounts posit that pre-activation depends on the relational links stored in lexical-semantic memory. In our study, the foreign-accented speaker produced words containing phonological errors, making it unlikely that comprehenders had well-established lexical representations of such phonological forms (namely, lexemes). Nevertheless, phonological predictions do not seem to be influenced by the speaker's accent (native or foreign).

Prediction-by-production represents a valuable framework for understanding how comprehenders implement speaker-specific phonological predictions. According to these models, prediction during language comprehension relies on processes and representations that are shared with language production (Huettig, 2015; Pickering & Gambi, 2018; Pickering & Garrod, 2013; Pickering & Strijkers, 2024), allowing comprehenders to generate predictions at different levels of representation, including the phonological level. Neurophysiology studies reveal similar time-frequency modulations between word planning in language production and prediction during comprehension (Gastaldon et al., 2020, 2023), suggesting that the language production network contributes to some extent to prediction in highly constraining sentences. Although the experimental paradigm implemented in the present study was not specifically designed to test a prediction-by-production account, our findings are compatible with this framework. Prediction-by-production accounts often emphasize the role of event simulation in the generation of predictions. For instance, Pickering & Garrod (2013) proposed that listeners covertly imitate the unfolding utterance to derive an internal representation of the speaker's percept, which constrains the prediction of upcoming speech. Similarly, Pickering & Gambi (2018) proposed that covert imitation enables the transformation of comprehension representations into production representations and allows comprehenders to take into account possible differences between themselves and the speaker. This is consistent with evidence showing that overt imitation of a foreign accent is associated with improved comprehension of foreign-accented sentences from background noise (Adank et al., 2010). In our study, participants may have relied on production mechanisms to internally simulate speech when the speaker's identity was cued, enabling the generation of predictions constrained by both the sentence context and the speaker's accent.

Pickering & Garrod (2013) proposed that comprehenders should prioritize prediction based on production mechanisms when they can identify with the speaker, since their predictions are more likely to be correct. Accordingly, in our study, participants were expected to engage more in phonological prediction with the native speaker compared to the foreign-

accented speaker. However, our findings did not confirm this expectation. One possible explanation lies in the systematic nature of the phonological manipulation: the foreign accent involved altering three phonemes, and each target word began with one of these phonemes. This structured pattern may have allowed participants to accurately anticipate the phonological form of predictable words. Another possibility is that participants adopted a more general processing strategy, anticipating that pronunciation errors might occur on the initial phoneme of the spoken target without predicting their exact realization. Under this account, when the foreign speaker identity was cued, an initial phoneme mismatching with the predicted word would not necessarily elicit a prediction error. Our data do not allow us to conclusively distinguish between these alternatives, and future work should clarify whether listeners make detailed phonological predictions for foreign-accented speakers or rely on broader expectations of reduced reliability. Additionally, although individuals tend to identify more with an in-group member, our paradigm was not designed to explicitly manipulate the likelihood of identification with the speaker; therefore, it may not adequately capture the conditions necessary for testing this particular aspect of the proposal by Pickering & Garrod (2013).

Other theoretical accounts argue that comprehenders can actively predict upcoming speech without assuming a role for language production mechanisms. Kuperberg & Jaeger (2016) proposed that language comprehension relies on a hierarchical generative architecture whose goal is to infer the message intended by the speaker. According to this view, predictions are generated through internal generative models, which are defined as a set of hierarchically organized internal representations. Internal generative models enable the probabilistic pre-activation of information at multiple levels of representation, thereby increasing the likelihood of accurately recognizing the incoming speech input. Internal representations can include information about the speaker's sound structure (Connine et al., 1991; Szostak & Pitt, 2013), and listeners may acquire distinct generative models through exposure to different statistical environments. Therefore, listeners may generate speaker-specific phonological predictions by relying on distinct internal generative models. The proposal by Kuperberg & Jaeger (2016) aligns closely with current models of speech perception, which suggest that listeners track the variability of the speech signal across different contexts to construct distributional (statistical) models of acoustic cues (Kleinschmidt, 2019; Kleinschmidt et al., 2018; Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). From this perspective, both speech perception and linguistic prediction can be understood as an inference process under uncertainty, where listeners must choose the appropriate generative model for the current speaker. Within this framework, our findings may indicate that perceptual adaptation to speech extends beyond the

mapping of the speech signal into phonemes, influencing predictions based on sentential constraints as well.

3.5. Limitations and future directions

The results of Study 1 provide evidence suggesting that prediction during language comprehension can target specific phonological forms. However, response times provide only a global measure of processing and do not describe *when* specific cognitive processes occur. To more precisely characterize the cognitive and neural mechanisms underlying predictive processing at sub-lexical levels, following research should integrate behavioral data with neurophysiological techniques (e.g., EEG, MEG), neurostimulation methods, or patient studies. Moreover, in our study, speech stimuli were produced either by one native or one foreign-accented speaker. Recent models of speech perception propose that listeners can generalize speaker-specific experiences over groups of speakers who share a similar social or linguistic background (Kleinschmidt & Jaeger, 2015). Future work should investigate the extent to which listeners can generalize speaker-specific experiences and whether this also applies to predictive processing in language comprehension. Finally, employing paradigms that explicitly manipulate or control cognitive load would help clarify the boundary conditions of speaker-specific effects in phonological prediction and reveal the extent to which comprehenders can flexibly implement such predictions.

Chapter 4 – Study 2: In the words of others: ERP evidence of speaker-specific phonological prediction

4.1. Introduction

In Study 1, we observed that cueing the speaker's identity is associated with faster lexical decision times for predictable words but not for unpredictable words. These results suggest that participants use the face cue to implement speaker-specific phonological predictions. However, response times offer only a global measure of processing: they tell us *when* a response was made, but not *what* happens *before* the response or *when* specific cognitive processes occur. To gain a more detailed understanding of the temporal dynamics of cognitive processes, it is essential to turn to electrophysiological methods such as Event-Related Potentials (ERPs). ERPs provide a window into the timing of the cognitive processes elicited by experimental events. They are obtained by segmenting electroencephalography (EEG) data into epochs that are time-locked to a specific event, such as the onset of a stimulus. The epochs are usually averaged separately by experimental condition in order to preserve stimulus-related brain activity, which is constant across trials, while filtering out unrelated brain activity or noise (Luck, 2014). Studies employing ERPs revealed distinct components associated with different aspects of language processing (Swaab et al., 2012). For instance, the N400 component typically shows a larger amplitude for unpredictable words compared to predictable words. This has led to the proposal that the N400 reflects the brain's response to prediction errors and the revision of prior expectations during language comprehension (Kuperberg, 2016; Nour Eddine et al., 2024). Several studies have investigated whether comprehenders predict the phonological form of a highly predictable word by examining the N400 component (DeLong et al., 2005, 2019, 2021; Ito et al., 2017, 2020; Laszlo & Federmeier, 2009; Nieuwland et al., 2018). However, the results across these studies remain largely inconsistent, even when the same experimental manipulation was used.

4.1.1. The present study

The current chapter presents a study published in *Psychophysiology* (Sala et al., 2025). In this study, we adapted the experimental paradigm of Study 1 to the ERP context in order to better characterize the processes involved in the prediction of phonological information. Participants were first familiarized with a native- and a foreign-accented speaker. In the experimental task, they silently read sentence frames that were either highly or weakly constraining toward a spoken target word, which was pronounced by either the native or the foreign-accented speaker.

The foreign-accented speaker made consistent phonological errors on the first phoneme of the spoken word. Crucially, during trial presentation, the speaker's identity was either cued or not by an image of the speaker's face. Unlike Study 1, the experimental paradigm did not include non-word trials or a lexical decision task. Instead, in 25% of the trials, participants were asked to indicate whether they expected the word produced by the speaker, regardless of how it was pronounced. This modification was motivated by the need to collect a larger number of trials per condition to reliably estimate ERP components (Boudewyn et al., 2018; Jensen & MacDonald, 2023; Swaab et al., 2012). We hypothesized that the face cue would enable participants to implement speaker-specific phonological predictions. Accordingly, we expected the spoken target to elicit a larger N400 predictability effect when the speaker identity is cued compared to when it is not cued, reflecting easier processing of predictable words due to phonological prediction. Similarly to Study 1, we also aim to explore whether phonological predictions differ between the native and the foreign-accented condition.

4.2. Methods

4.2.1. Participants

Forty-eight adults (39 females, M_{age} = 23.27 ± 3.05 y.o.) were recruited. Participants were right-handed native Italian speakers with no history of neurological, language-related, or psychiatric disorders. They were recruited from healthy volunteers and students at the University of Padova. Participants received 15 euros for their participation. Our initial sample size of 48 participants aimed to ensure sufficient power to detect at least the effect of face cue on predictable words after possible data loss. The power analysis, conducted using the method proposed by Judd et al. (2017), indicated that at least 38.9 participants would be required to detect a small-to-medium effect (Cohen's d = 0.3) with 80% power. The study was conducted in accordance with the ethical guidelines of the Declaration of Helsinki. All participants provided their informed consent before participating in the experiment. Permission to conduct the study was given by the Ethics Committee for Psychological Research of the University of Padova.

4.2.2. Materials

All stimuli, data and analysis scripts are publicly available in the project's Open Science Framework (OSF) repository (https://osf.io/brvkx/). The target stimuli included 168 spoken words (M_{length} = 5.86 ± 1.86 phonemes), all of which began with one of the following phonemes: /k/, /p/, or /r/. These phonemes did not appear in any other position within the same word. Each

target was preceded by a written sentence frame varying in semantic constraint, being either Highly Constraining or Low Constraining toward the target word. As in Study 1, we measured sentence constraint using an online sentence completion questionnaire completed by 22 participants, none of whom took part in the main experiment. Participants were asked to complete sentence frames with the first word that came to mind. Sentence constraint was defined as the proportion of responses involving the most frequent continuation (*HC*: $M_{\text{constraint}}$ = 0.94 ± 0.07; *LC*: $M_{\text{constraint}}$ = 0.17 ± 0.09). HC sentence frames were always followed by the most frequent continuation (*HC*: $M_{\text{cloze probability}}$ = 0.94 ± 0.07), while LC sentence frames were followed by a semantically plausible continuation (*LC*: $M_{\text{cloze probability}}$ = 0.03 ± 0.09). Sentence frames were similar in length across conditions (*HC*: M_{length} = 9.45 ± 2.20 words, range = 4-15 words; *LC*: M_{length} = 9.28 ± 2.03 words, range = 4-15 words; p = .455). Target stimuli were produced by either an artificial voice with a native Italian accent or an artificial voice with a novel foreign accent. Speech stimuli were created using the same procedure (and speakers) as in Study 1, and their duration was similar across accent conditions (*Native accent*: M_{duration} = 688 ±129 ms; *Foreign accent*: M_{duration} = 700 ± 125 ms; p = .399). The familiarization phase employed the same speech materials as in Study 1.

4.2.3. Procedure and design

Participants were seated comfortably in a soundproof room, equipped with a computer setup that included an LCD monitor, external speakers, and a keyboard. Stimuli were presented using PsychoPy (Peirce et al., 2019). The familiarization phase was the same as in Study 1: we presented a one-minute speech in which each speaker introduced herself, associated with a static image representing the speaker's face. The computer screen displayed either an Italian-looking female face (for the native speaker) or an Indian-looking female face (for the foreign-accented speaker).

In the experimental task, participants were asked to silently read the sentence frames displayed on the screen. The sentence frames were followed by a spoken word, produced either by the native or the foreign-accented speaker introduced during the familiarization phase. The foreign-accented speaker made consistent phonological errors on the initial phoneme of the spoken target. The spoken word could be either predictable or not, based on sentential constraint. An image of the speaker's face appeared 2000 ms before sentence onset, positioned on the lower part of the screen (4.5 cm below center), and remained continuously visible throughout the whole trial. In half of the trials, the face image was replaced by a control stimulus, created by scrambling together the faces of the two speakers. Both the face and the

control stimulus were 10 cm wide and 10 cm high. A fixation cross appeared 450 ms before the onset of the sentence and remained on screen for 300 ms. Sentence frames were presented word-by-word at a regular pace (300 ms word duration, 200 ms inter-word interval). Both the fixation cross and the sentence frame were presented 4.5 cm above the center of the screen. The last word of the sentence frame was followed by a fixation cross that remained on-screen for the rest of the trial. The experimental stimuli remained on screen for 2800 ms after the last word of the sentence frame. We planned to present the spoken target 800 ms after the presentation of the last word of the sentence frame; however, due to a technical error, this interval resulted in 930 ms. During trial presentation, participants were asked to avoid blinking and minimize movements. Each trial was followed by a 1750 ms interval in which participants could blink. In 25% of the trials, participants answered a written question asking whether they expected the word spoken by the speaker, regardless of how it was pronounced. They responded by pressing either the 'C' or 'M' key with their left or right index finger, respectively. In half of the question trials, the dominant hand was used to indicate an "expected" word; in the other half, it indicated an "unexpected" word.

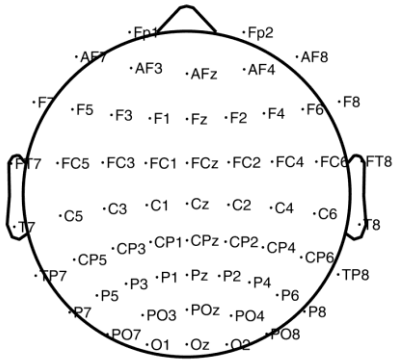
Prior to the experimental task, participants completed 8 practice trials that were not included in the experimental materials. The whole session lasted approximately 2 hours. Each participant completed a total of 336 trials, and they could take a short break every 21 trials. Each target word was presented twice: once following a High Constraint sentence frame and once following a Low Constraint sentence frame. To avoid close repetitions, the materials were divided into two blocks, with each target word spoken in a different accent across blocks and appearing only once per block. Each block included 84 spoken targets produced by the native-accented speaker and 84 by the foreign-accented speaker. In half of the trials, speaker identity was cued by an image of the speaker's face; in the other half, a visual neutral stimulus was presented, resulting in 42 trials per condition. The order of presentation of the two experimental blocks was counterbalanced across participants, and trial order within each block was randomized. Four experimental lists were created using a Latin square design to ensure that each stimulus occurred in all conditions across participants. Twelve participants were assigned to each list.

4.2.4. EEG data acquisition and pre-processing

EEG was recorded with a 64-channel active Ag/AgCl electrode system (Brain Products, ActiCap), following the international 10–20 convention. Sixty electrodes were used as active recording sites (Fp1, Fp2, AF3, AF4, AF7, AF8, Afz, F1, F2, F3, F4, F5, F6, F7, F8, Fz, FT7,

FT8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, T7, T8, C1, C2, C3, C4, C5, C6, Cz, TP7, TP8, CP1, CP2, CP3, CP4, CP5, CP6, CPz, P1, P2, P3, P4, P5, P6, P7, P8, Pz, PO3, PO4, PO7, PO8, Poz, O1, O2, Oz). Three electrodes were used to record blinks and saccades (external ocular canthi and below the left eye), and one electrode was positioned on the right earlobe. The reference was positioned on the left earlobe. Figure 4.1 illustrates the scalp position of the electrodes.

Figure 4.1. 2D electrode layout used in the current experiment. Three additional electrodes were used to record ocular activity and one was placed on the right earlobe.



The configuration was considered adequate only if electrode impedance was below 20 kΩ at the end of electrode placement. The EEG signal was amplified and digitized at a sampling rate of 500 Hz. Prior to the experimental task, a resting-state of 3 minutes was recorded, which is not analyzed further here. Before analyzing the EEG signal, we examined participants' behavioral data to ensure that they were reading the sentence frames and listening to spoken words. One participant was excluded from the analyses due to a high number of targets (69%) classified as "expected" in Low Constraint contexts. EEG preprocessing was performed in MATLAB (The MathWorks Inc., 2023) using EEGLAB (Delorme & Makeig, 2004) and Fieldtrip (Oostenveld et al., 2011). EEG signals were offline re-referenced to the average of the left and right earlobes. A 0.5–80 Hz band-pass filter was applied to attenuate slow drifts and high-frequency noise. Noisy or flat channels (up to three per participant) and segments with extreme muscle artifacts were manually removed. ICA (Independent Component Analysis) was applied after PCA-based dimensionality reduction to 60 components to identify and remove artifacts with well-known time-course and scalp distribution (blink, saccades and power-line noise at 50 Hz). To optimize the detection of ocular artifacts and power line-noise, ICA was applied to band-pass filtered data (1-55 Hz). ICA weights were then projected to the EEG data

filtered between 0.5 and 80 Hz. Following ICA correction, missing channels were interpolated using superfast spherical interpolation. Finally, the data were low-pass filtered at 30 Hz and segmented into epochs of 1200 ms, starting 200 ms before the onset of the spoken targets. A 200 ms pre-stimulus baseline correction was applied to all the extracted epochs. Trials contaminated by slow drifts, muscular activity or remaining eye-artifacts were excluded based on visual inspection (see Table 4.1). Two participants were removed from the analysis due to excessive trial rejection (> 20%), and three more participants were excluded due to pronounced alpha activity. Therefore, the final dataset included 42 participants (34 females; M_{age} = 23.05 ± 2.97 y.o.).

Table 4.1. Number of accepted trials for each experimental condition (M ± SD).

	High Constraint		**Low Constraint**	
	Face	**No Face**	**Face**	**No Face**
Native	41.02 ± 1.29	41.02 ± 1.26	41.19 ± 1.09	40.98 ± 1.32
Foreign	41.07 ± 1.47	40.83 ± 1.36	41.12 ± 1.40	41.02 ± 1.57

4.2.5. Statistical analyses

Statistical analyses were conducted using the open-source R software (R Core Team, 2023). Linear mixed models were used to analyze the single-trial mean voltage within the canonical N400 time-window (300-500 ms after word onset) in a cluster of centro-parietal electrodes (Cz, C3, C4, Pz, P3, P4) (Kutas & Federmeier, 2011; Nieuwland et al., 2018; Šoškić et al., 2022). To identify the best-fitting model for our data, we used a hierarchical model comparison approach, systematically comparing increasingly complex models (Heinze et al., 2018). Model comparison was based on AIC values (Akaike Information Criterion; Akaike, 1974), and especially delta AIC and AIC weights as measures of comparative model fit. The AIC and AIC weight give information on the relative evidence of models (i.e., likelihood and parsimony), and the model with the lowest AIC and the highest AIC weight is to be preferred (Wagenmakers & Farrell, 2004). To control for variability across participants and items, the baseline model included Constraint by Participant random slopes, and Participant and Item as random intercepts (Baayen et al., 2008). The order of the predictors in model comparison prioritized main effects over interactions, focusing first on well-established factors in the literature, specifically Accent and Constraint (Brunellière & Soto-Faraco, 2013; Goslin et al., 2012; Grey & van Hell, 2017; Kutas & Federmeier, 2011). Predictors were entered sequentially in the

following order: (i) Accent (Native vs. Foreign); (ii) Constraint (HC vs. LC); (iii) Face (Face vs. No Face); (iv) Accent*Constraint; (v) Constraint*Face; (vi) Accent*Face; (vii) Accent*Constraint*Face. Sum coding was used for estimating main effects (Brehm & Alday, 2022). Post-hoc comparisons were conducted using the "contrast" function of the *emmeans* package (Lenth, 2025), applying Bonferroni correction to p-values (Bonferroni, 1936).

4.3. Results

4.3.1. Behavioral results

On 25% of trials, participants were asked to evaluate whether they expected or not the spoken target, regardless of how it was pronounced. Table 4.2 presents the proportion of target words classified as "expected" in control trials for each experimental condition.

Table 4.2. Proportion of target words classified as "expected" in each experimental condition (mean ± SD).

Accent	High Constraint		Low Constraint	
	Face	No Face	Face	No Face
Native	0.94 ± 0.23	0.93 ± 0.25	0.07 ± 0.26	0.08 ± 0.27
Foreign	0.95 ± 0.23	0.94 ± 0.24	0.05 ± 0.23	0.07 ± 0.25

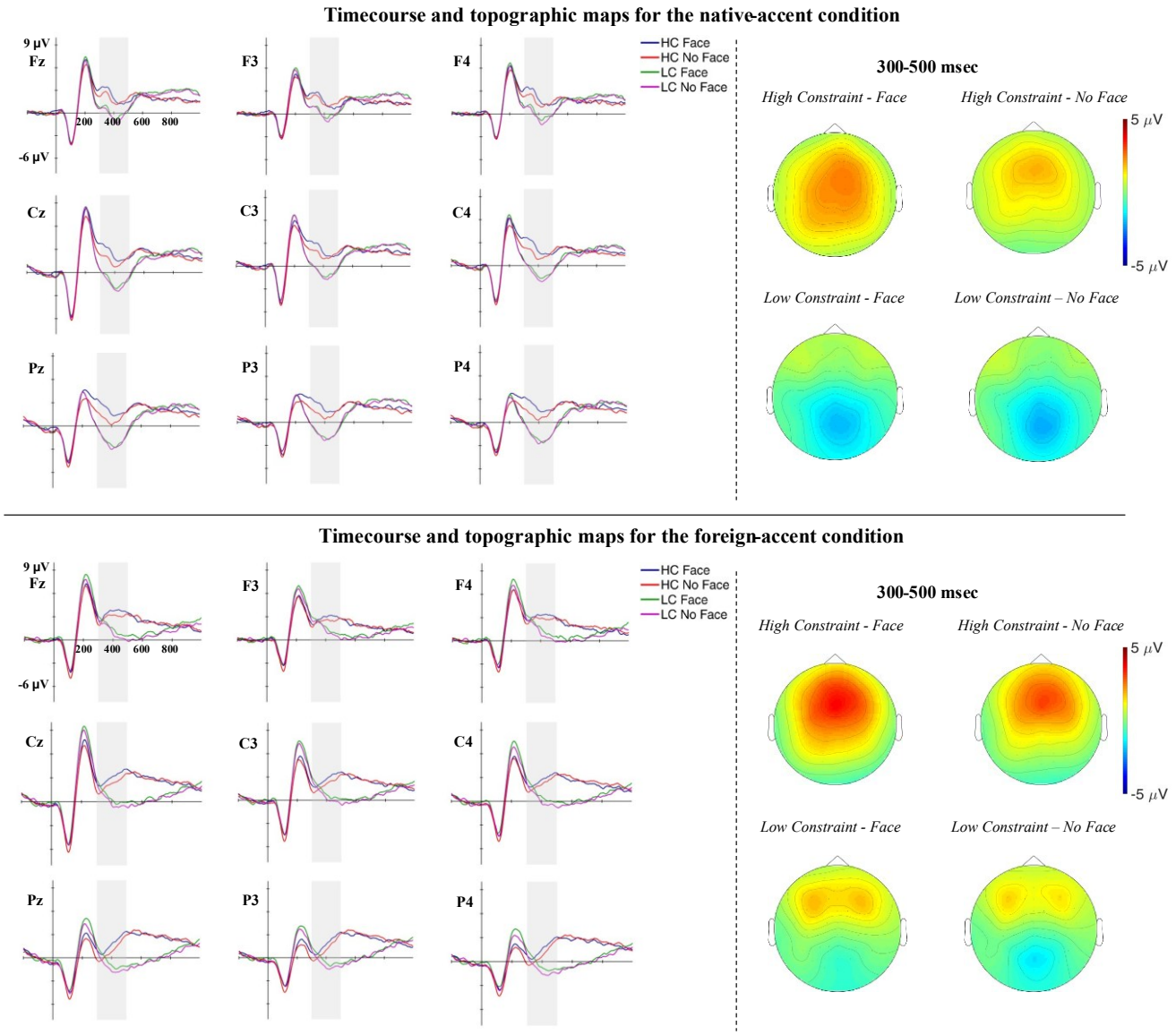
Participants' responses in the control trials show that most of the words in High Constraint contexts were classified as "expected", while most of the words in Low Constraint contexts were not, suggesting that they were attentive to the experimental materials.

4.3.2. ERP results

Visual inspection of the waveforms

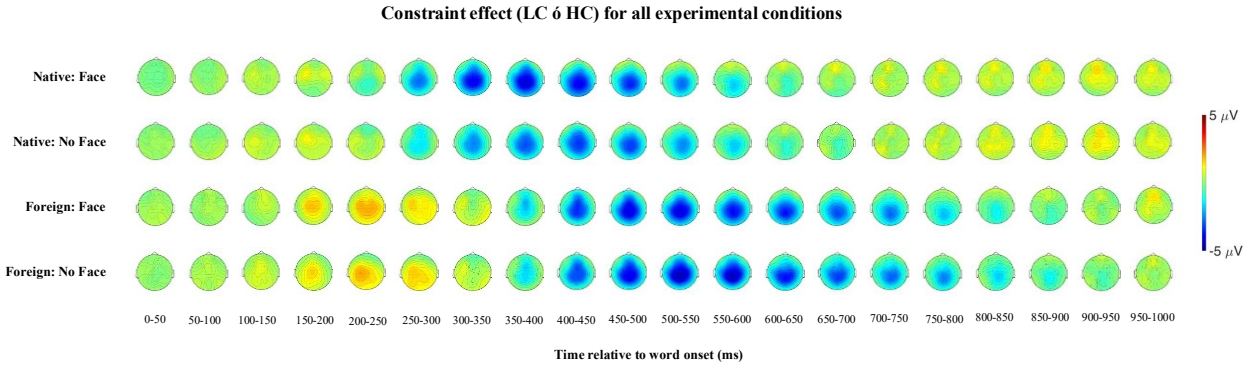
Figure 4.2 presents the grand-average ERP response for all experimental conditions in nine representative electrode sites (Fz, F3, F4, Cz, C3, C4, Pz, P3, P4), along with topographic maps in the N400 time-window.

Figure 4.2. Grand-average ERP waveforms and topographic maps for all experimental conditions. The shaded area highlights the N400 time-window (300-500 ms). The 9 electrodes were chosen to represent the EEG signal in frontal, central, and parietal regions, and across the left, midline, and right hemispheres based on our electrode layout.



As expected, unpredictable words seem to elicit a larger negativity compared to predictable words within the N400 time-window, reflecting the effect of the Constraint condition. The Constraint effect seems to be influenced by the presence of the face cue and this modulation appears to be maximal on centro-parietal electrodes. We also observed that the Constraint effect seems to be present before and after the N400 time-window, with possible differences between the native and the foreign-accented conditions (Figure 4.3).

Figure 4.3. Scalp distribution of the Constraint effect over time according to Accent and Face conditions.



In the native-accent condition, unpredictable words seem to elicit a centro-parietal negativity relative to predictable words starting from 250-300 ms. In the foreign-accent condition, unpredictable words show an increased positivity between 150–300 ms, followed by a centro-parietal negativity emerging from 350–400 ms. Due to this complex pattern, we conducted two types of analyses: an a priori analysis focused on the N400 time-window, and a post-hoc temporal Exploratory Factor Analysis (EFA) aimed at more precisely identifying the ERP components driving the observed effects. Separate Temporal EFAs were conducted for the native and foreign-accent conditions, as visual inspection of grand-average waveforms suggested differences in component structure or latency between accent conditions.

N400: 300-500 ms

Table 4.3 shows the model comparison for predicting the N400 amplitude. The model that provided the best fit (lower delta AIC and higher AIC weight) was Model 5:

*Amplitude ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + (Constraint|Participant) + (1|Item)*

Table 4.3. Model comparison for predicting the N400 amplitude. Deviance = residual deviance; dAIC = difference between AIC of each model and the model with lower AIC; AICw = AIC weight.

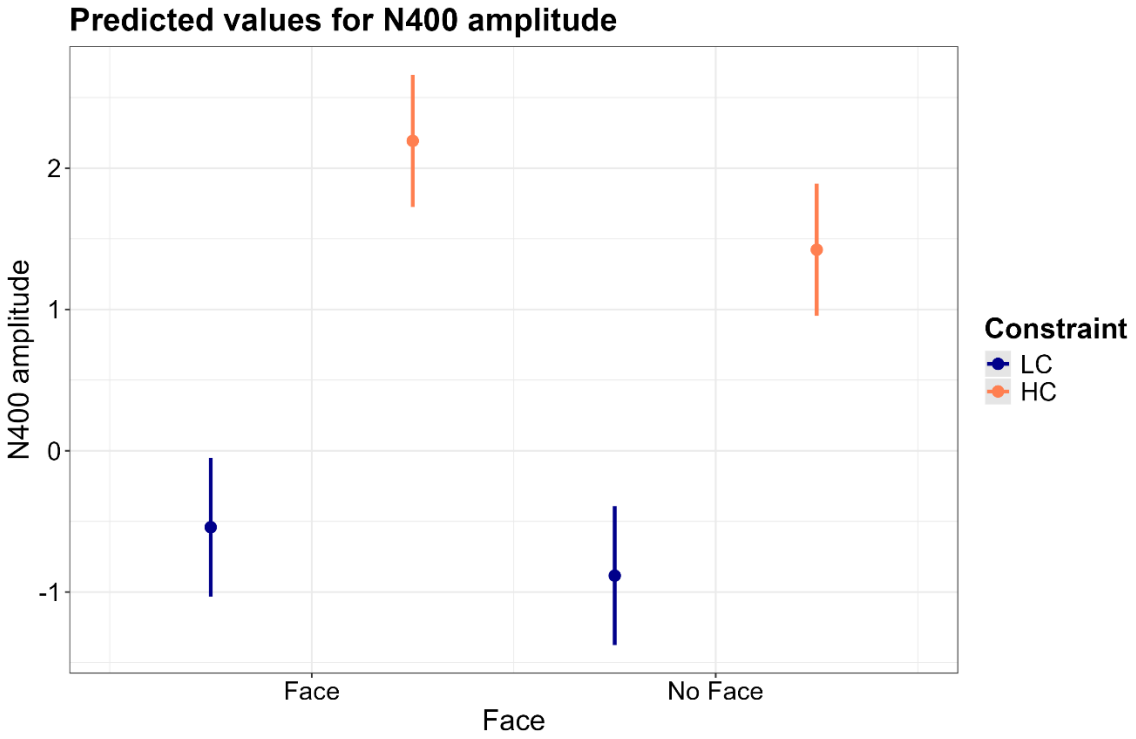
Models	**Deviance**	**dAIC**	**AICw**
M0. Amplitude ~ (Constraint|Participant) + (1|Item)	89293.77	166.67	0.0
M1. Amplitude ~ Accent + (Constraint|Participant) + (1|Item)	89239.48	114.38	0.0
M2. Amplitude ~ Accent + Constraint + (Constraint|Participant) + (1|Item)	89179.38	56.29	0.0
M3. Amplitude ~ Accent + Constraint + Face + (Constraint|Participant) + (1|Item)	89150.66	29.56	0.0
M4. Amplitude ~ Accent + Constraint + Face + Constraint*Accent + (Constraint|Participant) + (1|Item)	89121.36	2.26	0.15
M5. Amplitude ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + (Constraint|Participant) + (1|Item)	**89117.09**	**0.0**	**0.46**
M6. Amplitude ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + Accent*Face + (Constraint|Participant) + (1|Item)	89116.87	1.78	0.19
M7. Amplitude ~ Accent + Constraint + Face + Constraint*Accent + Constraint*Face + Accent*Face + Constraint*Accent*Face + (Constraint|Participant) + (1|Item)	89114.73	1.63	0.20

Table 4.4 shows the results of the best-fitting model (Model 5). As all main effects are implied in interactions, we focus on the latter. The interaction between Constraint and Accent indicates that the Constraint effect, namely larger N400 amplitude for unpredictable words compared to predictable words, is stronger for the native compared to the foreign accent condition. Crucially, the interaction between Constraint and Face indicates that cueing the speaker identity is associated with a larger Constraint effect compared to when speaker identity is not cued (Fig. 4.4). Post-hoc comparisons showed that cueing the speaker's identity is associated with smaller N400 amplitude for predictable (b = 0.77, z-ratio = 5.258, $p < .001$) but not for unpredictable words (b = 0.343, z-ratio = 2.343, $p = .076$).

Table 4.4. Results of the best-fitting model for N400 amplitude (Model 5).

	Estimate	CI (95%)	Std. Error	*t-value*	*p-value*
Intercept	0.548	[0.143 0.952]	0.204	2.683	.010
Accent [Foreign]	0.381	[0.280 0.483]	0.052	7.383	< .001
Constraint [HC]	1.260	[1.036 1.484]	0.113	11.135	< .001
Face [Face]	0.278	[0.177 0.380]	0.052	5.376	< .001
Constraint[HC] Accent[Foreign]*	-0.280	[-0.381 -0.179]	0.052	-5.417	< .001
Constraint[HC] Face[Face]*	0.107	[0.005 0.208]	0.052	2.064	.039

Figure 4.4. Model estimates for the interaction between Constraint and Face. The error bars indicate 95% confidence intervals.



4.4. Interim discussion

The present study investigated whether speaker identity (native vs. foreign) can be used to implement specific phonological predictions. To address this question, we employed an experimental paradigm that leverages the fact that non-native speakers frequently produce

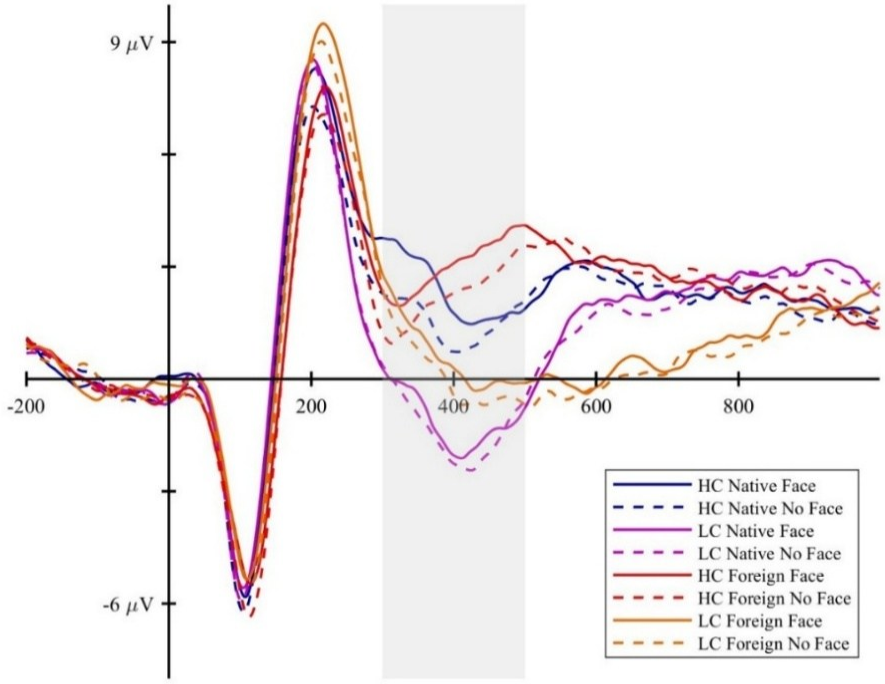
phonological errors. Participants read sentence frames that were missing a last word to be grammatical. The final word was produced by either a native or a foreign-accented speaker. The foreign-accented speaker made consistent phonological errors on the first phoneme of the spoken target. Depending on sentence constraint, the target word was either predictable or unpredictable. Crucially, speaker identity was either cued or not by an image of the speaker's face, thus manipulating the availability of information about the phonological features of the upcoming spoken target. We hypothesized that the face cue would be associated with a larger N400 predictability effect, reflecting a facilitation in the processing of predictable words due to phonological prediction.

The results confirmed our hypothesis: the N400 predictability effect, namely a larger N400 for unpredictable words compared to predictable words, was stronger when the speaker identity was cued compared to when it was not cued. Importantly, the face cue was associated with a smaller N400 for predictable words but not for unpredictable words, suggesting a facilitation in the processing of predictable words. These findings align with the results of Study 1, offering further evidence for the involvement of phonological representations in linguistic prediction and suggesting that prediction during language comprehension can target specific phonological forms.

4.5. Temporal Exploratory Factor Analysis

Figure 4.5 shows the grand-average ERP waveforms recorded at the electrode Cz for each experimental condition.

Figure 4.5. Grand-average ERP waveform in Cz electrode for each experimental condition. The shaded area highlights the N400 time-window (300-500 ms).



The grand-average ERPs seem to present different waveform shapes for native-accented versus foreign-accented words, especially from around 300 ms after word onset. In both conditions, positive deflections can be observed within the N400 time-window, although their latencies seem to differ across accent conditions. Additionally, the face cueing effect appears to emerge before the N400 time-window (in the foreign-accent condition, it can already be observed from around 100 ms, whereas in the native-accent condition it seems to emerge around 200 ms). This pattern suggests that ERP components different from the N400 may contribute to the face cueing effect.

To disentangle the ERP components underlying the effects observed in our study, we employed Temporal Exploratory Factor Analysis (EFA)[1]. This method was selected because traditional approaches to ERP data analysis, such as averaging the observed signal across a fixed time-window, often fail to account for the temporal and spatial overlap between components (Luck, 2014). Temporal EFA provides a data-driven approach to untangle

[1] In ERP research, both PCA (Principal Component Analysis; Dien, 2012; Dien & Frishkoff, 2005) and EFA (Exploratory Factor Analysis; Scharf et al., 2022) have been applied to estimate the underlying components of the ERP data. In practice, the outcomes of these two techniques often converge when applied to ERP data, given the high correlations between sampling points and the high number of variables (Dien & Frishkoff, 2005; Scharf & Nestler, 2019; Widaman, 1990, 2007).

overlapping signals based on patterns of covariance among sampling points (Dien, 2012; Dien & Frishkoff, 2005; Scharf et al., 2022). This technique allows for the extraction of a set of underlying factors summarizing sampling points with similar activity patterns across participants, electrodes, and conditions. Factors are assumed to represent an estimate of the ERP components underlying the observed signal, and are described by two coefficients: *factor loadings*, which indicate the contribution of the factor to a specific time point, and *factor scores*, which indicate the contribution of the factor to a specific observation (for more details, see Scharf et al., 2022).

We adopted the procedure reported by Scharf et al. (2022) to compute Temporal EFA. Single-trial data were averaged by participant, electrode, and conditions. Prior to performing Temporal EFA, the data were downsampled to 250 Hz. Since visual inspection of the waveforms suggested differences in latency or component structure between accent conditions, we conducted separate Temporal EFAs for the native and the foreign-accented condition. This approach is warranted when measurement invariance across conditions cannot be assumed (Beauducel & Hilger, 2018; Meredith, 1993; Möcks, 1986), such as when variations in the component structure or latency occur (Barry et al., 2016). Temporal EFA was applied to the trials-averaged waveforms, including all EEG channels and the baseline interval (Dien, 2012). The EFA model was estimated based on the covariance matrix of the sampling points. The number of factors to extract was determined using the Empirical Kaiser Criterion (Braeken & van Assen, 2017; Y. Li et al., 2020). To obtain an interpretable factor solution, Geomin rotation was applied with 30 random start values and a delta parameter of 0.01 (Yates, 1987). Only factors explaining more than 3% of the total variance were considered for the statistical analyses.

To analyze amplitude effects, we used the peak amplitude (factor scores multiplied by the peak factor loading) of the factors reflecting ERP components of interest as dependent variable. We selected a cluster of centro-parietal electrodes (Cz, C3, C4, Pz, P3, P4) for factors reflecting components with a centro-parietal distribution (P3b and N400), and a cluster of fronto-central electrodes (Cz, C3, C4, Fz, F3, F4) for factors reflecting components with a fronto-central distribution (N1, P2, P3a). We also examined factors reflecting slow waves extending beyond the N400 time-window. In such cases, the choice of electrode cluster (anterior vs. posterior) depended on the observed scalp distribution, given the heterogeneity in the topography of slow waves (Van Petten & Luka, 2012). The factors' peak amplitude was analyzed using repeated measures ANOVA including the following predictors: i) Constraint, ii) Face, iii) Constraint*Face. Post-hoc comparisons were conducted using the "contrast" function

of the *emmeans* package (Lenth, 2025), applying Bonferroni correction to p-values (Bonferroni, 1936).

4.5.1. Native-accent Temporal EFA

A total of 23 factors were extracted for the native-accent condition, accounting for 95% of total variance. The corresponding factor loadings are shown in Figure 4.6.

Figure 4.6. Unstandardized Factor Loadings from the Geomin-rotated Temporal EFA in the native-accent condition. Each line represents the factor loadings of a specific factor. Larger factor loadings indicate a greater contribution of the factor at a given sampling point. Factors are ordered according to the proportion of explained variance.

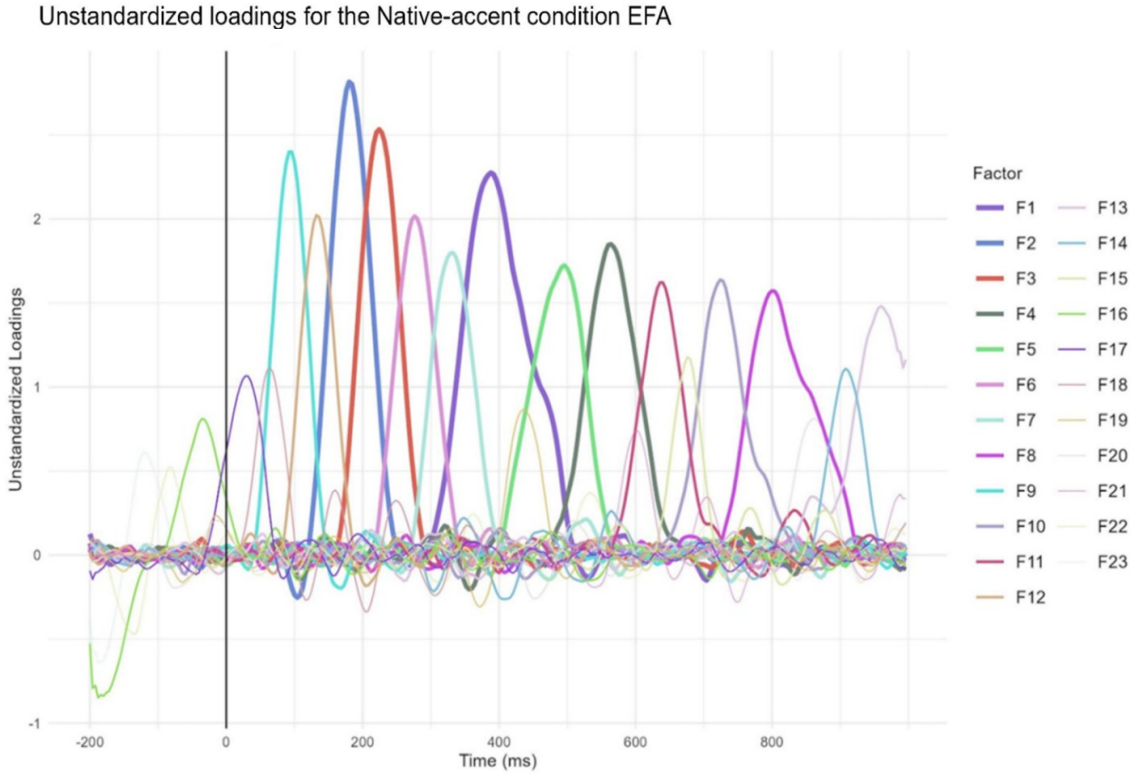
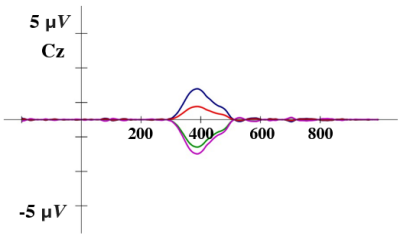
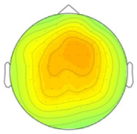
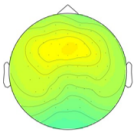
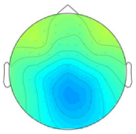
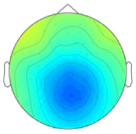
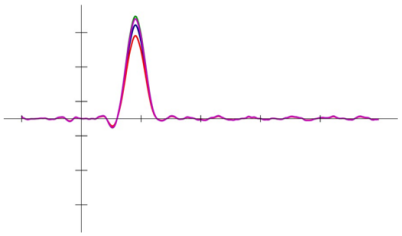
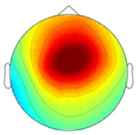
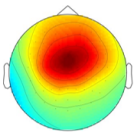
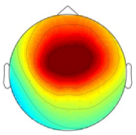
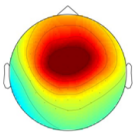
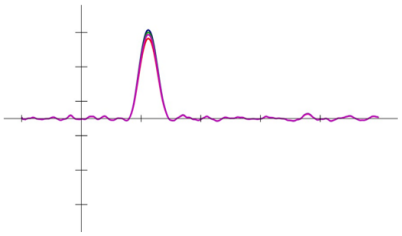
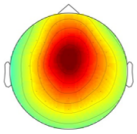
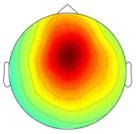
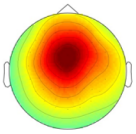
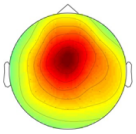
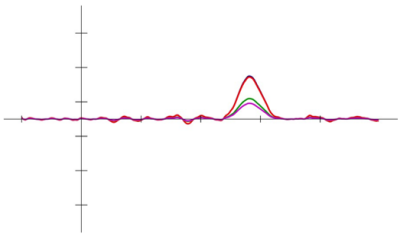
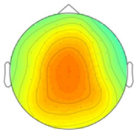
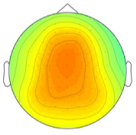
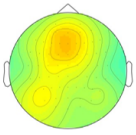
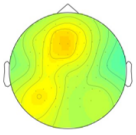
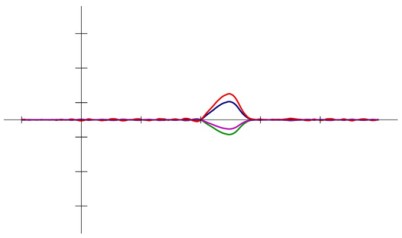
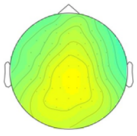
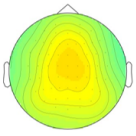
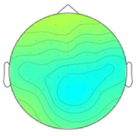
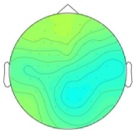
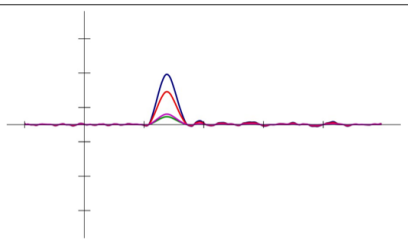
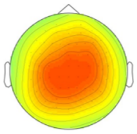
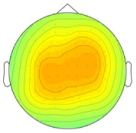
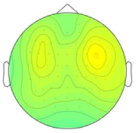
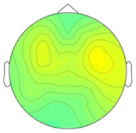
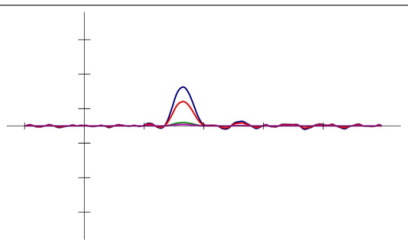
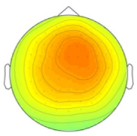
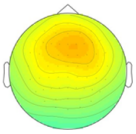
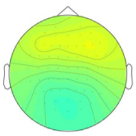
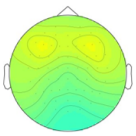


Figure 4.7 presents the time-course and scalp distribution of the factors explaining more than 3% of total variance.

Figure 4.7. For each factor that explains more than 3% of total variance, the time-course at electrode Cz and the scalp distribution of the peak amplitude are presented. VE: percentage of total variance explained.

Factor	VE	Timecourse	HC: Face	HC: No Face	LC: Face	LC: No Face
1	11.2					
2	10.4					
3	9.2					
4	6.7					
5	6.4					
6	6.0					
7	5.6					

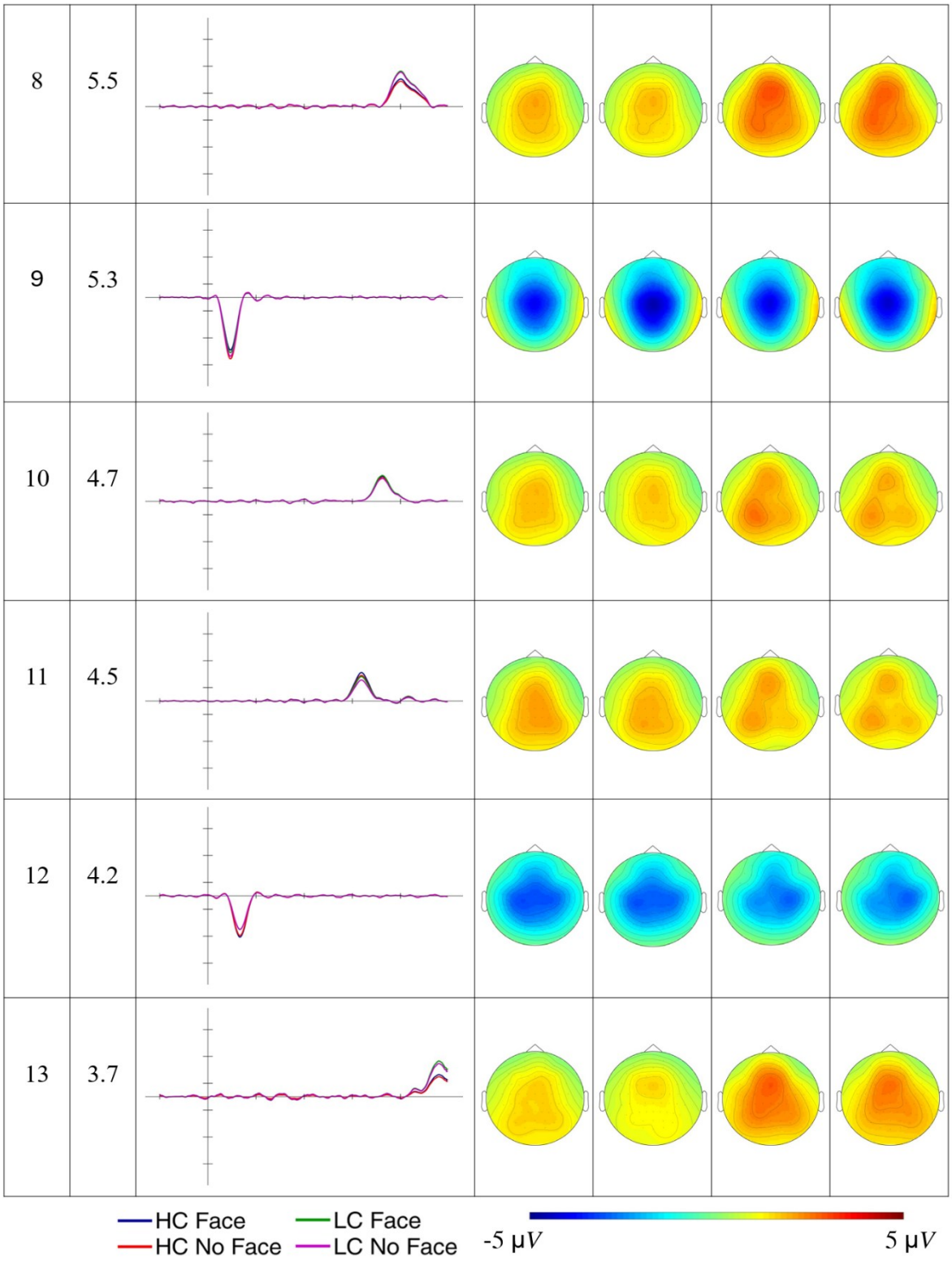
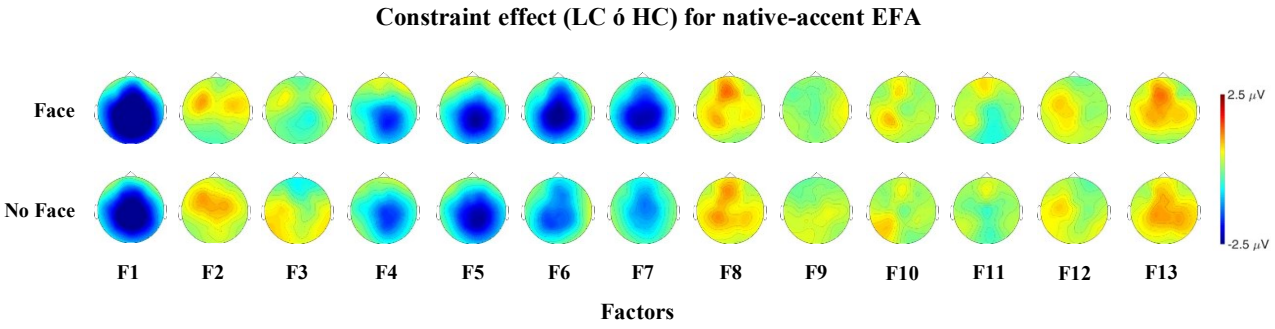


Figure 4.8 presents the scalp distribution of the Constraint effect for all factors accounting for more than 3% of total variance explained.

Figure 4.8. Scalp distribution of the Constraint effect for factors accounting for more than 3% of total variance explained, shown at the peak amplitude of each factor.



We focused our analysis on Factors from 1 to 9, as Factors 10 to 12 lacked a clear functional interpretation or showed no modulation by our experimental variables, while F13 extended beyond the analyzed time window. The outcomes of the statistical models are reported in Table 4.5.

Factor 9 presents a central negativity peaking at 92 ms, likely corresponding to the N1. Factor 2 shows a fronto-central positivity peaking at 180 ms, resembling a typical P2 response. Factor 9 does not seem to be influenced by the experimental conditions, whereas Factor 2 shows a Constraint effect, namely larger fronto-central positivity for unpredictable words compared to predictable words (b = -0.751, t-ratio = -4.424, $p < .001$). Factor 3 presents a fronto-central positivity peaking at 224 ms, showing both temporal and spatial similarities with Factor 2. It is possible that these factors reflect a temporal modulation of the P2 response, divided into separate factors through statistical decomposition. Factor 3 does not seem to be modulated by any of the experimental conditions.

Factors 6 and 7 present a positive deflection peaking around 300 ms, specifically at 276 and 332 ms, respectively. Factor 6 is maximal on central electrodes, while Factor 7 extends more toward fronto-central electrodes. Crucially, both factors seem to present a larger posterior positivity for predictable words relative to unpredictable words, with this effect being more pronounced when the face cue is present (Fig. 4.8). This pattern may reflect a P3b response. The interaction between Constraint and Face in Factors 6 and 7 confirms that the Constraint effect, namely larger centro-parietal positivity for predictable words compared to unpredictable words, is larger when the speaker identity is cued compared to when it is not cued. Post-hoc comparisons showed that cueing the speaker's identity elicited a larger centro-parietal positivity for predictable words (F6: b = 0.922, t-ratio = 3.747, $p = .002$; F7: b = 0.86, t-ratio = 3.372, $p = .007$), whereas no such effect was found for unpredictable words (F6: b = 0.064, t-ratio =

0.270, $p > .999$; F7: b = -0.169, t-ratio = -0.685, $p > .999$).

Factors 1 and 5 exhibit a negative deflection for unpredictable words peaking around 400 ms, specifically at 388 ms and 496 ms, respectively. Both factors seem to present a centro-parietal negativity for unpredictable words relative to predictable words, which aligns with the scalp distribution of the N400 (Fig. 4.8). The Constraint effect in both Factors 1 and 5 confirms that unpredictable words elicit a centro-parietal negativity compared to predictable words (F1: b = 2.99, t-ratio = 12.066, $p < .001$; F5: b = 1.99, t-ratio = 10.003, $p < .001$). Factor 1 presents a Face effect (b = 0.747, t-ratio = 3.922, $p < .001$), indicating that cueing the speaker's identity elicits a reduced centro-parietal negativity compared to when the speaker's identity is not cued. However, we also observed that both Factors 1 and 5 show a fronto-central positivity in response to predictable words, in addition to the posterior negativity in response to unpredictable words (Fig. 4.7). This may indicate the presence of temporally overlapping ERP components with different topographies, complicating the interpretation of the observed effects as modulations of the N400.

Following the N400 time-window (300-500 ms), Factors 4 and 8 seem to present different patterns of activity as a function of sentence constraint (Fig. 4.8). Factor 4 shows a centro-parietal negativity for unpredictable words compared to predictable words (b = 1.34, t-ratio = 5.031, $p < .001$), whereas Factor 8 presents a fronto-central positivity for unpredictable words compared to predictable words (b = -0.828, t-ratio = -3.571, $p < .001$). The face cue had no effect on these factors.

Table 4.5. Results of the ANOVA models; putative ERP component labels are reported.

	df	**MSE**	**F-value**	**η_p^2**	**p-value**
F1 (N400)					
Constraint	1, 41	2.58	145.58	.780	<.001***
Face	1, 41	1.52	15.38	.273	<.001***
Constraint*Face	1, 41	1.12	3.74	.084	.060
F2 (P2)					
Constraint	1, 41	1.21	19.57	.323	<.001***
Face	1, 41	1.62	2.97	.068	.092
Constraint*Face	1, 41	1.59	0.45	.011	.507
F3 (P2)					
Constraint	1, 41	1.11	0.08	.002	.779

Face	1, 41	1.16	2.23	.052	.143
Constraint*Face	1, 41	1.50	0.02	<.001	.878
F4 (Slow wave)					
Constraint	1, 41	2.99	25.31	.382	<.001***
Face	1, 41	1.08	0.22	.005	.642
Constraint*Face	1, 41	0.71	0.01	<.001	.937
F5 (N400)					
Constraint	1, 41	1.66	100.07	.709	<.001***
Face	1, 41	1.49	2.21	.051	.145
Constraint*Face	1, 41	1.12	0.15	.004	.696
F6 (P3b)					
Constraint	1, 41	2.05	65.09	.614	<.001***
Face	1, 41	1.52	6.73	.141	.013*
Constraint*Face	1, 41	0.94	8.19	.166	.007**
F7 (P3b)					
Constraint	1, 41	2.00	53.71	.567	<.001***
Face	1, 41	1.14	4.42	.097	.042*
Constraint*Face	1, 41	1.50	7.41	.153	.009**
F8 (Slow wave)					
Constraint	1, 41	2.26	12.75	.237	<.001***
Face	1, 41	0.87	0.65	.016	.426
Constraint*Face	1, 41	0.92	0.04	<.001	.843
F9 (N1)					
Constraint	1, 41	1.00	0.11	.003	.740
Face	1, 41	1.67	2.66	.061	.110
Constraint*Face	1, 41	0.82	0.38	.009	.540

* *p*-value < .05, ** < .01, *** < .001

4.5.2. Foreign-accent Temporal EFA

A total of 23 factors were extracted for the foreign-accent condition, accounting for 96% of the total variance. The corresponding factor loadings are shown in Figure 4.9.

Figure 4.9. Unstandardized Factor Loadings from the Geomin-rotated Temporal EFA in the foreign-accent condition. Each line represents the factor loadings of a specific factor. Larger factor loadings indicate a greater contribution of the factor at a given sampling point. Factors are ordered according to the proportion of explained variance.

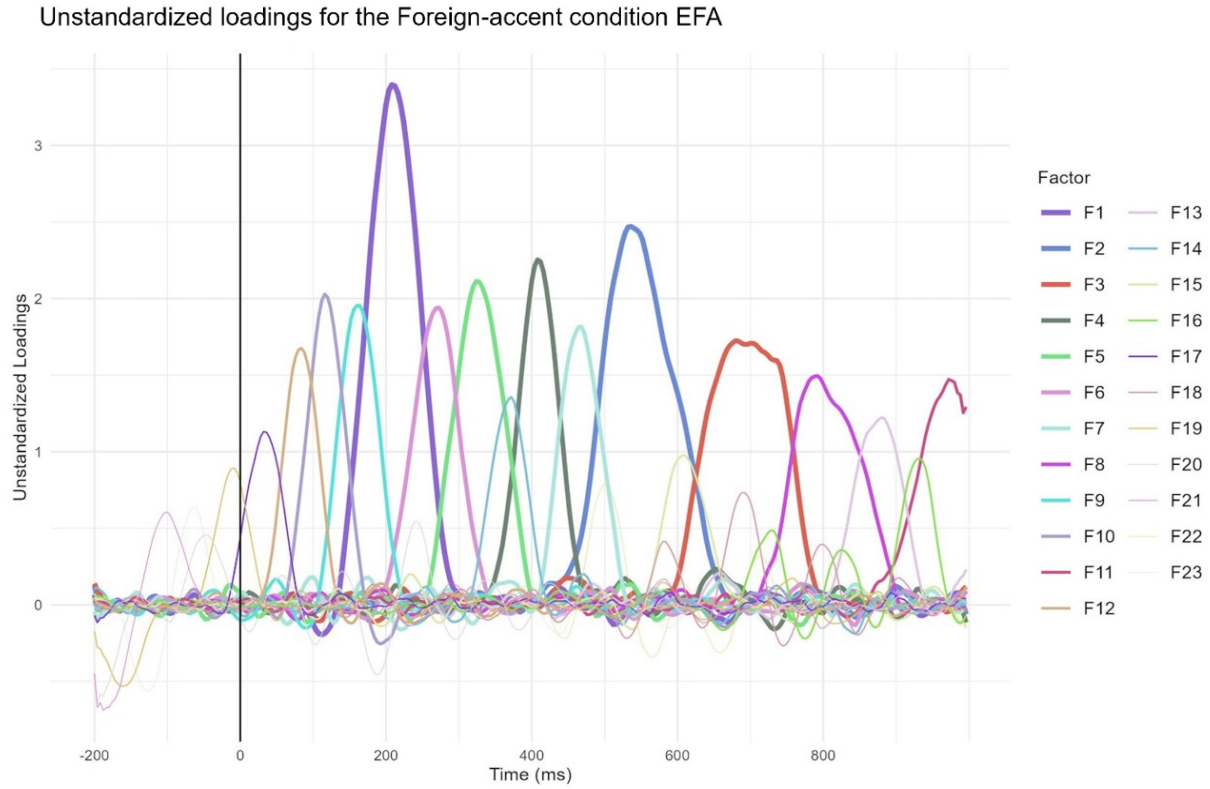
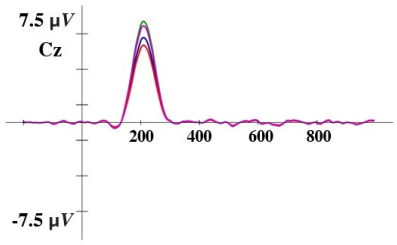
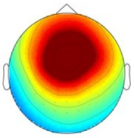
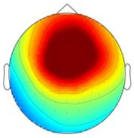
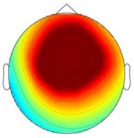
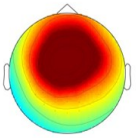
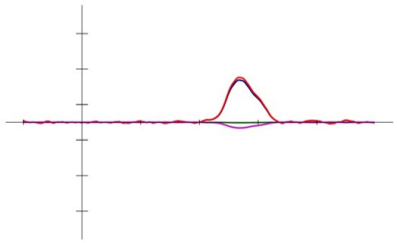
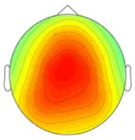
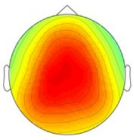
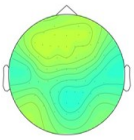
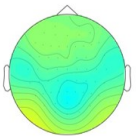
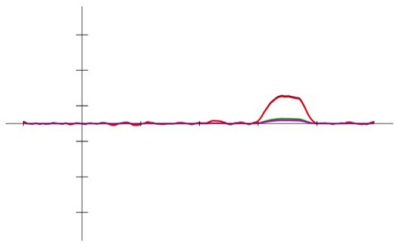
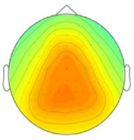
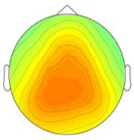
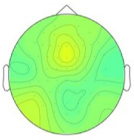
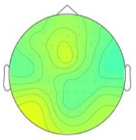
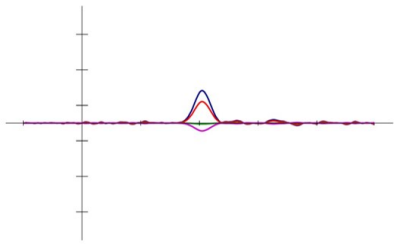
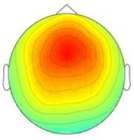
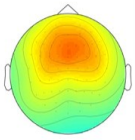
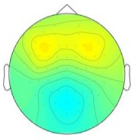
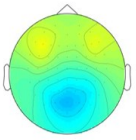
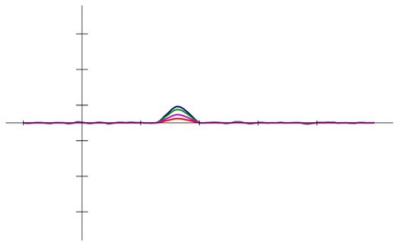
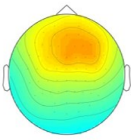
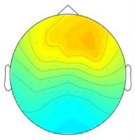
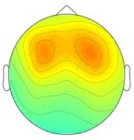
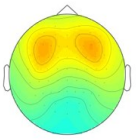
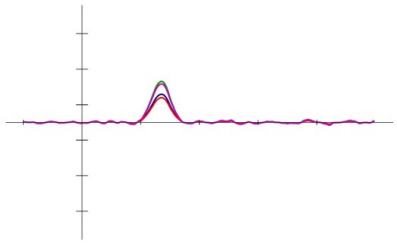
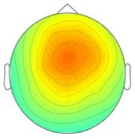
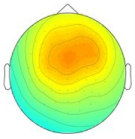
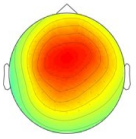
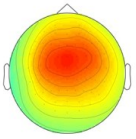
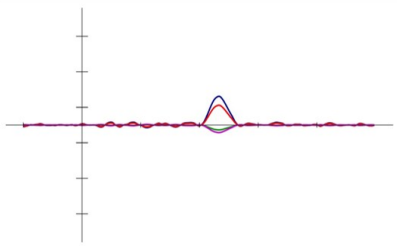
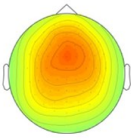
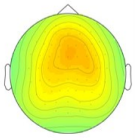
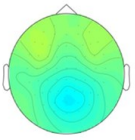


Figure 4.10 presents the time-course and scalp distribution of the factors explaining more than 3% of the total variance.

Figure 4.10. For each factor that explains more than 3% of total variance, the time-course at electrode Cz and the scalp distribution of the peak amplitude are presented. VE: percentage of total variance explained.

Factor	VE	Timecourse	HC: Face	HC: No Face	LC: Face	LC: No Face
1	16.5					
2	12.5					
3	8.5					
4	6.6					
5	6.5					
6	6.1					
7	5.6					

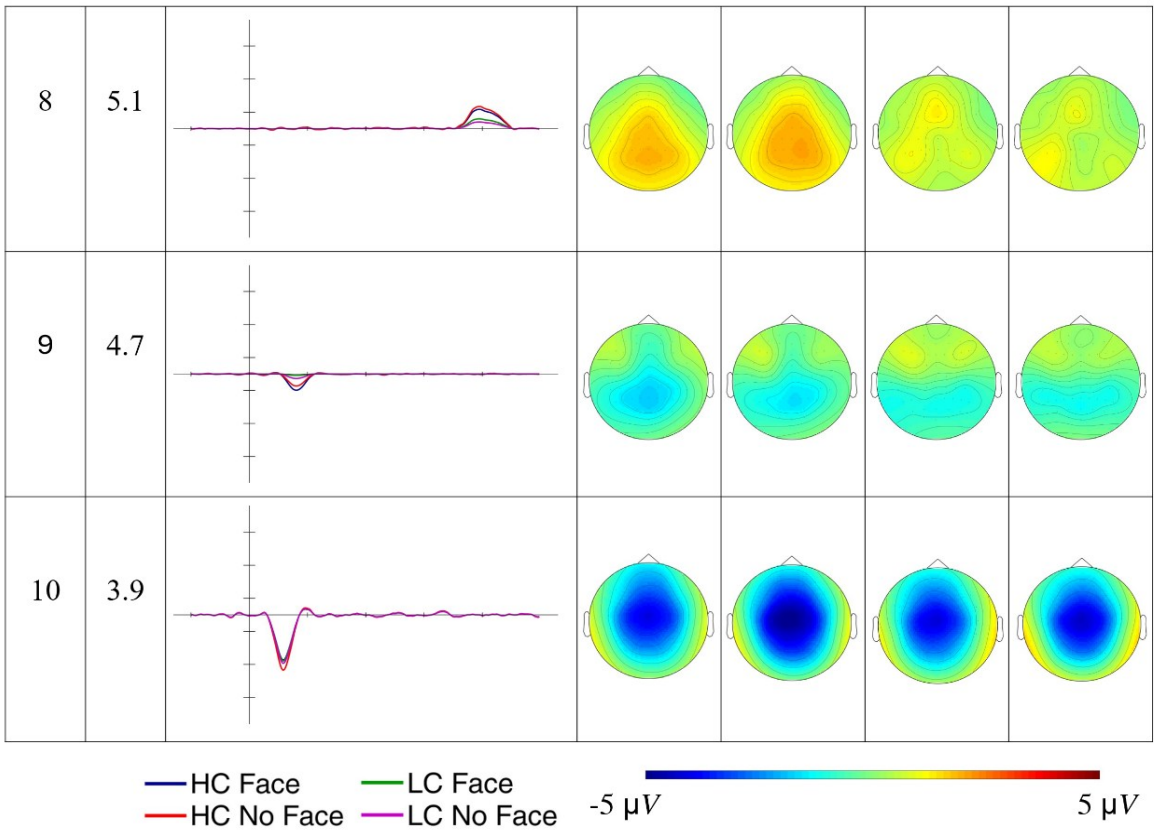
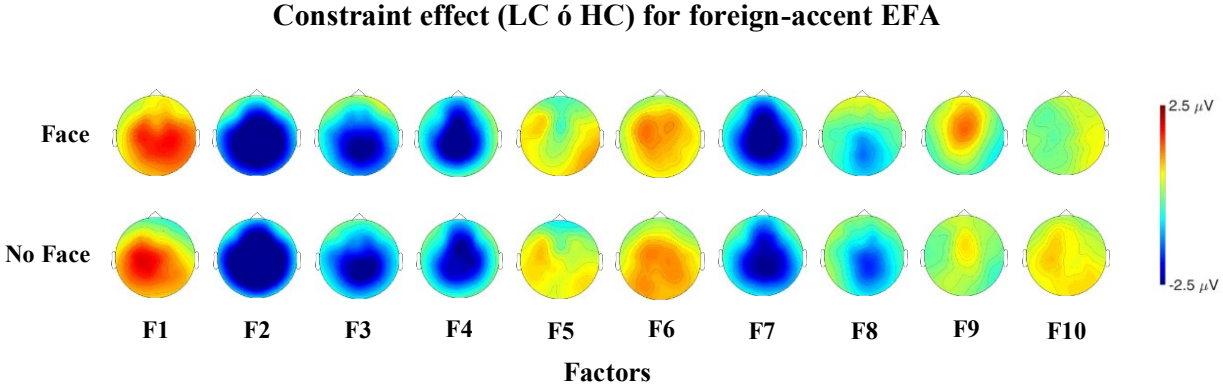


Figure 4.11 presents the scalp distribution of the Constraint effect for all factors accounting for more than 3% of total variance explained.

Figure 4.11. Scalp distribution of the Constraint effect for factors accounting for more than 3% of total variance explained, shown at the peak amplitude of each factor.



Factors from 1 to 10 were considered reflecting ERP components of interest and were analyzed. The outcomes of the statistical models are reported in Table 4.6.

Factor 10 presents a central negativity peaking at 116 ms, likely corresponding to the N1. This factor shows a Face effect, with reduced fronto-central negativity when the speaker's identity is cued compared to when it is not cued (b = 0.439, t-ratio = 2.416, p = .02). Factor 1 presents a fronto-central positivity peaking at 208 ms, resembling a typical P2 response. This factor shows a Constraint effect, with larger fronto-central positivity for unpredictable words compared to predictable words (b = -1.2, t-ratio = -4.757, p < .001).

Factors 6 and 5 present a positive deflection peaking around 300 ms, specifically at 272 and 324 ms, respectively. Both factors seem to be maximal on fronto-central electrodes, and Factor 6 seems to present a larger positivity for unpredictable words relative to predictable words (Fig. 4.11). Given their latency and scalp topography, these factors may reflect a temporal modulation of the P3a response. Factor 6 presents a Constraint effect, showing a larger fronto-central positivity for unpredictable words compared to predictable words (b = -1.0, t-ratio = -5.078, p < .001). Factor 5 presents a Face effect, showing a larger fronto-central positivity when the speaker's identity is cued compared to when it is not cued (b = 0.502, t-ratio = 3.03, p = .004).

Factors 4 and 7 exhibit a negative deflection for unpredictable words peaking around 400 ms, specifically at 408 ms and 468 ms, respectively. Both factors seem to present a centro-parietal negativity for unpredictable words relative to predictable words, which aligns with the scalp distribution of the N400 (Fig. 4.11). The Constraint effect in both Factors 4 and 7 confirms that unpredictable words elicit a centro-parietal negativity compared to predictable words (F4: b = 2.35, t-ratio = 7.547, p < .001; F7: b = 2.4, t-ratio = 9.645, p < .001). Both factors present a Face effect, indicating reduced centro-parietal negativity when the speaker's identity is cued compared to when it is not cued (F4: b = 0.708, t-ratio = 3.375, p = .002; F7: b = 0.432, t-ratio = 2.429, p = .02). As in the native-accent condition, the simultaneous presence of fronto-central positivity for predictable words and posterior negativity for unpredictable words (Fig. 4.10), complicates the interpretation of potential modulations of the N400 amplitude.

Following the N400 time-window (300-500 ms), Factors 2, 3 and 8 present a centro-parietal negativity for unpredictable words compared to predictable words (F2: b = 3.93, t-ratio = 12.205, p < .001; F3: b = 2.22, t-ratio = 9.55, p < .001; F8: b = 1.1, t-ratio = 5.499, p < .001). The face cue had no effect on these factors.

Table 4.6. Results of the ANOVA models; putative ERP component labels are reported.

	df	MSE	F-value	η_p^2	p-value
F1 (P2)					
Constraint	1, 41	2.66	22.63	.356	<.001***
Face	1, 41	1.51	3.35	.076	.074
Constraint*Face	1, 41	1.86	0.09	.002	.760
F2 (Slow wave)					
Constraint	1, 41	4.35	148.95	.784	<.001***
Face	1, 41	2.06	0.00	<.001	.991
Constraint*Face	1, 41	1.41	3.76	.084	.059
F3 (Slow wave)					
Constraint	1, 41	2.26	91.20	.690	<.001***
Face	1, 41	0.83	0.02	<.001	.901
Constraint*Face	1, 41	0.88	0.09	.002	.770
F4 (N400)					
Constraint	1, 41	4.06	56.96	.581	<.001***
Face	1, 41	1.85	11.39	.217	.002**
Constraint*Face	1, 41	0.82	0.88	.021	.353
F5 (P3a)					
Constraint	1, 41	2.51	0.71	.017	.405
Face	1, 41	1.16	9.18	.183	.004**
Constraint*Face	1, 41	1.35	0.29	.007	.590
F6 (P3a)					
Constraint	1, 41	1.64	25.78	.386	<.001***
Face	1, 41	0.89	3.18	.072	.082
Constraint*Face	1, 41	1.07	0.14	.003	.710
F7 (N400)					
Constraint	1, 41	2.61	93.03	.694	<.001***
Face	1, 41	1.33	5.90	.126	.020*
Constraint*Face	1, 41	1.33	0.69	.016	.413
F8 (Slow wave)					

Constraint	1, 41	1.67	30.24	.424	<.001***
Face	1, 41	1.02	0.01	<.001	.932
Constraint*Face	1, 41	1.40	1.59	.037	.214
F9 (Early negativity)					
Constraint	1, 41	1.36	16.97	.293	<.001***
Face	1, 41	1.18	0.08	.002	.775
Constraint*Face	1, 41	1.45	2.64	.060	.112
F10 (N1)					
Constraint	1, 41	1.24	3.76	.084	.059
Face	1, 41	1.39	5.84	.125	.020*
Constraint*Face	1, 41	1.23	2.99	.068	.091

* *p*-value < .05, ** < .01, *** < .001

4.5.3. Summary of Temporal EFA analysis results

Native-accent condition

We observed a main effect of Constraint in factors possibly reflecting the P2, P3b, N400 and slow waves. Unpredictable words elicited a larger P2 compared to predictable words. Predictable words elicited a larger P3b compared to unpredictable words, whereas the N400 was smaller for predictable words compared to unpredictable words. After the N400 time-window, unpredictable words showed a prolonged centro-parietal negativity relative to predictable words, and a subsequent fronto-central positivity. We observed a main effect of Face in a factor possibly reflecting the N400, with reduced N400 amplitude when the speaker identity was cued. Nevertheless, the presence of fronto-central positivity for predictable words complicates the interpretation of these effects as direct modulations of the N400. Finally, we observed an interaction between Constraint and Face in factors possibly reflecting a P3b response, indicating that cueing the speaker's identity elicits a larger P3b for predictable but not for unpredictable words.

Foreign-accent condition

We observed a main effect of Constraint in factors possibly reflecting the P2, P3a, N400 and slow waves. Unpredictable words elicited larger P2, P3a and N400 compared to predictable words. After the N400 time-window, unpredictable words showed a prolonged centro-parietal negativity relative to predictable words. In contrast to the native-accent condition, no late

fronto-central positivity was observed for unpredictable words relative to predictable words. We observed a main effect of Face in factors possibly reflecting the N1, P3a and N400. Cueing the speaker identity was associated with a smaller N1 and N400 compared to when speaker identity was not cued, whereas the P3a was larger when the speaker identity was cued. As in the native accent condition, the presence of fronto-central positivity for predictable words complicates the interpretation of effects possibly reflecting modulations of the N400. In contrast to the native-accent condition, no significant interactions between Constraint and Face were found in the factors examined.

4.6. Discussion

4.6.1. Facilitation in processing native and foreign-accented words

The present study aimed to investigate whether cueing the speaker's identity, and thus the speaker-specific phonology, allows comprehenders to anticipate the specific phonological form of a highly predictable word. The results of our main analysis aligned with this hypothesis, showing that cueing the speaker identity is associated with a reduced negativity in the N400 time-window for predictable but not for unpredictable words. We interpreted these findings as suggesting that phonological prediction facilitates the processing of predictable words.

Visual inspection of the waveforms revealed the presence of positive deflections within the N400 time-window in response to predictable words, suggesting that ERP components distinct from the N400 may have contributed to the observed effect. The positive deflections observed within the N400 time-window might be related to our experimental task, which required participants to indicate whether they expected the spoken target in a subset of trials. Brothers et al. (2015) used an experimental paradigm similar to the one adopted in the present study. In their study, participants read discourse contexts that ended with either a low-cloze or medium-cloze word (1 or 50% cloze probability, respectively). In every trial, they were required to indicate whether they expected the final word. The authors reported that ERP responses to predicted medium-cloze words began to diverge from those to unpredicted medium-cloze words around 200–300 ms after word onset. They interpreted this early difference as a modulation of the N250, an ERP component thought to reflect word form processing (Grainger & Holcomb, 2009), suggesting that participants predicted sublexical information. Nieuwland (2019) proposed an alternative interpretation of this effect, observing that predicted medium-cloze words elicited a pronounced positive deflection, possibly reflecting a P3b response.

There is a general agreement that a distinction can be made between two subcomponents of the P300 component: the P3a (or novelty P3) and P3b (or target P3). The P3a has a fronto-

central scalp distribution and is typically observed in response to infrequent task-irrelevant stimuli in the oddball paradigm (Friedman & Simpson, 1994; Polich, 2007; Spencer et al., 1999). Instead, the P3b presents a centro-parietal scalp distribution and is typically observed in response to task-relevant stimuli (Kok, 2001; Polich, 2007). The amplitude of the P3b is thought to reflect the extent to which the sensory input matches an internal representation during event categorization, with a larger amplitude reflecting a greater degree of correspondence (Kok, 2001). Psycholinguistic research has shown that the P3b can be elicited in response to different types of linguistic stimuli, such as collocations (Molinaro & Carreiras, 2010), idioms (Vespignani et al., 2010), antinomies (Roehm et al., 2007), and predictable verbs in studies involving acceptability judgment tasks (Freunberger & Roehm, 2016). Given that our experimental task required participants to categorize target stimuli, predictable words may have elicited a P3b response, which could have been particularly enhanced when the speaker's identity was cued. The face cue may have enabled participants to predict the specific phonological form of the upcoming word, thereby increasing the degree of correspondence between the internal representation of the upcoming word and the spoken target.

Temporal EFA analysis (Dien, 2012; Dien & Frishkoff, 2005; Scharf et al., 2022) showed that, in the native-accent condition, predictable words elicited a centro-parietal positivity relative to unpredictable words, possibly reflecting a P3b response. Crucially, the posterior positivity for predictable words was larger when the speaker identity was cued compared to when it was not, suggesting that comprehenders used the face cue to anticipate the phonological form of the upcoming spoken target. According to Nieuwland (2019), it is unclear whether P3b responses elicited in tasks requiring explicit decisions about linguistic stimuli reflect stimulus recognition, word-form processing or later decision-related processes. Our results provide some insights into this issue. In our study, participants were asked to indicate whether they expected the spoken target, regardless of how it was pronounced. While word predictability likely influenced participants' decisions, the face cue held no specific relevance for the task. Thus, the observed modulation of the P3b due to the face cue is more likely to reflect facilitated recognition or word-form processing, rather than decision-related processes.

In the foreign-accent condition, the face cue was associated with a modulation of the N1 and P3a responses to the spoken target. The N1 was reduced when the speaker identity was cued, regardless of word predictability. Traditionally, this component has been associated with early perceptual processing (Näätänen & Picton, 1987). Research on auditory perception suggests that the N1 response may index the detection of changes in the acoustic environment (Hyde, 1997), and its amplitude tends to increase when attentional demands on auditory stimuli

are heightened (Hillyard et al., 1973; Knight et al., 1981; Mangun, 1995; Ritter et al., 1988). The modulation of the N1 observed in the foreign-accent condition may be related to the phonological manipulation used in the present study. During speech perception, maintaining uncertainty about phonemic categories can be advantageous, especially when the signal is ambiguous, since later portions of the speech stream often provide clarifying information. This is evident in right-context effects, where later speech segments influence how earlier sounds are interpreted (Bard et al., 1988; Connine et al., 1991; Dahan, 2010; Grosjean, 1985). In our study, cueing the foreign speaker's identity may have led participants to delay the categorization of the initial phoneme of the spoken target, which is where the foreign-accented speaker diverged most notably from native phonology. Listeners might have allocated fewer attentional resources at word onset, as indicated by the suppression of the N1 response, and waited for additional information that could guide interpretation.

Foreign-accented words also elicited a P3a, which is typically interpreted as reflecting processes related to the orienting response (Cycowicz & Friedman, 1997; Friedman et al., 1998; Knight & Nakada, 1998). This component does not simply reflect the detection of a deviant or novel event per se, but rather the engagement of cognitive resources to evaluate the stimulus and determine an appropriate response (Justo-Guillén et al., 2019; Polich, 2007). In our study, foreign-accented words elicited a P3a rather than a P3b response, possibly due to the participants' limited familiarity with our phonological manipulation. The P3a showed an early modulation due to the predictability of the target word, with unpredictable words eliciting a larger P3a compared to predictable words. A later modulation emerged in response to the face cue, with a larger P3a elicited when the speaker's identity was cued compared to when it was not. The early modulation of the P3a may reflect an increased allocation of cognitive resources to recognize unpredictable words with non-standard phonology, as the sentence context does not provide cues for word recognition. The late modulation might be related to the phonological manipulation at word onset. When the foreign speaker's identity is cued, the system might delay categorizing the initial phoneme, awaiting additional information. Consequently, more cognitive resources may be allocated to extract information from later segments of the word that could support phonemic categorization at word onset.

In both accent conditions, unpredictable words elicited a sustained centro-parietal negativity relative to predictable words, extending beyond the canonical N400 time-window. Stowe et al. (2018) observed that sentence-final negativities typically emerge in tasks involving decisions about linguistic stimuli, suggesting a role in maintaining task-relevant information. This is consistent with ERP studies showing slow negative waves in conditions that engage

working memory (Lang et al., 1987; McCallum et al., 1988; Peronnet & Farah, 1989; Ruchkin et al., 1988, 1995). In low constraint contexts, where several sentence completions are plausible, the system may hold multiple candidates in working memory. This increased uncertainty could have made the decision-making process more complex compared to highly constraining contexts, where a single candidate is far more likely, resulting in a sustained negativity for unpredictable words.

Finally, our results showed a late frontal positivity for unpredictable words relative to predictable words in the native-accent condition, but not in the foreign-accent condition. This is consistent with previous empirical evidence showing that unpredictable yet semantically congruent sentence completions elicit late frontal positivities compared to predictable sentence completions (for a review, see Van Petten & Luka, 2012). Late frontal positivities have been proposed to reflect violations of specific lexical predictions rather than disconfirmed semantic expectations (Thornhill & Van Petten, 2012; Van Petten & Luka, 2012). In the native-accent condition, predictable words likely matched the expected lexical form, while unpredictable words did not. In contrast, in the foreign-accent condition, the non-standard phonology of the spoken target may have interfered with lexical access in both predictable and unpredictable words, making it more difficult to match the spoken target to the expected lexical form. Consistent with this interpretation, in the native-accent condition we observed a negativity for unpredictable relative to predictable words within the N400 time window, whereas in the foreign-accent condition this effect appeared later and was more sustained over time, possibly reflecting extended processing during word recognition due to non-standard phonology.

4.6.2. What do our data say about theories of linguistic prediction?

Taken together, the results of Studies 1 and 2 provide evidence supporting the involvement of phonological representations in prediction, at least in highly constraining and informative sentence contexts. In Study 1, we have shown that cueing the speaker identity (native vs. foreign), and hence the speaker-specific phonology, is associated with faster lexical decision times for predictable words. This effect did not seem to be modulated by the speaker's accent, suggesting that comprehenders implement speaker-specific phonological predictions for both the native and foreign-accented speaker. In Study 2, our analysis of the observed amplitude in the N400 time-window suggested that cueing the speaker's identity is associated with an easier processing of predictable words. Temporal EFA allowed us to clarify the ERP components underlying the face cueing effect and to investigate possible processing differences between accent conditions. In the native-accent condition, cueing the speaker's identity is associated

with a larger P3b response to predictable words, possibly reflecting a stronger match between the internal representation of the upcoming word and the spoken target due to phonological prediction. In the foreign-accent condition, we found no clear ERP components associated with a face cueing effect for predictable words. Rather, the face cueing effect appeared to occur regardless of word predictability and seemed to emerge earlier in time, as reflected by a reduced N1 and an enhanced P3a response. This pattern suggests that, when expecting a non-standard phonology, comprehenders may adopt a more flexible processing strategy, maintaining uncertainty about phonemic categories while waiting for further information to become available. In this condition, comprehenders seem to rely on a greater amount of bottom-up information to recognize the spoken word, given their reduced familiarity with the foreign speaker's phonemic categories, while in the native-accent condition they seem to generate specific predictions about the phonological form of the upcoming word.

Therefore, the results of Study 1 suggest that comprehenders implement phonological predictions for both the native and the foreign-accented speaker, while in Study 2, we found more robust evidence for phonological prediction in the native-accent condition. This apparent discrepancy might be related to differences in task demands. In Study 1, participants performed a lexical decision task, which required indicating as quickly and accurately as possible whether a spoken target presented after a written sentence frame was a real word or a non-word. In the native-accent condition, words were correctly pronounced, whereas in the foreign-accent condition, participants had to distinguish mispronounced real words from non-words. In contrast, Study 2 required participants to indicate whether they expected the last word of the sentence, regardless of how it was pronounced. The spoken target was always a real word, though its pronunciation varied depending on the speaker's accent. Participants may have implemented phonological predictions more flexibly in Study 1, as these predictions could help distinguish between mispronounced real words and non-words spoken by the foreign-accented speaker.

Our results align with current models of language comprehension that emphasize the sensitivity of predictive mechanisms to goal-directed behaviour (Kuperberg & Jaeger, 2016; Pickering & Strijkers, 2024). According to the model proposed by Kuperberg & Jaeger (2016), listeners can generate speaker-specific phonological predictions by constructing distinct generative models that reflect differences in the speech sound's structure. Internal generative models are built based on prior experience and are therefore likely to be more accurate and readily accessible in memory when the listener is familiar with the speaker, such as when a native speaker listens to another native speaker. In such cases, phonological predictions might

be implemented more efficiently compared to when the input comes from a less familiar source, such as a foreign-accented speaker. Crucially, Kuperberg & Jaeger (2016) proposed that the degree and level of predictive pre-activation depend on its expected utility, which in turn is shaped by the comprehenders' goals and their estimates of the relative reliability of prior knowledge and bottom-up input. Consequently, in contexts where anticipating the phonological form of upcoming words can enhance task performance, as in distinguishing mispronounced real words from non-words, comprehenders may more flexibly draw on their prior knowledge to implement phonological predictions.

The notion of prediction flexibility is also central to the parallel prediction-by-production model proposed by Pickering & Strijkers (2024). Traditionally, prediction-by-production models assumed that production representations are accessed sequentially (Pickering & Gambi, 2018; Pickering & Garrod, 2013). As a result, predictions were thought to proceed from higher to lower levels of representation, similarly to the hierarchical architecture proposed by Kuperberg & Jaeger (2016). By drawing on empirical evidence of parallel activation in language production, Pickering & Strijkers (2024) proposed a prediction-by-production model in which predictions can occur simultaneously at different levels of representation. The authors suggested that word learning involves constructing a single representation that combines meaning, grammar and phonology. This allows for the simultaneous retrieval of different linguistic components in language production and, consequently, during prediction in language comprehension. The concurrent availability of different linguistic components enables comprehenders to adaptively prioritize the prediction of information that is most relevant to the current task or communicative context. Within a parallel prediction-by-production framework, our results can be explained by assuming that comprehenders generate stronger parallel predictions for native-accented speech, as their phonological representations for such input are more robust due to greater familiarity. Conversely, parallel predictions might be reduced for foreign-accented speech, where phonological representations are less well established. Crucially, a parallel prediction-by-production system suggests that when the experimental task emphasizes sounds, as in Study 1, phonological predictions may be heightened. Therefore, task demands may enhance phonological predictions even when memory representations are weaker, such as with foreign-accented speech.

To conclude, the results of Studies 1 and 2 suggest that comprehenders track speech variability across speakers and exploit the flexibility of the perceptual system to implement specific phonological predictions. Our findings provide insights into both the levels of

representation and cognitive mechanisms involved in prediction during language comprehension. They support the notion that language comprehension involves the pre-activation of phonological information and suggest that predictions do not remain confined to an abstract representation that neglects phonological variability across speakers with a different accent. Moreover, our results indicate that prediction during language comprehension relies on mechanisms that extend beyond the passive spreading of activation between long-term stored representations. Linguistic prediction seems to involve flexible and finely tuned processes, which can account for inter-individual differences in word realization by targeting specific phonological forms.

4.7. Limitations and future directions

The results from Studies 1 and 2 support the conclusion that comprehenders can generate speaker-specific phonological predictions. Although our findings align with current models of language comprehension, further work is needed to clarify the cognitive mechanisms that support prediction across different representational levels. Specifically, paradigms that explicitly manipulate the availability of the production system (see Martin et al., 2018) are needed to assess the extent to which prediction-by-production models can account for our findings. In addition, while the present study showed evidence of phonological prediction following target word onset, examining neural activity during the interval preceding target onset could provide further insights into the temporal dynamics of predictive processing. Previous research has shown that highly constraining contexts elicit a sustained pre-target negativity relative to weakly constraining contexts (Grisoni et al., 2017; León-Cabrera et al., 2019). Investigating whether a similar effect is modulated by speaker accent or face cues would elucidate how phonological prediction operate before lexical information becomes available. Moreover, the prediction task employed in Study 2 may limit the generalizability of the results to more ecologically valid settings. Both studies also presented sentence frames that were either highly or weakly constraining toward a target word. Future work should investigate the relationship between prediction and speaker variability in more naturalistic contexts, where comprehension involves integrating multiple utterances and word predictability varies continuously.

Chapter 5 – Study 3: Using Temporal Response Function to Investigate Perceptual Adaptation to Speech and Prediction under Naturalistic Listening Conditions

5.1. Introduction

In the previous chapters, I presented two studies investigating whether comprehenders leverage contextual information to anticipate the specific phonological form of a predictable word. Yet, prediction is considered a crucial mechanism that operates across multiple stages of language processing, from the perception of phonemic categories to word recognition and sentence- or discourse-level comprehension. Traditionally, psycholinguistic research has sought to examine the processes involved at distinct levels of language processing separately; consequently, how different levels of processing interact during spoken language comprehension remains largely unexplored. Additionally, research on prediction in language comprehension often contrasts high versus low cloze probability words, even though real-time language comprehension involves processing words whose predictability varies continuously. In the following sections, I will present prediction as a probabilistic process unfolding across different stages of language processing and consider how it may be shaped by speaker variability. I will also examine how using the Temporal Response Function (TRF) to analyze electrophysiological data can provide insights into language processing under naturalistic listening conditions. Finally, I will introduce a study aimed at investigating how listeners cope with speaker variability in native speech, and whether this, in turn, modulates prediction during language comprehension.

5.1.1. Probabilistic prediction at different stages of language processing

A common feature of current language processing models is that prediction operates probabilistically: the brain continuously generates graded expectations about upcoming input to deal with uncertainty in both the speech signal and the linguistic content (e.g., Kleinschmidt & Jaeger, 2015; Kuperberg & Jaeger, 2016; Norris & McQueen, 2008). As described in the Introduction, recent models of speech perception propose that exposure to a given speaker enables listeners to build distributional models of acoustic cues, allowing them to anticipate how phonemic categories are likely to be realized (Kleinschmidt & Jaeger, 2015; Weatherholtz & Jaeger, 2016). This aligns with empirical evidence showing that listeners' expectations about a speaker's social and linguistic background influence how speech is perceived (Drager, 2011; Hay et al., 2006; Staum Casasanto, 2008; Walker & Hay, 2011). Although previous research

has largely addressed sociolinguistic variation, it remains unclear to what extent the speech perception system is able to track and adapt to acoustic differences across individuals with similar social and linguistic backgrounds (e.g., native male speakers of the same age group). Some studies have shown that individual words are recognized more easily when spoken by a familiar speaker than by an unfamiliar one (e.g., Nygaard et al., 1994), suggesting that the system learns speaker-specific properties to optimally process the variations present in the environment. Whether this facilitation extends to naturalistic speech processing, where contextual information can be used to facilitate comprehension, remains to be determined.

Word recognition is also guided by probabilistic predictions: as speech unfolds, the brain (pre)activates lexical candidates that are compatible with the input presented so far (Huettig et al., 2022; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986; Norris & McQueen, 2008). For example, upon hearing the phoneme sequence /kæp-/, listeners may anticipate words such as *captain* or *capital*, with the level of activation influenced by each word's contextual likelihood. As additional information becomes available, the set of lexical candidates is dynamically updated, with certain candidates gaining strength and others being ruled out, facilitating word recognition. The graded activation of lexical candidates during word recognition is supported by studies showing that words sharing their first phonemes with many other words (i.e., from large cohorts) are usually associated with slower response times in lexical decision tasks as compared to words drawn from smaller cohorts (e.g., Gareth Gaskell & Marslen-Wilson, 2002; Marslen-Wilson, 1987; Tyler et al., 2000). Finally, a large body of empirical evidence suggests that listeners continuously integrate the meaning of each word into an unfolding contextual representation, which in turn supports the generation of probabilistic predictions about upcoming words, even before bottom-up input from those words is encountered (Kuperberg & Jaeger, 2016). Previous work investigating how speaker variability influences prediction during language comprehension has primarily contrasted conditions in which listeners are exposed to native and foreign-accented speech, which often involves non-canonical pronunciations (Best et al., 2001; Clopper et al., 2005; Flege, 1988) as well as unusual prosody patterns (Gut, 2012). However, whether speaker variability in native speech affects prediction in language comprehension remains an open question.

5.1.2. Temporal Response Function and language processing

In Studies 1 and 2, we used a lexical decision task and Event-Related Potentials (ERPs) to examine whether comprehenders anticipate the specific phonological form of a predictable

word. However, this approach to studying prediction in language comprehension has inherent limitations. For instance, ERPs typically are used to measure brain responses to a word presented at a fixed position within a sentence, whereas language comprehension is a continuous process. Moreover, psycholinguistic studies usually compare words with high versus low cloze probability, even though word predictability varies continuously throughout a sentence or discourse. The Temporal Response Function (TRF) offers a way to overcome these limitations by modeling the relationship between a continuous stimulus feature and the corresponding neural activity over time (Brodbeck et al., 2018, 2022; Crosse et al., 2016; Di Liberto et al., 2015; Heilbron et al., 2022). This method relies on regularized regression to predict the M/EEG signal from a continuous stimulus feature (*forward modeling*), producing linear regression coefficients, or TRF weights, that can be interpreted in terms of stimulus-elicited brain activity[2]. Unlike ERPs, which require the repetition of experimental stimuli, TRF analysis enables the study of neural responses to naturalistic time-varying stimuli, such as continuous speech and music. For instance, modeling the speech envelope, a low-frequency signal reflecting amplitude fluctuations in speech over time, can reveal how the brain tracks low-level acoustic information (Kalashnikova et al., 2018; Kurthen et al., 2021; Lalor & Foxe, 2010). Given that complex stimuli such as speech typically include several dimensions, the multivariate form of TRF (*m*TRF) enables the simultaneous modeling of neural responses to multiple stimulus features (Crosse et al., 2016). Research using this approach showed that the human brain encodes speech at multiple levels, from low-level spectrotemporal information to phonetic articulatory features and phonemes (Di Liberto et al., 2015). Recent applications of *m*TRF included higher-order linguistic predictors, such as cohort entropy and word surprisal, to investigate real-time predictive processing during language comprehension (Brodbeck et al., 2018, 2022; Gillis et al., 2021; Gwilliams et al., 2022; Heilbron et al., 2022). Cohort entropy quantifies the degree of competition among lexical candidates that are compatible with the input from word onset to the current phoneme, capturing the extent to which the brain is uncertain about the upcoming phonemes of the current word. The TRF-modeled neural response (weights) of cohort entropy exhibits fronto-central activity approximately 100–300 ms after word onset (Gillis et al., 2021). Word surprisal quantifies the degree of unexpectedness associated with a word given its preceding context, reflecting prediction during sentence or discourse processing. This measure is typically estimated using computational models such as

[2] In addition to modeling neural responses from stimulus features, Temporal Response Function can also be applied to reconstruct stimulus features from neural activity (*backward modeling*). This approach will not be discussed further here; for more details, see Crosse et al. (2016).

large language models (e.g., GPT) (Gillis et al., 2021; Heilbron et al., 2022; Michaelov et al., 2024). The TRF-modeled neural response to word surprisal shows a prominent centro-parietal negativity, resembling the classic N400 component associated with prediction error (Gillis et al., 2021; Mesik & Wojtczak, 2023; Weissbart et al., 2020). This line of research has extended previous behavioral and electrophysiological evidence of probabilistic pre-activation of linguistic information, while also showing that the brain generates predictions at multiple levels of abstraction under naturalistic listening conditions (Heilbron et al., 2022).

5.1.3. The present study

As previously discussed, speech perception and prediction in language comprehension are probabilistic processes. However, the extent to which listeners track speaker-specific information in native speech, and whether speech variability among native speakers influences prediction in language comprehension, remains largely unknown. In this chapter, I will present a study aimed at addressing this research question using the Temporal Response Function, an M/EEG analysis technique that allows the investigation of language processing under naturalistic listening conditions. A previous version of this work was published as a preprint in *bioRxiv* (Piazza, Sala et al., 2025)[3]. However, we have recently identified an error in the forced alignment of the speech materials. Since data analysis is still ongoing, I will present the methodology and expected results of the study.

EEG data were recorded from native Italian speakers as they listened to continuous naturalistic speech. Participants listened to two narrative stories produced by male speakers in their early adulthood, all of whom were native speakers of Italian. The stories were presented under two experimental conditions: in the *Single-speaker* condition, the entire story was narrated by one speaker, enabling listeners to develop familiarity with the voice; in the *Multi-speaker* condition, distinct sections of the story were narrated by nine different speakers (one per section), introducing significant speech variability and potentially limiting voice-specific adaptation. Temporal Response Function will be used to capture neural responses to stimulus features related to speech perception and prediction in language comprehension. Specifically, we plan to model neural responses to phonemes, as a probe of speech perception; to cohort entropy, which reflects the degree of competition among lexical candidates given a sequence of phonemes and thus informs prediction in word recognition; and to word surprisal, which captures the unexpectedness of a word given its preceding context and thereby reflects

[3] Giorgio Piazza & Marco Sala equally contributed to this work as first authors.

prediction during sentence or discourse comprehension. Our hypothesis is that, in the Multi condition, listeners may need to accommodate differences in the production of speech sounds across speakers, leading to increased cognitive demands in the perception of phonemes compared to the Single condition. Greater uncertainty in perceiving phonemes may increase competition among lexical candidates during word recognition and reduce the strength of predictions about upcoming words.

5.2. Methods

5.2.1. Participants

Thirty-four adults (21 females; M_{age} = 24.7 ± 3.84 y.o.) were initially recruited for the study. Four participants were excluded due to technical problems, leaving the final cohort to 30 (20 females; M_{age} = 24.1 ± 3.52 y.o.). All participants were native Italian speakers with no reported history of neurological, psychiatric, or language-related disorders. They were recruited from healthy volunteers and students at the University of Padova, and received 25 euros for their participation. All participants provided their informed consent before participating in the experiment. The research adhered to the principles outlined in the Declaration of Helsinki. Permission to conduct the study was given by the Ethics Committee for Psychological Research of the University of Padova.

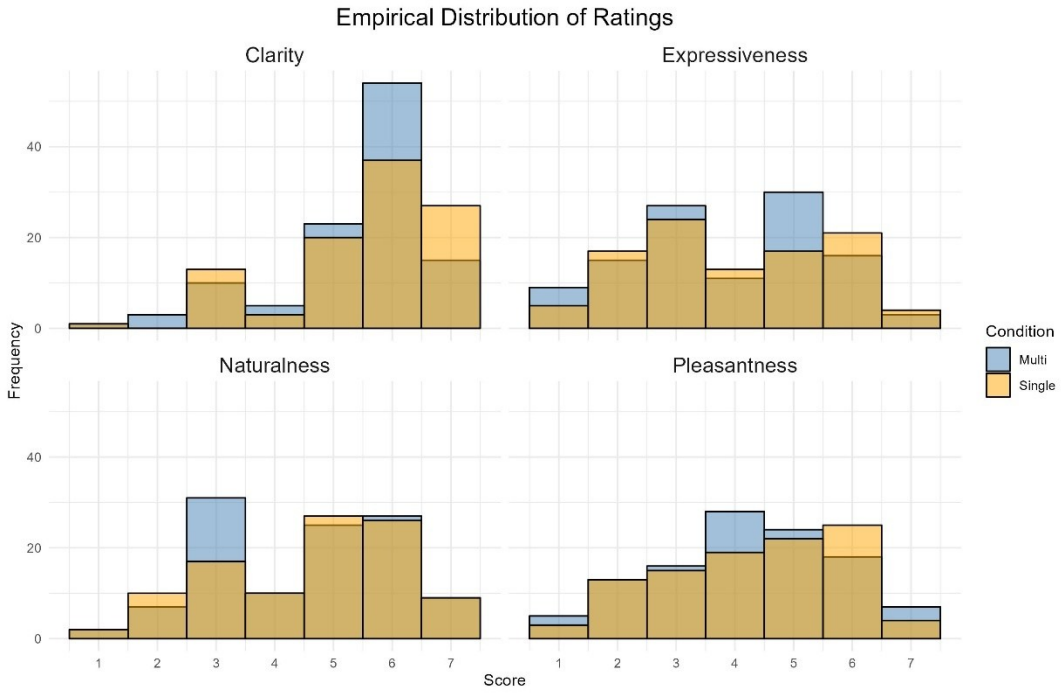
5.2.2. Materials

As part of this study, we used continuous recordings of Italian speech. Two narrative stories, describing trips to Ethiopia and North Korea, were pre-recorded by male speakers in their early adulthood, all being native speakers of Italian. Each story consisted of nine sections and was recorded under two conditions: in the *Single-speaker* condition (hence Single), the story was narrated by one speaker throughout; in the *Multi-speaker* condition (hence Multi), different sections of the story were narrated by distinct speakers (9 speakers in total). In the Multi condition, the same set of speakers was used across both stories, whereas in the Single condition, different speakers narrated the two stories to avoid making our results specific to any individual speaker. Since speakers were asked to speak naturally, the speech rate - and therefore the duration of the sections - varied across conditions. Given prior evidence showing that speech rate influences the neural encoding of speech (Piazza, et al., 2025; Verschueren et al., 2022), we computed the average duration of each section pair (e.g., Section 1 of the North Korea story in both the Single and Multi conditions) and applied minimal time-stretching or shortening (~5%) to balance the durations, while preserving pitch. As a result, the duration of each story

recording was approximately 22 minutes. Each story was normalized to an average level of 68 dB.

We conducted a preliminary online study on Prolific (www.prolific.com) to assess whether the audio recordings differed between conditions in speech features that could affect processing (without being the focus of our manipulation). 212 native Italian speakers listened to a ~2 minute story section and provided ratings on four dimensions: Clarity, Expressiveness, Naturalness, and Pleasantness, using a 7-point Likert scale (1 = 'Not at all'; 7 = 'Completely/Extremely'). Figure 5.1 shows the empirical distribution of ratings across the evaluated dimensions in both experimental conditions.

Figure 5.1. Empirical distribution of participants' ratings on a 7-point Likert scale (1 = 'Not at all', 7 = 'Completely/Extremely') across four dimensions: Clarity, Expressiveness, Naturalness, and Pleasantness. Ratings are shown separately for the two experimental conditions, "Multi" (blue) and "Single" (orange). Each panel displays the frequency of scores within each dimension.



We used Cumulative Link Mixed Models (CLMM) to compare the two experimental conditions ("Multi" vs. "Single") across the evaluated dimensions (Clarity, Expressiveness, Naturalness, and Pleasantness). CLMM are statistical models designed for analyzing ordinal data, where the variables have ordered categories and the distances between the categories are not known

(Christensen, 2023). These models estimate the probability that a response falls into a particular category or below, taking the ordinal nature of the data into account. This approach allowed us to assess whether the experimental conditions significantly differed in the distribution of ratings. Results showed null effects for the evaluated dimensions: Clarity (b = 0.396, z-value = 0.734, p = .463), Expressiveness (b = 0.108, z-value = 0.242, p = .809), Naturalness (b = 0.151, z-value = 0.486, p = .627), Pleasantness (b = 0.138, z-value = 0.253, p = .800).

5.2.3. Procedure and design

Participants were seated comfortably in a soundproof room, equipped with a computer setup that included an LCD monitor, external speakers, and a keyboard. PsychoPy software was used to present the stimuli (Peirce et al., 2019). Stimuli were presented at a comfortable volume through loudspeakers. Participants listened to continuous audio recordings while their EEG activity was recorded. They were instructed to sit upright while looking at a fixation cross displayed at the center of a computer screen (at ~ 80 cm from their eyes). Participants listened to two different stories, one for each experimental condition (Single and Multi), with the order of presentation counterbalanced across participants. Each story was assigned to both conditions across participants to ensure that condition effects were not due to differences between stories (e.g., one certain story is more interesting). Both stories were segmented into nine successive blocks, each lasting approximately 2 minutes and 30 seconds. After each block, participants answered five true/false comprehension questions, resulting in a total of 45 questions per story. The entire session lasted approximately two hours, including EEG setup and the experimental task.

5.2.4. EEG recording and pre-processing

EEG data were recorded using a 64-channel active Ag/AgCl electrode system (Brain Products, ActiCap), following the international 10–20 convention. Sixty-three electrodes were used as active recording sites (Fp1, Fp2, AF3, AF4, AF7, AF8, AFz, F1, F2, F3, F4, F5, F6, F7, F8, Fz, FT7, FT8, FT9, FT10, FC1, FC2, FC3, FC4, FC5, FC6, FCz, T7, T8, C1, C2, C3, C4, C5, C6, Cz, TP7, TP8, CP1, CP2, CP3, CP4, CP5, CP6, CPz, P1, P2, P3, P4, P5, P6, P7, P8, Pz, PO3, PO4, PO7, PO8, Poz, O1, O2, Oz, Iz), while one electrode was positioned on the right mastoid. The configuration was considered adequate only if electrode impedance was below 15 kΩ at the end of electrode placement. Signals were digitized at a sampling rate of 1000 Hz and online referenced to the left mastoid. We examined participants' responses to comprehension questions to ensure they paid attention to the stimuli during the experiment. All participants scored above

chance level, with an average accuracy of approximately 80% (range: 67% - 93%). As a result, no participants were excluded. Behavioral responses were not analyzed further.

EEG preprocessing was conducted in MATLAB (The MathWorks Inc., 2024), using Fieldtrip toolbox functions (Oostenveld et al., 2011) and EEGLAB (Delorme & Makeig, 2004). EEG data were offline re-referenced to the average of the left and right mastoids and downsampled to 100 Hz. A band-pass Butterworth filter (0.5–8 Hz, zero-phase, order 2+2) was applied to retain low-frequency activity relevant for TRF analysis. Noisy channels, defined as having variance three times above or below the median channel variance, were recalculated by spline interpolation using neighboring clean channels. Due to recording failures or excessive noise, one story section was excluded for ten participants, and two sections were excluded for two participants.

5.3. Planned analyses

5.3.1. Temporal Response Function modeling

We plan to assess the cortical encoding of linguistic features using Temporal Response Function (TRF). This method typically assumes that the neural response at a given time is a linear transformation of a specific stimulus feature. The TRF describes this transformation for a specified range of time lags relative to the instantaneous occurrence of the stimulus feature. In practical terms, when recording from multiple channels, the signal measured at a given channel and time is modeled as a combination of the current stimulus feature at different time lags, each weighted differently. The TRF weights reveal how the stimulus feature contributes to the observed EEG signal and can be interpreted both temporally (stimulus–EEG latencies) and spatially (scalp topographies), similarly to ERPs. Larger TRF weights, whether positive or negative, indicate that a particular feature at that time point plays a significant role in predicting the EEG signal. Because natural stimuli such as speech often exhibit strong temporal correlations, the stimulus-response relationship is usually estimated using regularized linear regression. This method applies a regularization parameter (λ) that stabilizes the solution and controls overfitting by assuming a certain level of temporal smoothness (Crosse et al., 2021). Since the choice of the regularization parameter λ is critical for obtaining reliable TRF estimates, it is often selected using cross-validation. For instance, Crosse et al. (2016) suggest using a “leave-one-out” approach, in which the stimulus and EEG time series are divided into multiple folds. For each candidate λ value, the model is trained on all folds except one, which is used for testing. This process is repeated until each fold has served once as the test set. Prediction accuracy is usually quantified for each electrode using a Pearson correlation (*r*-

value) between the predicted and observed EEG signal. An r-value of 1 indicates perfect correspondence between the signal predicted by the TRF model and the observed signal, while an r-value of 0 indicates no correspondence at all. Although prediction correlation values are generally low, typically around ~0.05 or ~0.1, this is expected given the high level of independent noise in EEG signals (Brodbeck et al., 2018; Di Liberto et al., 2021; Di Liberto et al., 2015). The optimal λ is determined by identifying the value that maximizes average prediction accuracy across all test segments and channels. The performance of the optimized model can then be tested either on held-out data, which is considered best practice, or on the cross-validation data, as both approaches provide unbiased estimates of prediction accuracy (see Crosse et al., 2016).

Different toolboxes allow the estimation of TRF models with multiple predictors (*multivariate* Temporal Response Function, or *m*TRF). In our study, we plan to model the neural responses for both the Single and Multi conditions using the *m*TRF-Toolbox in MATLAB (Crosse et al., 2016; see also Brodbeck et al., 2023, for a Python alternative). For each condition, the stimulus and EEG time series will be divided into nine folds (one per story section), and a leave-one-out procedure will be applied to optimize model hyperparameters and assess the model's predictive performance. The *m*TRF models will be estimated within a [-200, 600] ms time-lag window, as previous empirical evidence shows that this interval includes the EEG responses to speech of interest (Brodbeck et al., 2018; Chalehchaleh et al., 2025; Piazza et al., 2025). Since our models will include both sparse (phonetic features, cohort entropy, and word surprisal) and non-sparse (acoustic spectrogram) predictors, we will employ a banded regression approach. This method enables separate λ selection for each feature group, ensuring optimal regularization across feature types (Carta et al., 2024).

5.3.2. Temporal Response Function regressors

The *m*TRF models will include the following regressors: acoustic spectrogram, phonetic features, cohort entropy, and word surprisal. This will allow us to examine speech perception, as reflected in neural responses to phonemes, and prediction in language comprehension, as indexed by neural responses to cohort entropy and word surprisal, while controlling for acoustic differences across conditions (Chalehchaleh et al., 2025; Di Liberto et al., 2021; Gillis et al., 2021; Piazza et al., 2025). Prior to model fitting, both EEG data and (non-zero) stimulus vectors will be normalized, following CNSP and mTRF-Toolbox recommendations (Crosse et al., 2016; Di Liberto et al., 2024). The regressors are defined as follows:

- *Acoustic spectrogram:* The spectrogram regressor consists of 8 frequency bands

between 250 Hz and 8000 Hz, defined using the Greenwood equation, which describes the relationship between the anatomic location of the inner ear hair cells and the frequencies at which they are stimulated (Greenwood, 1961, 1990). The acoustic spectrogram shows how acoustic energy varies over time across different frequency bands, allowing for modeling differences in the spectrotemporal profile of a given phoneme across instances (Di Liberto et al., 2015; Piazza et al., 2025).

- *Phonetic features (and phonemes):* Forced alignment of the speech materials will be obtained using Montreal Forced Aligner (McAuliffe et al., 2017) and then verified manually with PRAAT (Boersma & van Heuven, 2001). Phonemic alignments will be stored as a 17-dimensional binary time-series, where each dimension corresponds to a distinct phonetic feature. This procedure allows us to describe each phoneme as a unique combination of articulatory features. Phonetic features specify whether a speech sound is voiced, voiceless (consonants), plosive, fricative, affricate, nasal, approximant, front, back, central, close, open (vowels), bilabial, labiodental, dento-alveolar, palatal, velar-glottal (Ladefoged, 2006). Phonetic features will be used to fit the *m*TRF model, and the resulting TRF weights will be projected onto phonemes using a linear transformation matrix that maps combinations of phonetic features to the corresponding phonemic categories (Di Liberto et al., 2021; Piazza et al., 2025).
- *Cohort entropy:* Cohort theory posits that lexical items compete for activation from incomplete input (Marslen-Wilson, 1987). Cohort entropy quantifies the degree of competition among lexical candidates that are compatible with the partial phoneme string from word onset to the current phoneme. It is expressed as the Shannon entropy of the active cohort of words and reflects the extent to which the brain is uncertain about the upcoming phonemes of the current word (Brodbeck et al., 2018). Cohort entropy at a given phoneme *i* can be defined as:

$$H_i = -\sum_{word}^{cohort} p_{word}\, log2\, p_{word}$$

where p_{word} represents the probability of each word within the cohort *C*, commonly approximated using corpus frequency data. In our study, we plan to extract word frequencies from the ItWac corpus (Baroni et al., 2009) and phonological forms from the WikiPronunciationDict, a multilingual pronunciation dictionary covering several languages, including Italian. A probabilistic tree structure will be constructed to estimate the likelihood of various phoneme sequences and compute entropy values at each

phoneme position within words (Brodbeck et al., 2018; Gillis et al., 2021; Gwilliams et al., 2022). Entropy values will be encoded in a sparse time vector, where non-zero values indicate phoneme onsets until entropy reaches zero at the word's uniqueness point.

- *Word surprisal:* This metric quantifies how unexpected a word is in a given context and is computed as the negative logarithm of the current word's probability. Surprisal values will be estimated using the Italian transformer-based large language model UmBERTo. This model is based on BERT and was trained on a large-scale Italian corpus (CommonCrawl ITA), using 110M parameters and a vocabulary size of 32k (~69 GB). For each section (chunk) of the story, UmBERTo will be used to compute the probability of each word based on the preceding context. Surprisal values will be mapped into a sparse time vector, with non-zero values indicating word onsets and the degree of surprisal for that word given its context.

5.3.3. Statistical analyses

Prior to testing whether speaker variability modulates the TRF weights of phonemes, cohort entropy, and word surprisal, we will conduct control analyses to confirm that linguistic features are reliably encoded in the EEG signal. To assess their unique contribution, we will systematically exclude each predictor from the full model and measure the resulting change in prediction accuracy (also called *prediction gain*) (Di Liberto et al., 2018; Di Liberto et al., 2015; Piazza, et al., 2025). In both conditions, we plan to use one-sample t-tests (one-sided) to test whether the prediction gain for each linguistic feature is greater than zero.

To investigate whether the TRF weights of the features of interest vary across conditions, we plan to use cluster-based permutation testing (CBPT). CBPT is a non-parametric statistical test that controls for multiple comparisons across electrodes and time points (Maris & Oostenveld, 2007). This data-driven approach allows us to avoid predefining regions of interest or time windows. In ERP research, there is extensive empirical evidence guiding the selection of a priori time windows and electrode clusters for testing specific components. For instance, N400 effects are typically examined in the 300–500 ms time-window after word onset in a cluster of centro-parietal electrodes (Kutas & Federmeier, 2011). In contrast, the TRF literature is still relatively limited, making a data-driven approach such as CBPT particularly appropriate.

5.4. Expected results

We expect to observe larger neural responses to phonemes in the Multi than in the Single condition, as listeners may need to accommodate differences in the production of speech sounds across speakers, leading to increased processing demands. Greater uncertainty in perceiving phonemes may increase competition among lexical candidates during word recognition, leading to larger neural responses to cohort entropy in the Multi than in the Single condition. Finally, speaker variability may limit the cognitive resources available and therefore reduce the strength of predictions about upcoming words, leading to smaller neural responses to word surprisal in the Multi than the Single condition.

The results of this study could provide relevant evidence in favor of the flexible nature of the perceptual system, showing that speaker-specific features are tracked and used to optimize speech perception. In addition, clarifying how speaker variability affects the generation of predictions in language comprehension may demonstrate that the extent to which linguistic representations are pre-activated depends on the estimated reliability of prior knowledge and sensory input. These findings would align closely with current models of language comprehension (Kuperberg & Jaeger, 2016; Nour Eddine et al., 2024) and with theories of human adaptive behavior (Clark, 2013; Doya et al., 2007; Friston, 2005), which propose that humans build probabilistic models based on prior experiences to facilitate information processing and improve adaptive behavior.

Although Study 3 may contribute to clarifying the processes involved in speech perception and prediction, some considerations should be kept in mind. The Single and Multi-speaker conditions may differ not only in terms of speech variability but also along other dimensions, such as cognitive demands and attentional engagement. To better isolate the effect of perceptual adaptation on speech processing, future analyses may examine how neural responses to linguistic features of interest evolve over time. Moreover, our paradigm may provide evidence that speaker-specific information is at least temporarily retained during naturalistic speech processing. However, assessing the extent to which this information is stored in long-term memory would require an experimental design in which participants are exposed to the speakers to varying degrees and then tested after a temporal delay. Additionally, although we account for acoustic differences between experimental conditions by modeling neural responses to the acoustic spectrogram, a more rigorous design would involve systematically rotating speakers across conditions to control for potential speaker-specific acoustic effects. Finally, while our paradigm aims to approximate naturalistic speech processing, future research should explore how speaker variability influences prediction in conversational contexts, where

the need to anticipate what a speaker will say may be even greater.

Conclusions

The studies presented in this thesis aimed to investigate whether and how speaker variability influences prediction in language comprehension. In Studies 1 and 2, we addressed a key question: is the human brain able to anticipate the specific phonological form of a predictable word when it is spoken by a native versus a foreign-accented speaker? We designed an experimental paradigm that controlled sentence-context processing demands across accent conditions, while using face stimuli to cue the speaker's identity and, consequently, their phonological features. Behavioral and ERP results consistently showed that comprehenders draw on both the linguistic and non-linguistic context to implement speaker-specific phonological predictions. These findings are relevant to current theories of language processing as they provide insights into both the content and the mechanisms of prediction during language comprehension. Specifically, they show that, at least in highly constraining and informative contexts, prediction targets specific phonological forms. Moreover, they suggest that prediction during sentence processing relies on flexible and finely tuned processes that extend beyond the passive spreading of activation between long-term stored representations. Although the results of Studies 1 and 2 align with current models of language comprehension, further research is needed to clarify the cognitive mechanisms that underlie the generation of phonological predictions. Additionally, both studies contrasted sentence frames that were either highly or weakly constraining toward a target word, even though naturalistic language processing involves integrating multiple utterances in which word predictability varies continuously.

Building on this limitation, in Study 3, we plan to employ the Temporal Response Function, an M/EEG analysis technique that models the relationship between continuous stimulus features and neural responses. In this study, we manipulated participants' exposure to native speakers by presenting narrative stories produced either by a single speaker or by multiple speakers. By modeling neural responses to linguistic features involved in speech perception and prediction, we aim to reveal whether listeners track speaker-specific information in naturalistic speech processing, and to what extent speaker variability in native speech affects the generation of predictions in language comprehension.

Overall, the present work underscores the adaptability of the human brain in processing linguistic information. Our findings suggest that the brain optimizes language processing by generating detailed predictions about upcoming input and modulating their strength according to how likely they are to be accurate. Whether predictive mechanisms in language comprehension are influenced by variability among native speakers remains an open question,

which we aim to address in future work. The findings presented here, together with future results, aim to contribute to the development of a unified framework that captures the mutual influences between processes involved in speech perception and language comprehension.

References

AbdulSabur, N. Y., Xu, Y., Liu, S., Chow, H. M., Baxter, M., Carson, J., & Braun, A. R. (2014). Neural correlates and network connectivity underlying narrative production and comprehension: A combined fMRI and PET study. *Cortex*, *57*, 107–127. https://doi.org/10.1016/j.cortex.2014.01.017

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520–529. https://doi.org/10.1037/a0013552

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation Improves Language Comprehension. *Psychological Science*, *21*(12), 1903–1909. https://doi.org/10.1177/0956797610389192

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Alarcos, E. (1953). Sistema fonemático del catalán. In *Estudis de lingüística catalana* (pp. 11–32). Ariel.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and Prediction in Human Sentence Processing. *Cognitive Science*, *33*(4), 583–609. https://doi.org/10.1111/j.1551-6709.2009.01022.x

Anderson, J. R. (1983). *The Architecture of Cognition*. Psychology Press. https://doi.org/10.4324/9781315799438

Angulo-Chavira, A. Q., Castellón-Flores, M. L. S., López-Santillán, H., & Arias-Trejo. N. (2023). Phono-semantic prediction during language comprehension: Effects of working memory. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560. https://doi.org/10.1080/01690965.2012.658072

Ardila, A. (2010). A Review of Conduction Aphasia. *Current Neurology and Neuroscience Reports*, *10*(6), 499–503. https://doi.org/10.1007/s11910-010-0142-2

Arias-Trejo, N., & Plunkett, K. (2009). Lexical–semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3633–3647. https://doi.org/10.1098/rstb.2009.0146

Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, *128*(2), 214–227. https://doi.org/10.1016/j.cognition.2013.03.008

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baggio, G. (2012). Selective alignment of brain responses by task demands during semantic processing. *Neuropsychologia*, *50*(5), 655–665. https://doi.org/10.1016/j.neuropsychologia.2012.01.002

Baggio, G. (2018). *Meaning in the Brain*. MIT Press.

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338–1367. https://doi.org/10.1080/01690965.2010.542671

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 336–345. https://doi.org/10.1037/0278-7393.12.3.336

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics*, *44*(5), 395–408. https://doi.org/10.3758/BF03210424

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Languages Resources and Evaluation*, *43*(3), 209–226.

Barreda, S. (2012). Vowel normalization and the perception of speaker changes: An exploration of the contextual tuning hypothesis. *The Journal of the Acoustical Society of America*, *132*(5), 3453–3464. https://doi.org/10.1121/1.4747011

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology*, *106*, 50–64. https://doi.org/10.1016/j.ijpsycho.2016.06.003

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Krivitsky, P. (2015). Package "lme4." *Convergence*, *12*(1), Article 2.

Beauducel, A., & Hilger, N. (2018). On Optimal Allocation of Treatment/Condition Variance in Principal Component Analysis. *International Journal of Statistics and Probability*, *7*(4), 50. https://doi.org/10.5539/ijsp.v7n4p50

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition*, *8*(6), 493–512. https://doi.org/10.3758/BF03213769

Becker, C. A. (1985). What do we really know about semantic context effects during reading? In D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 125–166). Academic Press.

Berndt, R. S., & Caramazza, A. (1980). A redefinition of the syndrome of Broca's aphasia: Implications for a neuropsychological model of language. *Applied Psycholinguistics*, *1*(3), 225–278. https://doi.org/10.1017/S0142716400000552

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171–204). New York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. https://doi.org/10.1121/1.1332378

Blumstein, S. A. (1973). *A phonological investigation of aphasic speech*. Walter de Gruyter GmbH & Co KG.

Boersma, P., & van Heuven, V. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5(9/10)*, 341–345.

Bölte, J., & Coenen, E. (2002). Is Phonological Information Mapped onto Semantic Information in a One-to-One Manner? *Brain and Language*, *81*(1–3), 384–397. https://doi.org/10.1006/brln.2001.2532

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni Del R Istituto Superiore Di Scienze Economiche e Commericiali Di Firenze*, *8*, 3–62.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, *2*(1). https://doi.org/10.16910/jemr.2.1.1

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6). https://doi.org/10.1111/psyp.13049

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, *125*, 104334. https://doi.org/10.1016/j.jml.2022.104334

Brodbeck, C., Bhattasali, S., Cruz Heredia, A. AL, Resnik, P., Simon, J. Z., & Lau, E. (2022). Parallel processing in speech perception with local and global representations of linguistic context. *ELife*, *11*, e72056. https://doi.org/10.7554/eLife.72056

Brodbeck, C., Das, P., Gillis, M., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2023). Eelbrain, a Python toolkit for time-continuous analysis with temporal response functions. *ELife*, *12*. https://doi.org/10.7554/eLife.85012

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. *Current Biology*, *28*(24), 3976-3983.e5. https://doi.org/10.1016/j.cub.2018.10.042

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory

processes. *Neuropsychologia*, *135*, 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149. https://doi.org/10.1016/j.cognition.2014.10.017

Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, *125*(1), 82–93. https://doi.org/10.1016/j.bandl.2013.01.007

Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*, Article 25. https://doi.org/10.3389/fnhum.2010.00025

Carta, S., Aličković, E., Zaar, J., Valdés, A. L., & Di Liberto, G. M. (2024). Cortical encoding of phonetic onsets of both attended and ignored speech in hearing impaired individuals. *PLOS ONE*, *19*(11), e0308554. https://doi.org/10.1371/journal.pone.0308554

Chalehchaleh, A., Winchester, M. M., & Di Liberto, G. M. (2025). Robust assessment of the cortical encoding of word-level expectations using the temporal response function. *Journal of Neural Engineering*, *22*(1), 016004. https://doi.org/10.1088/1741-2552/ada30a

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing Referential Domains during Real-Time Language Comprehension. *Journal of Memory and Language*, *47*(1), 30–49. https://doi.org/10.1006/jmla.2001.2832

Chiba, T., & Kajiyama, M. (1941). *The Vowel: Its Nature and Structure*. Tokyo-Kaseikan Pub. Co.

Christensen, R. (2023). ordinal—Regression Models for Ordinal Data. *R Package Version 2023.12-4.1*.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521–533. https://doi.org/10.1037/0278-7393.20.3.521

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing Uncertainty*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Clopper, C. G., & Bradlow, A. R. (2008). Perception of Dialect Variation in Noise: Intelligibility and Classification. *Language and Speech*, *51*(3), 175–198. https://doi.org/10.1177/0023830908098539

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America*, *118*(3), 1661–1676. https://doi.org/10.1121/1.2000774

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, *81*(4), 348–374. https://doi.org/10.1037/h0036656

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constrainst. *Journal of Memory and Language*, *30*(2), 234–250. https://doi.org/10.1016/0749-596X(91)90005-5

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, *70*(3), 403–411. https://doi.org/10.3758/PP.70.3.403

Connolly, J. F., & Phillips, N. A. (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266. https://doi.org/10.1162/jocn.1994.6.3.256

Connolly, J. F., Service, E., D'Arcy, R. C. N., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Neuroreport*, *12*(2), 237–243. https://doi.org/10.1097/00001756-200102120-00012

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664. https://doi.org/10.1016/j.cognition.2007.03.013

Creel, S. C., & Bregman, M. R. (2011). How Talker Identity Relates to Language Processing. *Language and Linguistics Compass*, *5*(5), 190–204. https://doi.org/10.1111/j.1749-818X.2011.00276.x

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic Processing of Accented Speech Across the Lifespan. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00479

Crocker, M. W. (1999). Mechanisms for sentence processing. In S. Garrod & M. Pickering (Eds.), *Language processing* (1st Edition). Psychology Press.

Crocker, M. W. (2000). Wide-Coverage Probabilistic Sentence Processing. *Journal of Psycholinguistic Research*, *29*(6), 647–669. https://doi.org/10.1023/A:1026560822390

Croot, K., Ballard, K., Leyton, C. E., & Hodges, J. R. (2012). Apraxia of Speech and Phonological Errors in the Diagnosis of Nonfluent/Agrammatic and Logopenic Variants of Primary Progressive Aphasia. *Journal of Speech, Language, and Hearing Research*, *55*(5), S1562–S1572. https://doi.org/10.1044/1092-4388(2012/11-0323)

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*, Article 604. https://doi.org/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*, *15*, Article 705621. https://doi.org/10.3389/fnins.2021.705621

Cycowicz, Y. M., & Friedman, D. (1997). A developmental study of the effect of temporal order on the ERPs elicited by novel environmental sounds. *Electroencephalography and Clinical Neurophysiology*, *103*(2), 304–318. https://doi.org/10.1016/S0013-4694(97)96053-3

Dahan, D. (2010). The Time Course of Interpretation in Speech Comprehension. *Current Directions in Psychological Science*, *19*(2), 121–126. https://doi.org/10.1177/0963721410364726

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

de Ruiter, J.-P., Mitterer, H., & Enfield, N. J. (2006). Projecting the End of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language*, *82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394. https://doi.org/10.1098/rstb.2012.0394

DeLong, K. A. (2009). *Electrophysiological explorations of linguistic pre -activation and its consequences during online sentence processing*. University of California.

DeLong, K. A., Chan, W., & Kutas, M. (2019). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology*, *56*(4), e13312. https://doi.org/10.1111/psyp.13312

DeLong, K. A., Chan, W., & Kutas, M. (2021). Testing limits: ERP evidence for word form preactivation during speeded sentence reading. *Psychophysiology*, *58*(2), e13720. https://doi.org/10.1111/psyp.13720

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, Predictive Parsing with Psycholinguistically Motivated Tree-Adjoining Grammar. *Computational Linguistics*, *39*(4), 1025–1066. https://doi.org/10.1162/COLI_a_00160

Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Cortical Measures of Phoneme-Level Speech Encoding Correlate with the Perceived Clarity of Natural Speech. *Eneuro*, *5*(2), ENEURO.0084-18.2018. https://doi.org/10.1523/ENEURO.0084-18.2018

Di Liberto, G. M., Nidiffer, A., Crosse, M. J., Zuk, N., Haro, S., Cantisani, G., Winchester, M. M., Igoe, A., McCrann, R., Chandra, S., Lalor, E. C., & Baruzzo, G. (2024). A standardised open science framework for sharing and re-analysing neural data acquired to continuous stimuli. *Neurons, Behavior, Data Analysis, and Theory*. https://doi.org/10.51628/001c.124867

Di Liberto, G. M., Nie, J., Yeaton, J., Khalighinejad, B., Shamma, S. A., & Mesgarani, N. (2021). Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage*, *227*, 117586. https://doi.org/10.1016/j.neuroimage.2020.117586

Diaz, M. T., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, *1146*, 85–100. https://doi.org/10.1016/j.brainres.2006.07.034

Dien, J. (2012). Applying Principal Components Analysis to Event-Related Potentials: A Tutorial. *Developmental Neuropsychology*, *37*(6), 497–517. https://doi.org/10.1080/87565641.2012.697503

Dien, J., & Frishkoff, G. A. (2005). *Principal components analysis of event-related potential datasets. In: Handy, T. (Ed.), Event-related Potentials: A Methods Handbook. MIT Press, pp. 189–208.*

Dikker, S., & Pylkkanen, L. (2011). Before the N400: Effects of lexical–semantic violations in visual cortex. *Brain and Language*, *118*(1–2), 23–28. https://doi.org/10.1016/j.bandl.2011.02.006

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early Occipital Sensitivity to Syntactic Category Is Based on Form Typicality. *Psychological Science*, *21*(5), 629–634. https://doi.org/10.1177/0956797610367751

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition*, *110*(3), 293–321. https://doi.org/10.1016/j.cognition.2008.09.008

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ding, J., Zhang, Y., Liang, P., & Li, X. (2023). Modulation of working memory capacity on predictive processing during language comprehension. *Language, Cognition and Neuroscience*, *38*(8), 1133–1152. https://doi.org/10.1080/23273798.2023.2212819

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT Press.

Drager, K. (2011). Speaker Age and Vowel Perception. *Language and Speech*, *54*(1), 99–121. https://doi.org/10.1177/0023830910388017

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021). The Same Ultra-Rapid Parallel Brain Dynamics Underpin the Production and Perception of Speech. *Cerebral Cortex Communications*, *2*(3), tgab040. https://doi.org/10.1093/texcom/tgab040

Falandays, J. B., Nguyen, B., & Spivey, M. J. (2021). Is prediction nothing more than multi-scale pattern completion of the future? *Brain Research*, *1768*, 147578. https://doi.org/10.1016/j.brainres.2021.147578

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Faust, M., & Kravetz, S. (1998). Levels of Sentence Constraint and Lexical Decision in the Two Hemispheres. *Brain and Language*, *62*(2), 149–162. https://doi.org/10.1006/brln.1997.1892

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999a). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., & Kutas, M. (1999b). Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373–392. https://doi.org/10.1016/S0926-6410(99)00036-1

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 202–224. https://doi.org/10.1037/0278-7393.27.1.202

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1017/S0048577202001373

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101

Feng, C., Damian, M. F., & Qu, Q. (2021). Parallel Processing of Semantics and Phonology in Spoken Production: Evidence from Blocked Cyclic Picture Naming and EEG. *Journal of Cognitive Neuroscience*, *33*(4), 725–738. https://doi.org/10.1162/jocn_a_01675

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20. https://doi.org/10.1023/A:1005290706460

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Qiu, Z. (2021). Predicting syntactic structure. *Brain Research*, *1770*, 147632. https://doi.org/10.1016/j.brainres.2021.147632

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 1–20. https://doi.org/10.1016/S0022-5371(79)90534-6

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522. https://doi.org/10.1121/1.427148

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70–79. https://doi.org/10.1121/1.396876

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and Foreign Accent Processing in English: Can Listeners Adapt? *Journal of Psycholinguistic Research*, *38*(4), 379–412. https://doi.org/10.1007/s10936-008-9097-8

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1276–1293. https://doi.org/10.1037/0096-1523.32.5.1276

Fodor, J. A. (1983). *The Modularity of Mind*. The MIT Press.

Forgie, J. W., & Forgie, C. D. (1959). Results Obtained from a Vowel Recognition Computer Program. *The Journal of the Acoustical Society of America*, *31*(11), 1480–1489. https://doi.org/10.1121/1.1907653

Forster, K. I. (1981). Priming and the Effects of Sentence and Lexical Contexts on Naming Time: Evidence for Autonomous Lexical Processing. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 465–495. https://doi.org/10.1080/14640748108400804

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, *14*(1), 3–28. https://doi.org/10.1016/S0095-4470(19)30607-2

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage.

Frank, S. L. (2024). Neural language model gradients predict event-related brain potentials. *Proceedings of the Society for Computation in Linguistics 2024*, 316–323.

Frank, S. L., & Bod, R. (2011). Insensitivity of the Human Sentence-Processing System to Hierarchical Structure. *Psychological Science*, *22*(6), 829–834. https://doi.org/10.1177/0956797611409589

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: evidence from brain potentials. *Language, Cognition and Neuroscience*, *31*(9), 1193–1205. https://doi.org/10.1080/23273798.2016.1205202

Friedman, D., Kazmerski, V. A., & Cycowicz, Y. M. (1998). Effects of aging on the novelty P3 during attend and ignore oddball tasks. *Psychophysiology*, *35*(5), 508–520. https://doi.org/10.1017/S0048577298970664

Friedman, D., & Simpson, G. V. (1994). ERP amplitude and scalp distribution to target and novel events: effects of temporal order in young, middle-aged and older adults. *Cognitive Brain Research*, *2*(1), 49–63. https://doi.org/10.1016/0926-6410(94)90020-5

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, *44*(3). https://doi.org/10.1111/cogs.12814

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. https://doi.org/10.3758/BF03193857

Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of*

Experimental Child Psychology, *173*, 351–370. https://doi.org/10.1016/j.jecp.2018.04.012

Gambi, C., Jindal, P., Sharpe, S., Pickering, M. J., & Rabagliati, H. (2021). The Relation Between Preschoolers' Vocabulary Development and Their Ability to Predict and Recognize Words. *Child Development*, *92*(3), 1048–1066. https://doi.org/10.1111/cdev.13465

Gambi, C., & Pickering, M. J. (2017). Models Linking Production and Comprehension. In *The Handbook of Psycholinguistics* (pp. 157–181). Wiley. https://doi.org/10.1002/9781118829516.ch7

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The Search for "Common Sense": An Electrophysiological Study of the Comprehension of Words and Pictures in Reading. *Journal of Cognitive Neuroscience*, *8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89

Gareth Gaskell, M., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266. https://doi.org/10.1016/S0010-0285(02)00003-8

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language*, *37*(1), 58–93. https://doi.org/10.1006/jmla.1997.2512

Gastaldon, S., Arcara, G., Navarrete, E., & Peressotti, F. (2020). Commonalities in alpha and beta neural desynchronizations during prediction in language comprehension and production. *Cortex*, *133*, 328–345. https://doi.org/10.1016/j.cortex.2020.09.026

Gastaldon, S., Bonfiglio, N., Vespignani, F., & Peressotti, F. (2024). Predictive language processing: integrating comprehension and production, and what atypical populations can tell us. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1369177

Gastaldon, S., Busan, P., Arcara, G., & Peressotti, F. (2023). Inefficient speech-motor control affects predictive speech comprehension: atypical electrophysiological correlates in stuttering. *Cerebral Cortex*, *33*(11), 6834–6851. https://doi.org/10.1093/cercor/bhad004

Gillis, M., Vanthornhout, J., Simon, J. Z., Francart, T., & Brodbeck, C. (2021). Neural Markers of Speech Comprehension: Measuring EEG Tracking of Linguistic Speech Representations, Controlling the Speech Acoustics. *The Journal of Neuroscience*, *41*(50), 10316–10329. https://doi.org/10.1523/JNEUROSCI.0812-21.2021

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183. https://doi.org/10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Gosselin, P. A., & Gagné, J.-P. (2011). Older adults expend more listening effort than young adults recognizing audiovisual speech in noise. *International Journal of Audiology*, *50*(11), 786–792. https://doi.org/10.3109/14992027.2011.599870

Grainger, J., & Holcomb, P. J. (2009). An ERP investigation of orthographic priming with relative-position and absolute-position primes. *Brain Research*, *1270*, 45–53. https://doi.org/10.1016/j.brainres.2009.02.080

Greenwood, D. D. (1961). Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane. *The Journal of the Acoustical Society of America*, *33*(10), 1344–1356. https://doi.org/10.1121/1.1908437

Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605. https://doi.org/10.1121/1.399052

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108. https://doi.org/10.1016/j.jneuroling.2016.12.001

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*, 377–388.

Grice, H. P. (1968). Utterer's meaning, sentence meaning and word meaning. *Foundations of Language*, *4*, 225–242.

Grisoni, L., McCormick Miller, T., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. *Journal of Neuroscience, 37(18), 4848-4858*. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Groppe, D. M., Choi, M., Huang, T., Schilz, J., Topkins, B., Urbach, T. P., & Kutas, M. (2010). The phonemic restoration effect reveals pre-N400 effect of supportive sentence context in speech perception. *Brain Research*, *1361*, 54–66. https://doi.org/10.1016/j.brainres.2010.09.003

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, *38*(4), 299–310. https://doi.org/10.3758/BF03207159

Gunnar, F. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Gut, U. (2012). The LeaP corpus: A multilingual corpus of spoken learner German and learner English. In T. Schmidt & K. Wörner (Eds.), *Hamburg studies on multilingualism: Vol. 14. Multilingual corpora and multilingual corpus analysis* (pp. 3–23). John Benjamins Publishing Company.

Gutnisky, D. A., & Dragoi, V. (2008). Adaptive coding of visual information in neural populations. *Nature*, *452*(7184), 220–224. https://doi.org/10.1038/nature06563

Gwilliams, L., King, J.-R., Marantz, A., & Poeppel, D. (2022). Neural dynamics of phoneme sequences reveal position-invariant code for content and order. *Nature Communications*, *13*(1), 6606. https://doi.org/10.1038/s41467-022-34326-1

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (Seventh Edition). Pearson.

Hakonen, M., May, P. J. C., Jääskeläinen, I. P., Jokinen, E., Sams, M., & Tiitinen, H. (2017). Predictive processing increases intelligibility of acoustically distorted speech: Behavioral and neural correlates. *Brain and Behavior*, *7*(9), e00789. https://doi.org/10.1002/brb3.789

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies 2001 - NAACL '01*, 1–8. https://doi.org/10.3115/1073336.1073357

Hale, J. (2016). Information-theoretical Complexity Metrics. *Language and Linguistics Compass*, *10*(9), 397–412. https://doi.org/10.1111/lnc3.12196

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic Properties of Stop Consonants. *The Journal of the Acoustical Society of America*, *29*(1), 107–116. https://doi.org/10.1121/1.1908634

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*(3–4), 373–405. https://doi.org/10.1016/j.wocn.2003.09.006

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & de Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. https://doi.org/10.1073/pnas.2201968119

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection - A review and recommendations for the practicing statistician. *Biometrical Journal*, *60*(3), 431–449. https://doi.org/10.1002/bimj.201700067

Hillinger, M. L. (1980). Priming effects with phonemically similar words: *Memory & Cognition*, *8*(2), 115–123. https://doi.org/10.3758/BF03213414

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error-based implicit learning signal during language comprehension. *European Journal of Neuroscience*, *54*(9), 7125–7140. https://doi.org/10.1111/ejn.15462

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences*, *1*(II). https://doi.org/10.33735/phimisci.2020.II.64

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

Huettig, F., Audring, J., & Jackendoff, R. (2022). A parallel architecture perspective on pre-activation and prediction in language processing. *Cognition*, *224*, 105050. https://doi.org/10.1016/j.cognition.2022.105050

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Hutchinson, J. B., & Barrett, L. F. (2019). The Power of Predictions: An Emerging Paradigm for Psychological Research. *Current Directions in Psychological Science*, *28*(3), 280–291. https://doi.org/10.1177/0963721419831992

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544

Hyde, M. (1997). The N1 Response and Its Applications. *Audiology and Neurotology*, *2*(5), 281–307. https://doi.org/10.1159/000259253

Irino, T., & Patterson, R. D. (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilised wavelet-Mellin transform. *Speech Communication*, *36*(3–4), 181–203. https://doi.org/10.1016/S0167-6393(00)00085-6

Ito, A. (2019). Prediction of orthographic information during listening comprehension: A printed-word visual world study. *Quarterly Journal of Experimental Psychology*, *72*(11), 2584–2596. https://doi.org/10.1177/1747021819851394

Ito, A. (2024). Phonological prediction during comprehension: A review and meta-analysis of visual-world eye-tracking studies. *Journal of Memory and Language*, *139*, 104553. https://doi.org/10.1016/j.jml.2024.104553

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Gambi, C., Pickering, M. J., Fuellenbach, K., & Husband, E. M. (2020). Prediction of phonological and gender information: An event-related potential study in Italian. *Neuropsychologia*, *136*, 107291. https://doi.org/10.1016/j.neuropsychologia.2019.107291

Ito, A., & Husband, E. M. (2017). How robust are effects of semantic and phonological prediction during language comprehension ? A visual world eye-tracking study. *IEICE Technical Report*, *117(149)*, 1–6.

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language,*

Cognition and Neuroscience, *32*(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Ito, A., & Sakai, H. (2021). Everyday Language Exposure Shapes Prediction of Specific Words in Listening Comprehension: A Visual World Eye-Tracking Study. *Frontiers in Psychology*, *12*, 607474. https://doi.org/10.3389/fpsyg.2021.607474

Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001

Jackendoff, R., & Audring, J. (2020). *The texture of the lexicon: Relational morphology and the parallel architecture*. Oxford University Press.

Jensen, K. M., & MacDonald, J. A. (2023). Towards thoughtful planning of ERP studies: How participants, trials, and effect magnitude interact to influence statistical power across seven ERP components. *Psychophysiology*, *60*(7), e14245. https://doi.org/10.1111/psyp.14245

Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 146–165). Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499. https://doi.org/10.1016/j.wocn.2005.08.004

Johnson, K., & Sjerps, M. J. (2021). Speaker Normalization in Speech Perception. In J. S. Pardo, L. C. Nygaard, R. R. Remez, & D. B. Pisoni (Eds.), *The Handbook of Speech Perception* (pp. 145–176). Wiley. https://doi.org/10.1002/9781119184096.ch6

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with More Than One Random Factor: Designs, Analytic Models, and Statistical Power. *Annual Review of Psychology*, *68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Justo-Guillén, E., Ricardo-Garcell, J., Rodríguez-Camacho, M., Rodríguez-Agudelo, Y., Lelo de Larrea-Mancera, E. S., & Solís-Vivanco, R. (2019). Auditory mismatch detection, distraction, and attentional reorientation (MMN-P3a-RON) in neurological and psychiatric disorders: A review. *International Journal of Psychophysiology*, *146*, 85–100. https://doi.org/10.1016/j.ijpsycho.2019.09.010

Kaiser, E., & Trueswell, J. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113–147. https://doi.org/10.1016/j.cognition.2004.01.002

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*(1), 13745. https://doi.org/10.1038/s41598-018-32150-6

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kerr, E., Ivanova, B., & Strijkers, K. (2023). Lexical access in speech production: Psycho- and neurolinguistics perspectives on the spatiotemporal dynamics. In R. Hartsuiker & K. Strijkers (Eds.), *Current issues in the psychology of language: Language production* (pp. 32–65). Routledge (Taylor & Francis).

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *73*(1), 322–335. https://doi.org/10.1121/1.388813

Kim, A., & Lai, V. (2012). Rapid Interactions between Lexical Semantic and Word Form Analysis during Word Recognition in Context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *24*(5), 1104–1112. https://doi.org/10.1162/jocn_a_00148

Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, *34*(1), 43–68. https://doi.org/10.1080/23273798.2018.1500698

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Kleinschmidt, D. F., Weatherholtz, K., & Florian Jaeger, T. (2018). Sociolinguistic Perception as Inference Under Uncertainty. *Topics in Cognitive Science*, *10*(4), 818–834. https://doi.org/10.1111/tops.12331

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1981). The effects of frontal cortex lesions on event-related potentials during auditory selective attention. *Electroencephalography and Clinical Neurophysiology*, *52*(6), 571–582. https://doi.org/10.1016/0013-4694(81)91431-0

Knight, R. T., & Nakada, T. (1998). A Review of EEG and Blood Flow Data. *Reviews in the Neurosciences*, *9*(1), 57–70. https://doi.org/10.1515/REVNEURO.1998.9.1.57

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577. https://doi.org/10.1017/S0048577201990559

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616. https://doi.org/10.1080/23273798.2015.1130233

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kurthen, I., Galbier, J., Jagoda, L., Neuschwander, P., Giroud, N., & Meyer, M. (2021). Selective attention modulates neural envelope tracking of informationally masked speech in healthy older adults. *Human Brain Mapping*, *42*(10), 3042–3057. https://doi.org/10.1002/hbm.25415

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A Look around at What Lies Ahead: Prediction and Predictability in Language Processing. In M. Bar (Ed.), *Predictions in the Brain: Using our past to generate a future* (pp. 190–207). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195395518.003.0065

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory States and Processes* (pp. 217–237). Erlbaum.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles & J. R. Jennings (Eds.), *Advances in Psychophysiology* (pp. 139–187). JAI.

Labov, W. (2022). *Sociolinguistics Patterns*. University of Pennsylvania Press.

Ladefoged, P. (1980). What Are Linguistic Sounds Made of? *Language*, *56*(3), 485. https://doi.org/10.2307/414446

Ladefoged, P. (2006). *A course in phonetics*. Thomson Wadsworth Corporation.

Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, *31*(1), 189–193. https://doi.org/10.1111/j.1460-9568.2009.07055.x

Lang, M., Lang, W., Uhl, F., Kornhuber, A., Deecke, L., & Kornhuber, H. H. (1987). Slow negative potential shifts in a verbal concept for mation task. *Human Neurobiology*, *6*, 183–190.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 Effects of Prediction from Association in Single-word Contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*(1), 74–88. https://doi.org/10.1016/j.bandl.2006.02.003

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Lenth, R. (2025). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.11.2-80003, https://rvlenth.github.io/emmeans/.*

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *Package "emmeans."*

León-Cabrera, P., Flores, A., Rodríguez-Fornells, A., & Morís, J. (2019). Ahead of time: Early sentence slow cortical modulations associated to semantic prediction. *Neuroimage, 189, 192-201*. https://doi.org/10.1016/j.neuroimage.2019.01.005

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewendon, J., Mortimore, L., & Egan, C. (2020). The Phonological Mapping (Mismatch) Negativity: History, Inconsistency, and Future Direction. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01967

Lewis, R. L. (2000). Falsifying Serial and Parallel Parsing Models: Empirical Conundrums and An Overlooked Paradigm. *Journal of Psycholinguistic Research*, *29*(2), 241–248. https://doi.org/10.1023/A:1005105414238

Li, J., & Futrell, R. (2023). A decomposition of surprisal tracks the N400 and P600 brain potentials. In M. Goldwater, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Meeting of the Cognitive Science Society.*

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in Mandarin Chinese. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(5), 531–547. https://doi.org/10.1037/xhp0000999

Li, X., & Qu, Q. (2024). Verbal working memory capacity modulates semantic and phonological prediction in spoken comprehension. *Psychonomic Bulletin & Review*, *31*(1), 249–258. https://doi.org/10.3758/s13423-023-02348-5

Li, Y., Wen, Z., Hau, K.-T., Yuan, K.-H., & Peng, Y. (2020). Effects of Cross-loadings on Determining the Number of Factors to Retain. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(6), 841–863. https://doi.org/10.1080/10705511.2020.1745075

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*(2), 148–167. https://doi.org/10.1037/0003-066X.37.2.148

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36. https://doi.org/10.1016/0010-0277(85)90021-6

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Lindborg, A., Musiolek, L., Ostwald, D., & Rabovsky, M. (2023). Semantic surprise predicts the N400 brain potential. *Neuroimage: Reports*, *3*(1), 100161. https://doi.org/10.1016/j.ynirp.2023.100161

Luck, S. J. (2014). *An Introduction to the Event-related Potential Technique* (second ed.). MIT Press.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98. https://doi.org/10.1016/0010-0285(92)90003-K

Magyari, L., & de Ruiter, J. P. (2012). Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology*, *3*, Article 376. https://doi.org/10.3389/fpsyg.2012.00376

Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, *32*(1), 4–18. https://doi.org/10.1111/j.1469-8986.1995.tb03400.x

Mani, N., & Plunkett, K. (2011). Phonological priming and cohort effects in toddlers. *Cognition*, *121*(2), 196–206. https://doi.org/10.1016/j.cognition.2011.06.013

Marin, S., Pouplier, M., & Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of /t/ and /k/ and its implications for speech error research. *The Journal of the Acoustical Society of America*, *127*(1), 445–461. https://doi.org/10.1121/1.3268600

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Marr, D. C. (1982). *Vision A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63. https://doi.org/10.1016/0010-0285(78)90018-X

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is Production: The missing link between language production and comprehension. *Scientific Reports*, *8*(1), 1079. https://doi.org/10.1038/s41598-018-19499-4

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Maule, J., & Franklin, A. (2020). Adaptation to variance generalizes across visual domains. *Journal of Experimental Psychology: General*, *149*(4), 662–675. https://doi.org/10.1037/xge0000678

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science*, *32*(3), 543–562. https://doi.org/10.1080/03640210802035357

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: trainable text-speech alignment using Kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association*.

McCallum, W. C., Cooper, R., & Pocock, P. V. (1988). Brain slow potential and ERP changes associated with operator load in a visual tracking task. *Electroencephalography and Clinical Neurophysiology*, *69*(5), 453–468. https://doi.org/10.1016/0013-4694(88)90068-5

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., & O'Regan, J. K. (1981). Expectations increase the benefit derived from parafoveal visual information in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 634–644. https://doi.org/10.1037/0096-1523.7.3.634

McDonald, S. A., & Shillcock, R. C. (2003). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science*, *14*(6), 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130. https://doi.org/10.1037/0096-3445.126.2.99

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Mesik, J., & Wojtczak, M. (2023). The effects of data quantity on performance of temporal response function analyses of natural speech processing. *Frontiers in Neuroscience*, *16*, 963629. https://doi.org/10.3389/fnins.2022.963629

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Miceli, G., Gainotti, G., Caltagirone, C., & Masullo, C. (1980). Some aspects of phonological impairment in aphasia. *Brain and Language*, *11*(1), 159–169. https://doi.org/10.1016/0093-934X(80)90117-0

Michaelov, J. A., Bardolph, M. D., Van Petten, C. K., Bergen, B. K., & Coulson, S. (2024). Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects. *Neurobiology of Language*, *5*(1), 107–135. https://doi.org/10.1162/nol_a_00105

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early Parallel Activation of Semantics and Phonology in Picture Naming: Evidence from a Multiple Linear Regression MEG Study. *Cerebral Cortex*, *25*(10), 3343–3355. https://doi.org/10.1093/cercor/bhu137

Möcks, J. (1986). The Influence of Latency Jitter in Principal Component Analysis of Event-Related Potentials. *Psychophysiology*, *23*(4), 480–484. https://doi.org/10.1111/j.1469-8986.1986.tb00659.x

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*, 120–132. https://doi.org/10.1016/j.neuroimage.2013.01.031

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3), 176–190. https://doi.org/10.1016/j.biopsycho.2009.12.006

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1273–1280. https://doi.org/10.1098/rstb.2008.0314

Munro, M. J., & Derwing, T. M. (1995). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Naeser, M. A., Palumbo, C. L., Helm-Estabrooks, N., Stiassny-Eder, D., & Albert, M. L. (1989). Severe nonfluency in aphasia: Role of the medical subcallosal fasciculus and other white matter pathways in recovery of spontaneous speech. *Brain*, *112*(1), 1–38. https://doi.org/10.1093/brain/112.1.1

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, *85*(5), 2088–2113. https://doi.org/10.1121/1.397861

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes. *Memory & Cognition*, *4*(5), 648–654. https://doi.org/10.3758/BF03213230

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226–254. https://doi.org/10.1037/0096-3445.106.3.226

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003–1019. https://doi.org/10.1037/0278-7393.15.6.1003

Newman, R. L., & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–121. https://doi.org/10.1016/j.biopsycho.2008.04.008

Newman, R. L., Connolly, J. F., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640–647. https://doi.org/10.1111/1469-8986.00065

Newman R. L., Forbes, K., & Connolly, J. F. (2012). Event-related potentials and magnetic fields associated with spoken word recognition. In M. Spivey & K. McRae (Eds.), *Cambridge handbook of psycholinguistics* (pp. 127–158). Cambridge University Press.

Nicenboim, B., Vasishth, S., & Rösler, F. (2020). Are words pre-activated probabilistically during sentence comprehension? Evidence from new data and a Bayesian random-effects meta-analysis using publicly available data. *Neuropsychologia*, *142*, 107427. https://doi.org/10.1016/j.neuropsychologia.2020.107427

Nickels, L. (2014). *Spoken Word Production and Its Breakdown In Aphasia*. Psychology Press. https://doi.org/10.4324/9781315804620

Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews*, *96*, 367–400. https://doi.org/10.1016/j.neubiorev.2018.11.019

Nieuwland, M. S., Arkhipova, Y., & Rodríguez-Gómez, P. (2020). Anticipating words during spoken discourse comprehension: A large-scale, pre-registered replication study using brain potentials. *Cortex*, *133*, 1–36. https://doi.org/10.1016/j.cortex.2020.09.007

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., Husband, M. E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, e33468. https://doi.org/10.7554/eLife.33468

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When Peanuts Fall in Love: N400 Evidence for the Power of Discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111. https://doi.org/10.1162/jocn.2006.18.7.1098

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Nour Eddine, S., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2024). A predictive coding model of the N400. *Cognition*, *246*, 105755. https://doi.org/10.1016/j.cognition.2024.105755

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. https://doi.org/10.3758/BF03206860

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech Perception as a Talker-Contingent Process. *Psychological Science*, *5*(1), 42–46. https://doi.org/10.1111/j.1467-9280.1994.tb00612.x

Ogar, J., Slama, H., Dronkers, N., Amici, S., & Luisa Gorno-Tempini, M. (2005). Apraxia of Speech: An overview. *Neurocase*, *11*(6), 427–432. https://doi.org/10.1080/13554790500263529

Öhman, S. E. G. (1966). Coarticulation in VCV Utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America*, *39*(1), 151–168. https://doi.org/10.1121/1.1909864

Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, *98*(1), 112–117. https://doi.org/10.1016/j.bandl.2006.04.006

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology*, *60*(2–3), 121–150. https://doi.org/10.1016/S0301-0511(02)00045-5

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, *67*(4), 426–448. https://doi.org/10.1016/j.jml.2012.07.003

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328. https://doi.org/10.1037/0278-7393.19.2.309

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169–181. https://doi.org/10.1016/j.cortex.2015.03.006

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peronnet, F., & Farah, M. J. (1989). Mental rotation: An event-related potential study with a validated mental rotation task. *Brain and Cognition*, *9*(2), 279–288. https://doi.org/10.1016/0278-2626(89)90037-7

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*, *109*(6), 2988–2998. https://doi.org/10.1121/1.1370525

Peterson, G. E., & Barney, H. L. (1951). Control Methods Used in a Study of the Vowels. *The Journal of the Acoustical Society of America*, *23*(1_Supplement), 148–148. https://doi.org/10.1121/1.1917300

Petten, C. Van. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, *8*(4), 485–531. https://doi.org/10.1080/01690969308407586

Piazza, G., Carta, S., Ip, E. Y. J., Pérez-Navarro, J., Kalashnikova, M., Martin, C. D., & Di Liberto, G. M. (2025). Are you talking to me? How the choice of speech register impacts listeners' hierarchical encoding of speech. *Imaging Neuroscience*, *3*. https://doi.org/10.1162/imag_a_00539

Piazza, G., Sala, M., Guerrini, R., Winchester, M. M., & Peressotti, F. (2025). *No Risky Bets: The Brain Avoids All-In Predictions During Naturalistic Multitalker Listening*. Preprint posted on *bioRxiv* (accessed 21 September 2025). https://doi.org/10.1101/2025.07.18.665546

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02), 169–190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. https://doi.org/10.1016/j.tics.2006.12.002

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pickering, M. J., & Garrod, S. (2014). Neural integration of language production and comprehension. *Proceedings of the National Academy of Sciences*, *111*(43), 15291–15292. https://doi.org/10.1073/pnas.1417917111

Pickering, M. J., & Strijkers, K. (2024). Language Production and Prediction in a Parallel Activation Model. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12775

Pisoni, D. B. (1992). Talker normalization in speech perception. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 143–151). Ohmsha.

Pisoni, D. B. (1997). Some thoughts on 'normalization' in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–33). Academic Press.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Porretta, V., Buchanan, L., & Järvikivi, J. (2020). When processing costs impact predictive processing: The case of foreign-accented speech and accent experience. *Attention, Perception, & Psychophysics*, *82*(4), 1558–1565. https://doi.org/10.3758/s13414-019-01946-7

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Lawrence Erlbaum Associates.

Poulton, V. R., & Nieuwland, M. S. (2022). Can You Hear What's Coming? Failure to Replicate ERP Evidence for Phonological Prediction. *Neurobiology of Language*, *3*(4), 556–574. https://doi.org/10.1162/nol_a_00078

Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *The Journal of the Acoustical Society of America*, *101*(4), 2267–2277. https://doi.org/10.1121/1.418247

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*(5), 187–189. https://doi.org/10.1016/S1364-6613(03)00092-5

R Core Team. (2023). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org*.

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, *31*(1), 94–105. https://doi.org/10.1080/23273798.2015.1077979

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 358–374. https://doi.org/10.1016/S0022-5371(83)90236-0

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 651–666. https://doi.org/10.1037/0096-1523.23.3.651

Remez, R. E., Ferro, D. F., Dubowski, K. R., Meer, J., Broder, R. S., & Davids, M. L. (2010). Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, & Psychophysics*, *72*(8), 2054–2058. https://doi.org/10.3758/BF03196682

Riès, S. K., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., Kuperman, R. A., Auguste, K. I., Brunner, P., Schalk, G., Lin, J. J., Parvizi, J., Crone, N. E., Dronkers, N. F., & Knight, R. T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences*, *114*(23), E4530–E4538. https://doi.org/10.1073/pnas.1620669114

Ritter, W., Simson, R., & Vaughan, H. G. (1988). Effects of the amount of stimulus information processed on negative event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *69*(3), 244–258. https://doi.org/10.1016/0013-4694(88)90133-2

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To Predict or Not to Predict: Influences of Task and Strategy on the Processing of Semantic Relations. *Journal of Cognitive Neuroscience*, *19*(8), 1259–1274. https://doi.org/10.1162/jocn.2007.19.8.1259

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255. https://doi.org/10.1016/j.neuropsychologia.2016.03.022

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object Shape and Orientation Do Not Routinely Influence Performance During Language Processing. *Psychological Science*, *24*(11), 2218–2225. https://doi.org/10.1177/0956797613490746

Ruchkin, D. S., Canoune, H. L., Johnson, R., & Ritter, W. (1995). Working memory and preparation elicit different patterns of slow wave event-related brain potentials. *Psychophysiology*, *32*(4), 399–410. https://doi.org/10.1111/j.1469-8986.1995.tb01223.x

Ruchkin, D. S., Johnson, R., Mahaffey, D., & Sutton, S. (1988). Toward a Functional Categorization of Slow Waves. *Psychophysiology*, *25*(3), 339–353. https://doi.org/10.1111/j.1469-8986.1988.tb01253.x

Sakayori, S., Kitama, T., Chimoto, S., Qin, L., & Sato, Y. (2002). Critical spectral regions for vowel identification. *Neuroscience Research*, *43*(2), 155–162. https://doi.org/10.1016/S0168-0102(02)00026-3

Sala, M., Vespignani, F., Casalino, L., & Peressotti, F. (2024). I know how you'll say it: evidence of speaker-specific speech prediction. *Psychonomic Bulletin & Review*, *31*, 2332–2344. https://doi.org/10.3758/s13423-024-02488-2

Sala, M., Vespignani, F., Gastaldon, S., Casalino, L., & Peressotti, F. (2025). In the Words of Others: ERP Evidence of Speaker-Specific Phonological Prediction. *Psychophysiology*, *62*(9), e70135. https://doi.org/10.1111/psyp.70135

Samuel, A. (1996). Phoneme Restoration. *Language and Cognitive Processes*, *11*(6), 647–654. https://doi.org/10.1080/016909696387051

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 915–930. https://doi.org/10.1037/0278-7393.18.5.915

Scharf, F., & Nestler, S. (2019). Exploratory structural equation modeling for event-related potential data—An all-in-one approach? *Psychophysiology*, *56*(3), e13303. https://doi.org/10.1111/psyp.13303

Scharf, F., Widmann, A., Bonmassar, C., & Wetzel, N. (2022). A tutorial on the use of temporal principal component analysis in developmental ERP research – Opportunities and challenges. *Developmental Cognitive Neuroscience*, *54*, 101072. https://doi.org/10.1016/j.dcn.2022.101072

Schiffer, S. (1972). *Meaning*. Clarendon Press.

Schiller, N. O., Boutonnet, B. P.-A., De Heer Kloots, M. L. S., Meelen, M., Ruijgrok, B., & Cheng, L. L.-S. (2020). (Not so) Great Expectations: Listening to Foreign-Accented Speech Reduces the Brain's Anticipatory Processes. *Frontiers in Psychology*, *11*, Article 2143. https://doi.org/10.3389/fpsyg.2020.02143

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *14*(2), 344–354. https://doi.org/10.1037/0278-7393.14.2.344

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*(2), 232–252. https://doi.org/10.1016/0749-596X(85)90026-9

Searle, J. R. (1969). *Speech acts*. Cambride University Press.

Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*(7079), 936–942. https://doi.org/10.1038/nature04519

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696. https://doi.org/10.1073/pnas.1323812111

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Sivonen, P., Maess, B., Lattner, S., & Friederici, A. D. (2006). Phonemic restoration in a sentence context: Evidence from early and late ERP effects. *Brain Research*, *1121*(1), 177–189. https://doi.org/10.1016/j.brainres.2006.08.123

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, *32*(3), 329–350. https://doi.org/10.1017/S0047404503323024

Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., & Ković, V. (2022). How to do Better N400 Studies: Reproducibility, Consistency and Adherence to Research Standards in the Existing Literature. *Neuropsychology Review*, *32*(3), 577–600. https://doi.org/10.1007/s11065-021-09513-4

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology*, *36*(3), 409–414. https://doi.org/10.1017/S0048577299981180

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *47*(2), 276–309. https://doi.org/10.1037/h0078826

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, *112*(1), 1–36. https://doi.org/10.1037/0096-3445.112.1.1

Staub, A. (2015). The Effect of Lexical Predictability on Eye Movements in Reading: Critical Review and Theoretical Interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. https://doi.org/10.1111/lnc3.12151

Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425–436. https://doi.org/10.1037/0278-7393.32.2.425

Staum Casasanto, L. (2008). Does social information influence sentence processing? *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). McGraw-Hill.

Stevens, K. N., & Blumstein, S. E. (1977). Onset spectra as cues for consonantal place of articulation. *The Journal of the Acoustical Society of America*, *61*(S1), S48–S48. https://doi.org/10.1121/1.2015732

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1–38). Erlbaum.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Stocker, A. A., & Simoncelli, E. P. (2006). Sensory adaptation within a bayesian framework for perception. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems* (Vol. 18, pp. 1291–1298). MIT Press.

Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, *176*, 232–247. https://doi.org/10.1016/j.cognition.2018.03.011

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *The Journal of the Acoustical Society of America*, *85*(5), 2135–2153. https://doi.org/10.1121/1.397863

Strijkers, K. (2016). A Neural Assembly–Based View on Word Production: The Bilingual Test Case. *Language Learning*, *66*(S2), 92–131. https://doi.org/10.1111/lang.12191

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*(4), 484–503. https://doi.org/10.1080/23273798.2015.1120878

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex*, *20*(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N. J. Smelser & D. N. Gerstein (Eds.), *Behavioral and social science: Fifty years of discovery: In commemoration of the fiftieth anniversary of the "Ogburn Report: Recent Social Trends in the United States."* (pp. 208–248). National Academy Press.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. (2012). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 397–439). Oxford University Press.

Szostak, C. M., & Pitt, M. A. (2013). The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, & Psychophysics*, *75*(7), 1533–1546. https://doi.org/10.3758/s13414-013-0492-3

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics (2nd. ed.)*. Harper & Row.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Taylor, W. L. (1953). "Cloze Procedure": A New Tool for Measuring Readability. *Journalism Quarterly*, *30*(4), 415–433. https://doi.org/10.1177/107769905303000401

The MathWorks Inc. (2023). *MATLAB version: 23.2.0 (R2023b)*. The MathWorks Inc.

The MathWorks Inc. (2024). *MATLAB version: 24.2.0 (R2024b)*. The MathWorks Inc.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. https://doi.org/10.1016/j.ijpsycho.2011.12.007

Traxler, M. J. (2014). Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, *18*(11), 605–611. https://doi.org/10.1016/j.tics.2014.08.001

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1266–1282. https://doi.org/10.1037/0278-7393.26.5.1266

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct Attachment Is Not a Form of Lexical Ambiguity Resolution. *Journal of Memory and Language*, *39*(4), 558–592. https://doi.org/10.1006/jmla.1998.2600

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition*, *5*(1), 84–89. https://doi.org/10.3758/BF03209197

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual Integration and Lipreading Abilities of Older Adults with Normal and Impaired Hearing. *Ear & Hearing*, *28*(5), 656–668. https://doi.org/10.1097/AUD.0b013e31812f7185

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. *Psychonomic Bulletin & Review*, *7*(2), 320–326. https://doi.org/10.3758/BF03212988

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic Integration in Sentences and Discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657–671. https://doi.org/10.1162/089892999563724

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*(2), 284–307. https://doi.org/10.1016/j.jml.2004.11.003

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). Speech Understanding Oppositely Affects Acoustic and Linguistic Neural Tracking in a Speech

Rate Manipulation Paradigm. *The Journal of Neuroscience*, *42*(39), 7442–7453. https://doi.org/10.1523/JNEUROSCI.0259-22.2022

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive Mechanisms in Idiom Comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682–1700. https://doi.org/10.1162/jocn.2009.21293

Vilares, I., & Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. https://doi.org/10.1111/j.1749-6632.2011.05965.x

Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes*. CRC Press.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Walker, A., & Hay, J. (2011). Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology*, *2*(1). https://doi.org/10.1515/labphon.2011.007

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory Sequence: Confusion of Patterns Other Than Speech or Music. *Science*, *164*(3879), 586–587. https://doi.org/10.1126/science.164.3879.586

Weatherholtz, K., & Jaeger, T. F. (2016). Speech Perception and Generalization Across Talkers and Accents. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.95

Weber, A., Grice, M., & Crocker, M. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), B63–B72. https://doi.org/10.1016/j.cognition.2005.07.001

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical Tracking of Surprisal during Continuous Speech Comprehension. *Journal of Cognitive Neuroscience*, *32*(1), 155–166. https://doi.org/10.1162/jocn_a_01467

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168. https://doi.org/10.1016/S0304-3940(03)00599-8

Widaman, K. F. (1990). Bias in Pattern Loadings Represented by Common Factor Analysis and Component Analysis. *Multivariate Behavioral Research*, *25*(1), 89–95. https://doi.org/10.1207/s15327906mbr2501_11

Widaman, K. F. (2007). Common factors versus components: principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor Analysis at 100:*

Historical Developments and Future Directions (pp. 173–203). Lawrence Erlbaum Associates.

Wilbur, R. B., & Nolkn, S. B. (1986). The Duration of Syllables in American Sign Language. *Language and Speech*, *29*(3), 263–280. https://doi.org/10.1177/002383098602900306

Wilson, M., & Emmorey, K. (2006). Comparing Sign Language and Speech Reveals a Universal Limit on Short-Term Memory Capacity. *Psychological Science*, *17*(8), 682–683. https://doi.org/10.1111/j.1467-9280.2006.01766.x

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. https://doi.org/10.3758/BF03206432

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. https://doi.org/10.1016/S1364-6613(97)01070-X

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/S0960-9822(01)00432-8

Yates, A. (1987). *Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis*. State Univ. New York Press.

Zahorian, S. A., & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America*, *94*(4), 1966–1982. https://doi.org/10.1121/1.407520

Zhao, Z., Ding, J., Wang, J., Chen, Y., & Li, X. (2024). The flexibility and representational nature of phonological prediction in listening comprehension: evidence from the visual world paradigm. *Language and Cognition*, *16*(2), 481–504. https://doi.org/10.1017/langcog.2023.38

List of publications

Papers and conference presentations stemmed from the work presented in the thesis.

PAPERS

Sala, M., Vespignani, F., Casalino, L., & Peressotti, F. (2024). I know how you'll say it: evidence of speaker-specific speech prediction. *Psychonomic Bulletin & Review*, 31, 2332-2344. https://doi.org/10.3758/s13423-024-02488-2

Sala, M., Vespignani, F., Gastaldon, S., Casalino, L., & Peressotti, F. (2025). In the Words of Others: ERP Evidence of Speaker-Specific Phonological Prediction. *Psychophysiology*, *62*(9), e70135. https://doi.org/10.1111/psyp.70135

Piazza*, G., **Sala*, M.**, Guerrini, R., Winchester, M. M., & Peressotti, F. (2025). No Risky Bets: The Brain Avoids All-In Predictions During Naturalistic Multitalker Listening. Preprint posted on *bioRxiv* (accessed 21 September 2025). https://doi.org/10.1101/2025.07.18.665546

* Indicates shared first authorship

CONFERENCES

Sala, M., Piazza, G., Guerrini, R., & Peressotti F. Adapting predictions: Investigating the role of speech adaptation in predictive processing using Temporal Response Function. Mini-talk at the *31st Congresso AIP (Associazione Italiana Psicologia) – Sezione Sperimentale*, 11-13 September 2025, Torino, Italy.

Sala, M., Vespignani, F., Gastaldon, S., Casalino, L., & Peressotti F. Who will speak? Speaker Identity Shapes Phonological Predictions. Oral contribution at the *24th ESCoP (European Society for Cognitive Psychology) Conference*, 3-5 September 2025, Sheffield, UK.

Sala, M., Piazza, G., Guerrini, R., & Peressotti F. From Adaptation to Anticipation: How Speaker Familiarity Shapes Speech Predictions. Poster presented at the *17th ISP (International Symposium of Psycholinguistics) Conference*, 26-28 May 2025, Barcelona, Spain.

Sala, M., Vespignani, F., Gastaldon, S., Casalino, L., & Peressotti, F. Adapting expectations to speaker variability: speaker identity shapes phonological prediction. Poster presented at the *19th ABIM (Alpine Brain Imaging Meeting) Conference*, 12-16 January 2025, G.

Sala, M., Vespignani, F., Gastaldon, S., Casalino, L., & Peressotti, F. Contextual modulation of phonological predictions: Exploring the role of Speaker Identity. Poster presented at the

HLSC (Highlights in the Language Sciences) Conference, 8-11 July 2024, Nijmegen, Netherlands.

Sala, M., Vespignani, F., Gastaldon, S., Casalino, L., & Peressotti, F. In the words of others: How speaker identity shapes phonological prediction. Oral contribution at the *21st PiF (Psycholinguistics in Flanders) Conference*, 27-28 May 2024, Brussels, Belgium.

Sala, M., Vespignani, F., Casalino, L., & Peressotti, F. Does speaker's accent modulate phonological prediction? Mini-talk at the *29th Congresso AIP (Associazione Italiana Psicologia) – Sezione Sperimentale*, 18-20 September 2023, Lucca, Italy.

Sala, M., Vespignani, F., Casalino, L., & Peressotti, F. Does speaker's accent modulate phonological prediction? Poster presented at the *29th AMLAP (Architectures and Mechanisms for Language Processing) Conference*, 31 August - 2 September 2023, Donostia - San Sebastian, Spain.

Sala, M., Vespignani, F., & Peressotti, F. Does the linguistic identity of the speaker modulate speech prediction? Poster presented at the *Cognitive Science Arena Conference*, 18-20 February 2023, Brixen - Bressanone, Italy.

Acknowledgements

These pages mark the conclusion of a three-year journey in which I have grown not only as a researcher but also as a person. Along the way, I have been fortunate to meet people to whom I owe my deepest gratitude.

I am sincerely grateful to my supervisor, Prof. Francesca Peressotti, and my co-supervisor, Prof. Francesco Vespignani, for their support throughout this journey. My heartfelt thanks go to Francesca for always believing in me and in the journey we shared. She gave me freedom and guidance to pursue the research questions that inspired me using the approaches I considered most appropriate, while at the same time teaching me to be open to other perspectives. I am particularly grateful to her for always placing great value on the human side of the supervisor–student relationship, and for her patience in addressing my endless doubts. I would like to express my gratitude to Francesco for introducing me to the fundamentals of cognitive electrophysiology and for sparking my passion for this field. His technical and theoretical contributions to this work have been invaluable.

I am grateful to the collaborators with whom I worked on the studies included in this thesis, in particular Simone Gastaldon, Laura Casalino and Giorgio Piazza. My thanks go to Simone for his patience in teaching me how to collect EEG data, as well as for always being willing to share his expertise. To Laura, I am thankful for her support in the laboratory and for her positive presence, which made even the most stressful days more manageable. I am also grateful to Giorgio for our extensive and stimulating discussions on psycholinguistics and EEG data analysis, which I hope he found as enriching as I did.

I would like to thank my PhD colleagues from the 38th cycle for the support we have given each other throughout these years and for the friendship we built alongside our work. Thanks to them, this PhD has been not only a research journey but also a shared human experience.

I am also thankful to the people who were part of my Swiss experience during my visiting period at the University of Geneva. I sincerely thank Prof. Alexis Hervais-Adelman and the Dynamics of Brain and Language Lab for welcoming me and for making my time there an enriching experience, both professionally and personally. Their collaborative spirit and enthusiasm for research created a stimulating environment that inspired me to pursue new directions in my work. In particular, I would like to thank Martina Dordijevic and Enrico Varano for being not only wonderful colleagues but also dear friends.

Finally, I am deeply grateful to my family for their love and support: my father, Sergio,

my mother, Nunzia, and my brother, Riccardo. Their constant belief in me has been a source of strength and inspiration along this journey.

Marco Sala
September 2025