ABSTRACT

Title of Dissertation: FROM SOUND TO MEANING: QUANTIFYING CONTEXTUAL EFFECTS IN RESOLUTION OF L2 PHONOLEXICAL AMBIGUITY.

Anna Lukianchenko, Doctor of Philosophy, 2014

Directed By: Professor Kira Gor
Second Language Acquisition Program

Unlike native speakers, nonnative speakers perceive speech sounds through the prism of their native language (L1), which sometimes results in phonological ambiguity in their second language (L2). For example, Japanese learners experience difficulty discriminating English /r/ and /l/ sounds, which may lead to a lexical confusion between English words minimally different on this phonological contrast (e.g., “rock” and “lock”). While L1 comprehenders can rely on context to disambiguate meaning of spoken words, it is unclear whether L2 comprehenders have access to the same mechanisms, owning to their i) smaller vocabularies and weaker semantic associations between words; ii) use of shallow syntactic processing and decreased sensitivity to morphosyntactic violations; and iii) slowed meaning integration and prediction mechanisms.

Across four experiments, both behavioral and electrophysiological, this dissertation research aims to examine how information gleaned from the phonological level is brought together with that derived from the larger linguistic context. In particular, we are interested in identifying the extent to which different kinds of contextual information (semantic, morphological and syntactic) can potentially be utilized by L2

Russian speakers (as compared to native Russian speakers) for shaping interpretation of individual words, especially if they are perceptually ambiguous.

The results indicate that approximate and unstable nature of L2 phonological representations leads to phonolexical ambiguity in the L2, causing minimal pairs to become temporarily perceptually indistinguishable. Unlike phonolexically unambiguous words, ambiguous incongruent words do not incur processing costs associated with contextual integration, suggesting that L2 comprehenders disambiguate meaning through accessing their semantic, syntactic and morphological characteristics despite low-resolution phonological information. Syntactic and semantic contextual constraints appear to produce a stronger context effect than morphological constraints for both L1 and L2 groups. Although L2 representations may differ from those in L1 in that they may lack phonological specification and detail, the mechanisms associated with the use of contextual information for meaning resolution in auditory sentence comprehension are essentially the same in the L1 and the L2. The outcome of this dissertation work has widespread implications, both theoretical and pedagogical.

FROM SOUND TO MEANING:
QUANTIFYING CONTEXTUAL EFFECTS IN RESOLUTION OF L2 PHONOLEXICAL AMBIGUITY.

By

Anna Lukianchenko

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Kira Gor, Chair
Professor Nan Jiang
Professor Colin Phillips
Professor Scott Jackson
Professor Min Wang

Dedication

For and despite my son, Nicolas.

Acknowledgements

Although only my name appears on the title page of this dissertation, it is a product of a collective influence, both professional and personal, on me of so many amazing people who have crossed my path. I thank you all for your support and encouragement and for making my years in the graduate school such an enjoyable experience.

First and foremost, I want to express my deepest gratitude to my advisor and mentor Kira Gor. Thank you for your guidance, inspiration and support throughout these years. Thank you for always being available, for reading everything no matter how last minute I sent it, for allowing me to take detours on my way to graduation and at the same time keeping things in check, for giving me space and freedom when I needed it, for all your time and patience. I feel like I have known you for a very long time—long enough that it can't be measured in the number of emails that we have exchanged, cups of tea we have drunk at our weekly meetings, words we have written. I have learned so much from you, and I feel really lucky that one day you took me under your wing.

I would particularly like to thank Nan Jiang. Without you, I wouldn't be where I am now. Doing a Ph.D. was not in my plans until I took you course on second language acquisition. Thank you for encouraging me to apply and for kindling my passion for science. It proved to be a life-changing decision. I will never forget it.

I would like to express my great appreciation to Colin Phillips. I admire your powerful intellect, your shrewdness and your sense of vision to the point that sometimes I feel intimidated. Thank you for pushing me and seeing a potential in me, for making me think critically about linguistic problems, showing ways for improvement, and for reminding me that one should never stop questioning.

I am grateful to Scott Jackson for agreeing to be on my committee and for his help with the design of the ERP study. Thank you for sharing your expertise and your R genius—I have learned a lot from the multiple times I attended your R talks.

I thank Min Wang for willingly agreeing to serve as dean's representative for my dissertation defense.

It is my pleasure to acknowledge all my professors and colleagues at UMD for being so great at what they do and for generously sharing their knowledge. In particular, I would like to express admiration to William Idsardi and Robert DeKeyser, who represent the highest mark of scholarship and simply know the answer to any question. I am thankful to Naomi Feldman and Yasmeen Faroqi Shah, for investing their time in our project and the long discussions. I thank Jeffrey Chrabaszcz, Ellen Lau, Ilia kurenkov, Glynis MacMillan, Dan Parker and Kristen Wahlgren for various help with the ERP study and for sharing their knowledge and skills with me.

Perhaps, I have learned the most from my collaborators Candise Lin, Matthew Winn, and Svetlana Cook. Thank you very much for setting new standards of excellence—I will always look up to you. I am endlessly grateful to my best friends and colleagues Erika Hussey, Ilina Kachinske, and Sunyoung Lee-Ellis, who are not only great scientists but also great personalities. You have influenced me in many different ways and you continue to inspire me. A big thank you goes to all my classmates and colleagues in the SLA program and in the UMD Language Science community, with whom we shared good and tough times, but who made my experience in the graduate school unforgettable. I am also indebt to my non-linguist and non-academic friends in the U.S., Russia and worldwide for giving me a different perspective on life, for helping me get through the difficult times, for the comraderie and caring.

This dissertation would not have been possible without my participants. Let me acknowledge here your help and patience sitting at the computer and pressing buttons for hours—I hope that your time was not spent in vain, and that the results of this study will contribute to the cumulative scientific effort.

I would not have been able to finish this dissertation without the financial support. I am grateful to the UMD Graduate School for awarding me the Summer Research fellowship and the Graduate Dean's Dissertation fellowship; to Language Learning Journal for awarding me the Language Learning dissertation grant, and to IGERT and SLA program for their continuous support which enabled me to conduct research.

My deepest love and gratitude goes to my family, old and new. I thank my entire extended family for providing a loving environment. I thank Amy, James and Michael for accepting me the way I am, and for their endless support and generous help with pretty much everything. I thank my parents, Elena and Vitaly, and my dear babushka for their unconditioned love and their faith in me, no matter what. You raised me, educated me, supported me, taught me, and loved me all this time. And I love you in return.

Finally, I wish to thank my husband Jeffrey and my son Nicolas. I am lucky to be married to someone who shares my passion for research and science. Thank you for your support and for giving me confidence when I have doubts, for reminding me every day about what's most important in life, for picking me up when I felt down and for pushing me forward when I felt I couldn't do it anymore, for your sacrifices and your love. And thank you, Nico, for keeping me company throughout my dissertation writing. For your kicks when I was pregnant and then for keeping me awake all night, making sure I get my writing done with one hand while holding you in the other. You were the biggest motivation to get my dissertation finished, and I therefore dedicate it to you.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Overview

Ambiguity is present at many levels of language processing, including semantic (*bear* may refer to an animal species or to the act of supporting), syntactic (*bear* can be a noun or a verb), orthographic (when *bear* is heard, it may be interpreted as the word “bear” or “bare”), and perceptual (in the presence of noise, *bear* may be easily confused with *pear*). This dissertation is concerned with yet another kind of ambiguity, to which we refer as “phonolexical”. We believe that such ambiguity is characteristic of second language (L2) processing and originates from ambiguous phonological representations in the L2. Unclear, fuzzy phonological representations may render spoken word recognition problematic by making similar-sounding words highly confusable (or even homophonous). For example, Japanese speakers of L2-English experience difficulty discriminating /r/ and /l/, leading to confusion between English words minimally different on this phonological feature (as in *rock* and *lock*). In other words, if perceptual correlates of phonological representations are unstable in the L2, lexical representations may also become insufficiently differentiated. Phonolexical ambiguity is different from perceptual ambiguity (for example, due to noise) in that it is systematic and affects certain phonological contrasts, usually those that are not represented in the speaker’s native language (L1).

Despite pervasive ambiguity in speech, the human brain is remarkably capable of converging on the intended interpretation of a word in a matter of mere milliseconds, a feat that is achieved through the real-time interaction of the bottom-up and top-down information. A key to the speed and efficiency of spoken word recognition is likely to be contingent not only on the ability to synthesize information as it arrives from these two sources, but also on the ability to actively anticipate the upcoming input. Indeed, decades of research in cognitive science have established that the brain's ability to generate predictions about future events is a fundamental principle underlying many cognitive processes (Hawkins, 2004), including language comprehension (for an overview, see Chow, 2013; DeLong, 2009; Federmeier, 2007; Lau, 2009). For example, linguistic predictions, built up incrementally over the course of a sentence or other higher-order language context, can be used to pre-select certain semantic features or even pre-activate lexical items (Delong, Urbach, & Kutas, 2005; Federmeier & Kutas, 1999; Laszlo & Federmeier, 2009) and to predict syntactic relationships between words (Marslen-Wilson & Tyler, 1980; Tyler & Warren, 1987).

This dissertation makes an implicit assumption that although not all top-down information has to be of predictive nature, more often than not, sentential context is actively involved in constructing representations, on the basis of which hypotheses about upcoming signal are formed. Over a series of experiments, we manipulate the predictive power of different types of context (semantic, morphological and syntactic) in order to investigate how top-down contextual information contributes to resolution of phonological ambiguity in auditory speech comprehension. While it is a well-established fact that native speakers can utilize rich contextual information to compute online linguistic predictions during language comprehension, little is known about how sensitive nonnative speakers are to such contextual expectations due to i) smaller vocabularies and weaker semantic associations between words; ii) their use of shallow syntactic processing and decreased sensitivity to morphosyntactic violations. We adopt a synergistic approach, which consists in synthesizing evidence from a variety of experimental techniques (both behavioral and neurocognitive) and from different levels of linguistic analysis (acoustic-phonetic, word level, sentence level) in order to elucidate the mechanisms underlying the use of contextual cues for ambiguity resolution in the L2 compared to the L1.

1.2 Outline of the dissertation

This dissertation is based on a set of beliefs about what distinguishes native versus nonnative speech comprehension. One of the critical beliefs is centered around the idea that phonology plays a crucial role in spoken word recognition and acts as an activation code to the mental lexicon. Chapter 2 compares and contrasts the differences in native versus nonnative phonological acquisition, summarizes the properties of the two systems and outlines how they affect bottom-up processing in the two populations. Because bottom-up information has to be combined with top-down information in order for speech comprehension to happen, we review evidence on L1 and L2 speakers' use of contextual information for meaning resolution while outlining differences and similarities. Another belief is based on the idea that the level at which lexical processing and word recognition are carried out serves as a functional locus of the interaction between the bottom-up and the top-down input. To this end, we review some influential models of spoken word recognition and outline the assumptions they make.

Chapter 3 identifies the main motivating questions and hypotheses that the study aims to address. A brief description of the target phonological feature is provided and preliminary evidence from pilot experiments is reviewed.

Chapter 4 examines the consequences of L2 phonolexical ambiguity for word identification and explores how different kinds of sentence-level contextual cues are utilized for processing of ambiguous and unambiguous words by L2 compared to L1 listeners during sentence comprehension. Three behavioral experiments—lexical decision task (LDT) in context, translation judgment task (TJT), and self-paced listening (SPL) task—are reported. The goal of the LDT is to identify the contextual cues that are most effective for resolving ambiguity by quantifying their relative effects (e.g., facilitation or inhibition) on word recognition. The TJT provides evidence in favor of phonolexical ambiguity at word level. The SPL task aims to test how phonolexically ambiguous words are processed during real-time auditory comprehension, and which contextual information constrains word identification the most.

Chapter 5 presents electrophysiological (EEG) evidence on processing of phonolexical ambiguity in morphology. The EEG study capitalizes on the findings from the behavioral experiments and examines the temporal parameters of morpho-phonological processing in auditory sentence comprehension by L1 and L2 listeners.

Finally, the last chapter, Chapter 6, synthesizes the empirical findings reported in this dissertation and discusses their theoretical and practical implications.

2 Spoken word recognition in L1 and L2

2.1 L1 spoken word recognition

2.1.1 The role of phonology: Speed and efficiency

Phonology plays a crucial role in spoken word recognition as it determines which acoustic/phonetic properties of sounds are used in the language to signal lexical and grammatical distinctions and extracts relevant segmental and suprasegmental information from the speech signal to guide such decisions. Language-specific phonological behavior is largely driven by innate learning programs and is accomplished very early in life. As early as six months of age, infants have already established vowel prototypes (Kuhl, Williams, Lacerda, Stevens, and Lindblom, 1992) and continue phonological "tuning" to their native language sounds throughout the first year of life while gradually losing sensitivity to nonnative sounds and contrasts (Kuhl, 2004; Polka and Werker, 1994; Werker and Tees, 1984; Werker, 1995; for review, see Aslin, Jusczyk, and Pisoni, 1998). As infants begin to build a vocabulary and learn word meanings and forms, they add more phonetic detail to refine their phonological representations of lexical items and grammatical morphemes. This process is necessitated by the functional need to differentiate among a growing number of similar-sounding words as well as word forms (especially in morphologically rich languages) which enter the developing lexicon (Jusczyk, 1986; Mills, Prat, Zangl, Stager, Neville, and Werker, 2004; Swingley and Aslin, 2002). Such early phonological specialization and refinement is conceived of as a necessary mechanism needed to cope with the variability in the speech signal that is due to allophonic variation and inherent variance, which arise from a number of causes, including vocal tract differences across speakers, dialectal variation, speech rate, phonological processes (coarticulation, neutralization, etc.), environmental context and noise. The developing phonological system has “to learn” to make a distinction between what is acoustically/phonetically different and phonologically different on the one hand, and between what is acoustically/phonetically different but phonologically irrelevant, on the other.

Another demand imposed on the phonological system concerns the need to interpret as much of the incoming auditory information as possible and as quickly as possible (even in non-optimal conditions), because speech unfolds in real time, and one cannot go back in time to the part of the utterance that was not well-heard when the utterance has already been uttered. Thus, the phonological system “learns” to act fast and use numerous acoustic-phonetic cues to the identity of sounds that are relevant for meaning distinction. Indeed, electrophysiological studies demonstrate that phonemic categorization happens very early in speech processing—about 100 to 280 ms after stimulus onset (Dehaene-Lambertz, 1997; Näätänen et al., 1997). At this stage, all of the irrelevant phonetic details are ignored (i.e., they are not included in the final outcome of the phonological analysis) while functionally significant information used to encode meaning contrasts is preserved (Dehaene-Lambertz, Dupoux, and Gout, 2000; Kazanina, Phillips, and Idsardi, 2006). Importantly, the brain’s response (e.g., in the form of mismatch negativity) to a change in the phonemic status of a sound can be elicited in the absence of the subject’s attention (Näätänen, 1995), which indicates an obligatory, effortless and automatic nature of phonemic categorization as far as the native language is concerned.

Such properties of the phonological system are critical for successful word recognition. When sound hits the ear and enters the auditory system, relevant acoustic-phonetic information is extracted and is mapped onto corresponding phonological representations in the brain. While the units of phonological representations (phonetic segments, phonemes, distinctive features, syllables, spectral templates, or component articulatory gestures) are still widely debated, most influential models of spoken word recognition agree that the phonological information acts as a sort of activation code to the mental lexicon (Logogen model by Morton (1969); Lexical Access from Spectra (LAFS) by Klatt (1979); TRACE model by McClelland and Elman (1986), and Elman (1989); the Cohort model by Marslen-Wilson (1987); Shortlist by Norris (1994), and Norris, McQueen, Cutler, and Butterfield, (1997),]; the Distributed Cohort Model (DCM) by Gaskell and Marslen-Wilson (1997); ARTWORD by Grossberg and Myers (2000); PARSYN by Luce, Goldinger, Auer, and Vitevitch. (2000)). Using the metaphor of Altmann (1997), "a sequence of sounds is much like the combination to a safe; the tumblers in a combination lock fall into place as the correct sequence of rotations is performed" (p. 71). Similarly, lexical representations in the lexicon are activated on the basis of sequences of sounds unfolding in time. It is generally agreed that in the process of matching up the auditory input with the lexical representations, not just one, but multiple candidates get activated in the lexicon and start competing with each other for final selection. This notion of multiple access and competition of lexical candidates is the central component of connectionist models, such as TRACE, and models in the Cohort tradition. What is most crucial for the connectionist models is the total amount of phonological overlap between the input and the target lexical representation relative to the overlap with other candidates. Thus, for example, the input /bleɪs/ will be recognized as a token of *place* because there is a high degree of overlap between /bleɪs/ and /pleɪs/ (they are said to be phonological neighbors), and because there is no existing word *blace* and no other close competitor. In contrast, the Cohort model (Marslen-Wilson, 1987, 1989; Marslen-Wilson, Moss, and van Halen, 1996; Marslen-Wilson and Zwitserlood, 1989) makes strong claims about directionality and sequentiality of the lexical selection process, arguing that the phonological information contained at the beginning of a word defines the cohort of word candidates. For example, hearing /raɪ/ will activate all the words in the listener's mental lexicon that share this onset sequence, such as *rye, rice, right, ripe, rhyme*, etc. This initial pool of activated word candidates constitutes the initial word cohort. As more of the word is heard, the size of the cohort is reduced until there remains only one candidate that still matches the sensory input. On hearing /raɪt/, for example, the lexical processor will choose the word *right* as the best-fitting candidate among all the activated candidates (it is the only word that passes the "goodness of fit" criterion). This fit reflects the quality and the quantity of the match between the sensory input and the lexical form representation (Frauenfelder, Scholten, and Content, 2001). A critical factor determining when a word can be recognized is, therefore, the point at which it becomes unique—its uniqueness point (UP). Naturally, the word's UP is a function not just of the word itself, but also of the size of the cohort, i.e., the number of the word's possible competitors. Thus, many short words do not become unique until the word offset because they tend to have many competitors, but the UP in longer words can often happen before the end of the word (Luce, 1986; Tyler, 1984). By the same token, the word's UP occurs later in the word if the word is part of an inflectional paradigm (in highly inflected languages).

Empirical evidence for multiple lexical access and the contingency of lexical choice comes from cross-modal priming tasks. In a study by Zwitserlood (1989), native Dutch speakers heard incomplete sequences, such as the string /kapɪt/, which can be the onset of the Dutch words *kapitein* ("captain") and *kapitaal* ("capital"). The sound stopped at the /t/ in /kapɪt/ and the listeners saw a visual probe, which was associatively related to either captain (e.g., ship) or capital (e.g., money). Reaction times needed to make lexical decisions about the probes were compared with reaction times to the same probes embedded in the middle of control words, which were not semantically related. Priming effects, relative to the control condition, were observed for both *ship* and *money* when they were inserted after the /t/ in /kapɪt/. In the case when the probes were presented at the end of the word, e.g., after /n/ in /kapɪteɪn/, only the probe related to the target sound was facilitated (i.e., the word meaning *ship* in Dutch, but not *money*). In other words, both candidates (captain and capital) were activated as long as they were compatible with the auditory input. Once the auditorily presented word reached its UP, the competitor was dropped off, or deactivated. Additional evidence in favor of multiple activation of semantic codes comes from studies examining lexical ambiguity (Onifer and Swinney, 1981; Seidenberg, Tanenhaus, Leiman, and Bienkowski, 1982; Swinney, 1979; Tanenhaus, Leiman, and Seidenberg, 1979). In very general terms, such studies demonstrate that different meanings of a polysemous word (e.g., the word *bank* can refer to either the edge of a river or a financial institution) get activated when the listener encounters a polysemous word in sentence comprehension tasks.

In light of the evidence demonstrating that sensory input "activates" not just one word, but a whole class of possible lexical candidates, it follows that successful word recognition should be contingent on the ability of the phonological system to encode and categorize acoustic-phonetic information efficiently and accurately in order to prevent spurious activation of irrelevant candidates and, thus, overload the lexical system. With the goal to examine the role of phonological representations in word recognition, some studies set out to inspect the effect of the phonological mismatch between the auditory form and the lexical form. Marslen-Wilson and Zwitserlood (1989), for example, showed that a phonological mismatch of only one phoneme blocks lexical access to the target. They observed that in a cross-modal priming task, a prime such as *honing* (which means "honey" in Dutch) facilitated recognition of the target *bij* ("bee" in Dutch) whereas rhyming primes such as *woning* (means "dwelling") or *foning* (a nonword) were ineffective in priming the target word, although in a study by Connine, Blasko, and Titone (1993) nonword rhyming primes (such as foning) were found to activate rhyming targets. Andruski, Blumstein, and Burton (1994) manipulated the voice onset time (VOT) in the prime's initial consonant (e.g., reducing the VOT in the voiceless stop /p/ in *pear* will make the word sound more like a *bear*) and observed that responses to targets (e.g., *fruit*) were faster after *pear* than after an unrelated word (*bear*) and that the priming effect became smaller as the VOT in the prime's initial consonant decreased. Moreover, they also found that lexical decision times to pairs where primes had a voiced-stop counterpart (e.g., *pear—bear*) were slower than they were to pairs where primes had no voiced-stop lexical competitor (*king—ging*).

Because of the temporal nature of speech (the “left to right” analysis of the spoken word), disruptive effects of phonological mismatch are more noticeable in word-initial than in word-final position. However, even though the mismatching segment can arrive late in the word, when it does arrive, it can impede word recognition. Using a fragment priming technique, Soto-Faraco, Sebastián-Gallés, and Cutler (2001) demonstrated that fully matching primes facilitate decisions to a target word, whereas mismatching primes with a mismatch embedded in the middle of the word inhibit responses. In a similar vein of research, Slowiaczek and Hamburger (1992) and Hamburger and Slowiaczek (1996) showed that a phonological mismatch at the end of the word can also interfere with the word recognition processes. Specifically, they showed that, the greater the initial phonological overlap is between the two words, the greater interference costs there are for the lexical system. Thus, word-initial phonological overlap of only one phoneme (e.g., *green—goal*) results in a facilitatory effect on recognition of the target word, while a 3-phoneme overlap inhibits lexical access (e.g., *green—grief*). Besides word position, word length appears to matter for the size of the disruptive effect of the phonological mismatch: it has been shown that it affects recognition of short words to a greater extent (Gow, 2001) than long words, which can still be recovered (Connine, Blasko, and Titone, 1993). Another determinant factor in the size of the phonological mismatch effect on word identity concerns phonetic similarity between the word’s original sound and the substitute sound (McQueen, 2007): the more dissimilar the two sounds are (i.e., separated by a larger number of distinctive features, as in /t—n/ compared to /t—p/), the more disruptive the phonological mismatch effect is for lexical access (Marslen-Wilson et al., 1996).

Examining the results of the above studies, one could argue that, if word recognition had to rely solely on accurate, discrete phonemic information to access correct lexical candidates, the process would have been impossible because, unlike printed language, spoken language does not consist of separate, discrete events like letters on the page. Individual speech sounds like vowels and consonants overlap in their articulation, and their interpretation is contingent on the interpretation of adjacent sounds. Let us consider the following example from Altmann (1997). If a native speaker of English pronounces the phrase '*a thin book'* in a non-deliberate manner, instead of articulating the /n/ in *thin* and closing off the word with the tip of the tongue against the back of the upper teeth, the speaker might pronounce /m/ instead of /n/, thus assimilating the word-final sound in *thin* to the following /b/ sound in the word book so that the resulting phrase may sound something like 'a *thim book'*. Speech is abundant in examples of similar assimilation (or coarticulation) processes (e.g., vowels influenced by the voicing of the subsequent consonant, or by its nasality), which are, in essence, a temporal overlap of the acoustic cues to two distinct phonetic segments (Fowler, 1984; Marslen-Wilson, 1989; Warren and Marslen-Wilson, 1987; 1988). Gaskell and Marslen-Wilson (1996) used cross-modal repetition priming to examine the effects of phonological assimilation on lexical access. In one condition, these changes were characteristic of natural assimilation processes (e.g., /b/ for /d/ before a subsequent labial sound /p/, as in '*That was a wickib prank'*); in another condition the same phonological substitutions violated assimilation rules (as in '*That was a wickib game'*). The activation of the underlying base word, *wicked*, by the phonologically changed auditory prime, *wickib*, was reflected in the time taken to make a lexical decision to the base word, which was presented to participants visually at the offset of the prime word. They observed faster reaction times in the first condition compared to the second, suggesting that phonological changes in an unviable context for assimilation disrupt lexical access. This argument was later supported with a phoneme monitoring task (Gaskell and Marslen-Wilson, 1998).

Evidence from studies using cross-spliced stimuli (e.g., switching around the final consonants in *scoop* and *scoot* while preserving the initial segment of the words such that the vowel contains acoustic information consistent with the original word-final consonant) shows that the phonological system of an adult native speaker can use this coarticulation information to its advantage as anticipatory cues to the identity of an upcoming segment: as the listener hears one segment, he will also hear partial cues to the next sound (Marslen-Wilson, 1989). In auditory lexical decision tasks, identity-spliced words were responded to more rapidly than were cross-spliced words (McQueen, Norris, and Cutler, 1999), and cross-spliced words (e.g., *soak*) were responded to more slowly than cross-spliced nonwords (e.g., *shoak*) (Streeter and Nigro, 1979; Whalen, 1982, 1983). In speech gating tasks with stimuli contrasting in place of articulation (*scoop—scoot*), manner of articulation (*spout—spouse*), or voice onset time (*lock—log*), subjects' responses started to diverge well before closure of the alignment point. Moreover, spliced words had later isolation and recognition points than unspliced words (Warren and Marslen-Wilson, 1987, 1988), suggesting that splicing is more disruptive for anticipatory cues to word identity because it brings coarticulatory information into potential conflict with the new incoming phonetic information belonging to a different word. Dahan, Magnuson, Tanenhaus, and Hogan (2001) conducted a visual-world version of a subcategorical mismatch study, where participants' eye movements were monitored as they followed spoken instructions to click on a picture of a named referent (e.g., *net*). On some trials, the referent's name had been spliced with an existing word, on others—a nonword. People were found to fixate on the target picture more slowly when the onset of the target word came from a competitor word (e.g., *neck*) than from a nonword (e.g., *nep*).

The common conclusion about the above studies is that the interfering effect of cross-splicing on lexical access depends on lexical factors, i.e., whether the target word is a word or a nonword, and whether its constituents come from words or non-words). This suggests that subcategorical information (acoustic-phonetic cues resulting from natural phonological processes) actively participates to the ongoing phonemic categorization process and, therefore, contributes to lexical selection processes. Although there exist different accounts of how the phonological system copes with such acoustic-phonetic variability in the signal, such as tolerance-to-mismatch accounts, underspecification accounts, and inferential accounts (for review, see Marslen-Wilson, Nix, and Gaskell, 1995; Gow, 2001; Mitterer, Csépe, and Blomert, 2006), it is agreed that the presence of even a small amount of mismatching information (e.g., violation of a phonological assimilation rule) can be enough to disrupt word recognition.

Overall, evidence points to an active, dynamic and efficient phonological system which is established early in life and which is able to integrate various sources of information—phonetic, featural, segmental and suprasegmental—in a as much detail as necessary, and deliver this information to the lexical system in an immediate and continuous manner. Once this information reaches the lexical level, it is used to restrict the class of possible lexical choices while the word is still unfolding in time. Therefore, we can say that the role of phonology in word recognition is to guide and constrain lexical access and eliminate spurious competition among lexical candidates, thereby allowing for speed and efficiency of word recognition.

2.1.2 Context effects in L1 spoken word recognition

Understanding natural speech necessarily entails not only attending to low-level, phonological information but also to higher-level processes (e.g., lexical processes, syntactic processes, etc.) and their interaction, and therefore engages numerous higher-order factors which critically mediate the analysis of spoken input (Poeppel, Idsardi, and van Wassenhove, 2008). The terms ‘bottom-up’ and ‘top-down’ are sometimes used to distinguish between the processing of information derived from perceptual sources and information derived from higher-order processing levels, respectively. In bottom-up processing, low-level units are progressively included into higher-level ones, while in top-down processing, higher-level units are believed to contribute continuously to the way in which low-level units are processed. For example, in natural language situations, there is often some amount of noise present (several people talking, talker differences, environmental noise, etc.), which can affect the incoming speech signal in the way that some phonetic features can be degraded or lost. Normally, this should not affect speech comprehension and communication in healthy, normal-hearing people listening to their native language. This is because there is much redundancy inherently built in the speech signal to make communication reliable, and the lexical processor is able to combine the degraded information from sound segments and the context and to select the word that represents the best fit, i.e., consistent with both. Because native speakers have robust phonological decoding/encoding strategies, on the one hand, and effective semantic, syntactic, and morphosyntactic processing strategies, on the other hand, this results in efficient and automatic bottom-up and top-down procedures. Moreover, the speed with which speech comprehension happens suggests that information from these different levels of processing must somehow interact, and interact early on and in a fast manner. It has been proposed that the mental lexicon is the "place" where information derived from the sensory level is integrated with higher-level contextual (semantic and syntactic) information (Marslen-Wilson, 1989). While no one disputes the fact that higher-order contextual information plays a crucial role in recognition of spoken words in a sentence, there is a greater controversy about the nature and the precise time course of information integration that exists among psycholinguists. In essence, several approaches are possible here: a) higher-order properties of speech become available very early in the lexical selection process and can "assist" the perceptual analysis of the input at the preselection stage; b) sensory level information receives priority in word identification whereas higher-order cues become available later on in the process of lexical selection and serve a function of amplifying the perceptual cues emerging in the signal; and c) bottom-up analysis of the incoming speech happens incrementally and in parallel with the top-down analysis, and integration of information coming from both sources occurs as the word unfolds in time. Different research methodologies and stimulus materials have been used to support the above claims; that is why findings are not easily generalizable.

The first type of evidence of the higher-order constraints in speech comprehension comes from studies that investigate the question of whether contextual processes can exert an influence on phonetic-acoustic analyses. Such studies usually use acoustically manipulated speech (e.g., masking in noise, phoneme distortion, acoustically synthesized stimuli) in order to quantify contextual effects in speech recognition. For example, Remez, Rubin, Pisoni, and Carrell (1981) used sine-wave speech sentences to demonstrate how their perceived intelligibility depends on the listener's prior knowledge of the linguistic content of the sentence. They argued that the “pop-out” effect (when the sine-wave speech sentence suddenly becomes intelligible) is a top-down process produced by higher-level knowledge and predictions that the brain is making concerning the incoming sounds that can potentially be heard as speech. Another example in support of the direct role of context in spoken word recognition comes from studies using speech in noise (SPIN) paradigm, where speech intelligibility is evaluated in contextually constrained and unconstrained sentences with different signal-to-noise ratios (SNRs). Results from these studies demonstrate that identification of words is more accurate in contextually constrained sentences than in unconstrained sentences with the same SNR (Cervera and González-Alvarez, 2011; Kalikow, Stevens, and Elliot, 1977). One more paradigm in which contextual factors were directly demonstrated to override information derived at a phonetic-phonological level is the phoneme restoration task. Using this task, Warren (1970) observed that, when a noise replaced a certain phoneme in a word embedded in a meaningful sentence, listeners were unaware that the phoneme was missing, and were unable to accurately localize the substituted noise within the sentence. A more recent study by Liederman, Gilbert, McGraw Fisher, Mathews, Frye, and Joshi (2010) modified the standard restoration paradigm by masking the initial phoneme in the target words (which were also minimal pairs, e.g., *peas* vs. *bees*) with 100 milliseconds of pink noise and embedding them in contextually congruent or incongruent sentences. Upon the presentation of the sentence, listeners had to indicate which word they heard. It was found that, in the case of the contextual information incongruent with the target word, the listeners reported hearing the word consistent with the sentence interpretation rather than the target more often (e.g., *bee* instead of *pea)*.

Other studies take a somewhat different approach to examining how the context and the sensory information interact. Rather than overlapping or masking speech sounds with noise, they manipulate the phonological make-up of the word and observe how context contributes to resolving such perceptual ambiguity. For example, Miller, Green, and Schermer (1984), Connine (1987), and Connine, Blasko, and Hall (1991) used a voicing continuum that ranged from a voiced stop to a voiceless stop (e.g., *bath* and *path*). Each token was embedded in sentence contexts semantically biased toward the voiced or the voiceless counterpart, as in "*She needs hot water for the ___"* versus "*She likes to jog along the* ___". Using a phoneme labeling technique, it was found that listeners were more likely to label tokens from the midrange of the voice-voiceless continua as forming a word consistent with the semantic bias—a finding similar to what we have seen in the aforementioned studies. Analogous results were also demonstrated with regard to the syntactic constraints on word identification (Isenberg, Walker, and Ryder, 1980). A continuum of /tə-ðə/ tokens was constructed and embedded in sentences either before the word *go* or *gold* (e.g., "*We tried ___ go/gold*," and *"___ go/gold is essential*"). Listeners were asked to identify the critical tokens as *to* or *the*. It was found that they reported hearing *to* more often in the *go* context and hearing *the* more often in the *gold* context, indicating the effect of syntactic category assignment on word identification in the presence of perceptual ambiguity. Importantly, however, identification of endpoint stimuli (containing phonologically unambiguous tokens) was not systematically influenced by the context in all the above studies. Also of interest is the fact that consistent responses were not faster than inconsistent responses in the case of midrange (ambiguous) tokens, but in the case of endpoint (phonetically unambiguous) tokens, context-consistent responses showed a reaction time advantage compared to context-inconsistent responses (Connine, 1987). In summary, the reaction time pattern for the phonetically ambiguous condition suggests that two lexical items should be equally available and that the semantic context should play a crucial role in the final lexical decision. At the endpoints of the continua, however, the unambiguous phonetic-phonological information contained at the word onset forces the lexical processor to commit to a lexical candidate early on in the process even though the lexical candidate may be anomalous with respect to the sentence context. In this case, reaction time lag in the context-inconsistent condition reflects a later, additional analysis (reanalysis) of the lexical hypothesis, i.e., the context-bias cost.

Because we know (see previous section) that even incomplete auditory input can activate a whole set of lexical candidates, the question of interest then is whether the contextual information can help "preselect" the relevant lexical candidates while the auditory input is still unfolding. The second class of studies we review below asks exactly this question and relies on methodologies examining the timing of such integration.

In the previous section, we introduced the definition of the word's uniqueness point—the point at which the word can be unambiguously chosen by the lexical system over a set of the word's competitors. We also claimed that the larger the number of competing lexical candidates, the more postponed in time is the word's UP. Besides the cohort size, however, sentential context has also been shown to influence a word's UP. Confirming evidence comes from studies showing that words are recognized earlier in utterance contexts than when the same words appear in isolation (Marslen-Wilson and Tyler, 1980), suggesting that, with prior context, the correct lexical candidate is chosen when the sensory input is still ambiguous. For example, in gating tasks (Grosjean, 1980), in which subjects have to identify a word based on progressive presentation of word fragments ("gates"), fewer gates are necessary to identify the word correctly in a highly constraining context as opposed to low constraining context (Craig, Kim, Rhyner, and Chirillo, 1993; Salasoo and Pisoni, 1985; Tyler, 1984; Tyler and Wessels, 1983; 1985). This means that information from higher levels of processing (such as semantic, syntactic, pragmatic) must become available early in the word-recognition process to be combined with incomplete sensory information in order to narrow in on the target word; otherwise there would have been no contextual advantage. Additional evidence in favor of the facilitative role of context on word recognition comes from studies examining lexical ambiguity resolution. We have already mentioned that multiple meanings of a polysemous word get activated when the listener hears the word (e.g., hearing the word *bank* will activate meanings related to *river* or *money*). If an ambiguous word is presented in isolation, the only basis for meaning selection is the dominance of the various meanings (Simpson and Burgess, 1985). Conversely, when an ambiguous word appears in a sentence context, meaning selection is guided by both meaning frequency and contextual constraints, with the contextually appropriate meaning of the ambiguous word activated more strongly than the inappropriate one (Lucas, 1999; Moss and Marslen-

Wilson, 1993; Rayner, Pacht, and Du, 1994; Tabossi, Colombo, and Job, 1987; Tabossi, 1988; Tabossi and Zardon, 1993).

Proponents of the Cohort model explain these findings from the position of the "bottom-up priority" principle (Marslen-Wilson and Tyler, 1980; Marslen-Wilson, 1987; 1989). They reject the concept of contextual preselection and claim that the signal is the primary means by which listeners recover meaning, while context plays a secondary (but nonetheless strong and rapid) role. They argue that contextual information is not used to determine which words are considered for recognition, but is used subsequently to influence later stages of word selection. Sentential information, therefore, has the effect of strongly amplifying the cues already emerging in the signal so that activation of the correct lexical candidate is significantly increased, while activation of the competitors is greatly decreased. The argument is based on several sources of evidence dating back to Swinney (1979). He was the first one to use a cross-modal associative priming task to demonstrate that both meanings of an ambiguous word are activated at the offset of that word regardless of whether the preceding context biased interpretation of the word in one or the other direction, but that shortly after the presentation of the word, only the contextually appropriate meaning remains active. Zwitserlood (1989) continued this line of research and probed different word positions to determine how soon contextual constraints affect the activation of different word candidates in a cross-modal lexical decision task. To test this, she presented cross-modal probes that were associatively related to contextually appropriate and inappropriate words at various positions before and concurrently with the target word. The results show that at the first two probe positions—before the target word and at the onset of the target word—there is no advantage of context for lexical access: response times to related probes following a strongly constraining context and an unrelated control context do not differ compared to unrelated probes. Therefore, responses at these word positions replicate the pattern previously found for words presented in isolation. It is only at the third probe position that the context effect starts to emerge, and only in strongly constraining context as opposed to weakly biasing context condition. At this probe position, the auditory fragment duration is equal to the word's UP established in the prior gating task, suggesting that the auditory signal is already sufficient to differentiate a word from its competitors. A study by Moss and Marslen-Wilson (1993) casts some doubts on Zwitserlood's conclusions. They used a similar paradigm as Zwitserlood, but varied the nature of the semantic relationship between the visual probe and the preceding auditory fragment. They found that closely associated target probes (e.g., *hen*) were primed by the preceding auditory fragment (e.g., *chicken*) even when it was much shorter than the word's UP (e.g., *chi-* instead of *chicken*) compared to the responses to unrelated words. More importantly, the effect was context-independent, i.e., it was present even when the preceding context did not provide any contextual support (e.g., "*When she was looking through the photographs, Tracey found a rather odd one of some chi-*"). In contrast, RTs to target probes that were less strongly related (e.g., *chicken* and *farm*, or *chicken* and *beak*) to the auditory prime produced a different pattern. In this case, priming effect turned out to be context-dependent. For example, only *farm* showed a RT advantage after a sentence about places where chickens might be bred, but only *beak* showed a RT advantage after a sentence about chickens catching worms at early target positions. Such findings suggest that, after all, a strongly constraining context plays a more active, predictive role in lexical access than simply amplifying the outcome of the bottom-up analysis at a post-access decision stage.

Electrophysiological evidence also speaks in favor of the more active role of context in lexical access of spoken words. Several studies have investigated this issue by manipulating perceptual and contextual information in the sentence so that it either matched or mismatched, and showed that event-related brain potentials (ERPs) start to vary even before a word's UP has been reached depending on the contextual appropriateness of that word. For example, Van Petten, Coulson, Rubin, Plante, and Parks (1999) established the UPs for a set of spoken words with the help of the gating technique and used these words as congruous and incongruous sentence completions in an ERP study. For example, a constraining sentence "*It was a pleasant surprise to find that the car repair bill was only seventeen* ___" could end with either a) a cohort congruous word (e.g., *dollars*); b) a rhyme incongruous word (e.g., *scholars*); c) a cohort incongruous word (e.g., *dolphins*); or (d) a fully incongruous word (e.g., *burdens*). The time course of elicited brain-related potentials time-locked to the words' UPs was then examined, and it was found that all three incongruous conditions elicited a significantly larger N400 than the congruous words. However, fully incongruent and rhyme incongruent words elicited an N400 response about 200 milliseconds before the words' UPs, whereas the onset of the N400 response to the cohort incongruous words was delayed by some 200 milliseconds. The results suggested that the onset of the N400 effect reflects the moment at which the acoustic input first diverges from the semantically defined expectation, which the authors refer to as *discrepancy point*. Therefore, they concluded that higher-order semantic processes begin to operate on the partial and incomplete results of perceptual analyses without waiting for the completion of the word identification process. Liu, Shu, and Wei (2006) replicated results of Van Petten and colleagues with an ERP study on spoken word recognition in Chinese. They also observed that, relative to the rhyme incongruous condition, the N400 response in the cohort incongruous condition was delayed by some 200–300 milliseconds. In addition, they demonstrated that the onset of the N400 diverged earlier in highly constraining (200–300 ms window) compared to low constraining (300–400 ms window) sentences, and diverged earlier in the maximal onset mismatch and first-syllable mismatch (200–300 ms) than in the minimal onset mismatch (300–400 ms) condition[1]. Connolly and Phillips (1994) also observed a divergence in the timing of the brain response to cohort incongruent versus plain phonologically incongruent but semantically congruent words. Whereas cohort incongruent words (e.g., *luggage* instead of *luck* in the sentence "*The gambler had a streak of bad luggage*") elicited an N400 response, words with an initial phoneme, different from that of the highest cloze probability word but still semantically plausible (e.g., *glove* instead of *hand* in "*Don caught the ball with his glove.*"), elicited an earlier brain response, which the authors termed the phonological mismatch negativity (PMN) effect (also known as N200). Words that were semantically incongruent and also had an unexpected initial phoneme given the sentence's context (e.g., "*The dog chased our cat up the queer.*") elicited both N200 and N400. The authors claimed that the N200 response reflects context effects in very early acoustic/phonological processing at the juncture of the lexical access stage and the earliest point in the lexical selection stage,

---

[1] Minimal onset mismatch was generated by altering one or two distinctive features of the onset of the first syllable in the congruous words. For example, the phonemes /ʒ/ and /ʃ/ differ only in one feature—voicing. Maximal onset mismatch was created by altering two or more distinctive features of the onset such that the nonword became less similar to the congruous word, e.g., /ʒ/ and /l/.

while the N400 amplitude is modulated by semantic expectancy, and is more dependent on memory, context, and integration processes.

In summary, evidence from electrophysiological studies points to the fact that listeners start predicting the upcoming word candidates based on sentential constraints and even wider discourse constraints (Berkum, Zwitserlood, Brown, and Hagoort, 2003) even before they encounter the word itself. That is why they need less acoustic signal to realize that an unfolding word is not going to fit the context than they need to identify the word in isolation based on its uniqueness point. This means that lexical selection process cannot uniquely depend on the information derived from the sensory input. Instead, collective evidence reviewed in this chapter seems to suggest that the lexical processor has to resolve two sets of constraints—sensory and contextual. Sensory constraints originate at the bottom-up level of processing where the incoming acoustic signal undergoes acoustic-phonetic analysis and its goodness of fit is evaluated. Contextual constraints derive from the goodness of fit of the lexical item to the unfolding context or discourse. Together, these constraints must converge to define a unique intercept—the best fitting lexical candidate, and the speed and the earliness of such convergence suggests some sort of functional parallelism in the speech comprehension system.

2.2 L2 spoken word recognition

2.2.1 The role of phonology: Ambiguity and fuzziness

Unlike native speech perception, which is robust, automatic, and efficient even in non-optimal conditions, L2 speech perception is notoriously problematic even in highly proficient and experienced L2 listeners. Phonetic segments that are phonologically contrastive in the L2, but not in the L1, are often miscategorized and misconstrued by L2 speakers, which leads to inadequate identification of L2 phonemes and renders spoken L2 comprehension difficult (Strange and Shafer, 2008).

Perhaps, the best-known documented evidence in the second language acquisition (SLA) literature is the difficulty that Japanese listeners experience with the perceptual discrimination of the English /r/ and /l/ contrast, as in *rock* versus *lock*, because these are perceived as allophonic variants of the same phoneme in Japanese. Although such difficulty is more pervasive at the beginning stages of language acquisition, advanced, highly functional Japanese speakers also demonstrate a perceptual deficit (Goto, 1971; Lively, Logan, and Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, and Yamada, 1994; McClelland, Thomas, McCandliss, and Fiez, 1999; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura, 1981; Takagi and Mann, 1995).

Another well-described example concerns the perceptual difficulty of the Catalan /e/-/ɛ/ contrast for Spanish speakers. Unlike Spanish, Catalan has two mid vowels of different height, one high /e/ and one low /ɛ/, which are used to distinguish between words, e.g., /te/ “take” and /tɛ/ “tea.” With a variety of research methods and experimental paradigms, Spanish speakers were systematically demonstrated to perform rather poorly on the tasks involving this Catalan contrast compared to the control group of Catalan-dominant speakers (Bosch, Costa, and Sebastián-Gallés, 2000; Navarra, Sebastián-Gallés, and Soto-Faraco, 2005; Pallier, Bosch, and Sebastián-Gallés, 1997; Pallier, Colomé, and Sebastián-Gallés, 2001; Sebastián-Gallés, and Soto-Faraco, 1999; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer, and Díaz, 2006). This evidence is particularly striking because such perceptual discrimination difficulty is observed even in highly proficient Spanish-dominant Spanish-Catalan bilinguals, who received an early and extensive exposure to Catalan and who use both languages in their everyday life. Similar results were obtained by Lee-Ellis, Idsardi, and Phillips (2009) and Lukyanchenko and Gor (2011), who reported a degraded sensitivity to certain phonological contrasts in heritage speakers' less dominant language.

Existing theories of L2 speech perception attempt to explain L2 listeners' phonological difficulties by drawing on the idea of cross-linguistic differences and similarities expressed by Evgeny Polivanov and Lev Shcherba back in late 30-s of the 20th century (Polivanov, 1931; Shcherba, 1939). This view is related to the notion that the native phonological system which is acquired very early in life acts as a 'sieve,' filtering out the phonetic properties in the L2 speech signal that are not relevant for the L1 system (Polivanov, 1931; Trubetzkoy, 1969). Along these lines, Best's Perceptual Assimilation Model (PAM) (Best, 1995), a more recent PAM-L2 model (Best & Tyler, 2007), the Speech Learning Model (Flege, 1993), and the Native Language Magnet (NLM) Model (Kuhl, 1991) likewise claim that perceived similarity between L1 and L2 sounds impacts the way the new L2 sound is assimilated into the shared phonological space. According to the NLM model, L1 sounds act as perceptual magnets 'absorbing' L2 sounds into the same L1 category so that L2 sounds happen to be 'caught' in the perceptual space of the L1 prototype. The PAM distinguishes different patterns of L2 sound discrimination based on the degree of similarity between L1 and L2 sounds. Thus, discrimination is the easiest when contrasting L2 segments are assimilated to separate L1 categories ('two-category pattern'). It deteriorates when the two L2 segments differ in their perceived L1 category goodness ('category goodness pattern'). When phonologically distinctive segments in L2 are perceived as equally good exemplars of a single L1 category ('single category pattern'), discrimination is most difficult (e.g., Japanese listeners mapping English /r/ and /l/ or Spanish listeners mapping Catalan /e/ and /ɛ/ onto a single L1 category).

It is important to know how the native phonological system affects perception of L2 contrasts and to predict relative difficulties in the acquisition of L2 phonology. However, listeners suffer from the difficulties of sound perception only to the extent that their word recognition and speech comprehension is affected (Broersma and Cutler, 2011). For example, if the L1 phonological system sometimes prevents accurate L2 perception (e.g., *rock* versus *lock*), such lack of phonemic dissociation will, naturally, have implications for lexical access and retrieval. However, while the representational aspects of non-native language phonology have been vastly explored in numerous studies, the impact of the phonological deficit on L2 lexical access has been under-investigated (Trofimovich, 2008), except for a number of studies reviewed below.

First evidence comes from studies investigating speech perception in balanced bilinguals. For example, Pallier and colleagues (2001) used a medium-range auditory repetition-priming paradigm to investigate word recognition by Spanish-dominant and Catalan-dominant bilinguals in a lexical decision task and found that the Spanish-dominant listeners, but not the Catalan-dominant listeners, showed a repetition-priming effect for minimal pairs involving a phonological contrast distinctive in Catalan but not in Spanish (e.g., /netə/ "granddaughter" vs. /nɛtə/ "clean"). The authors concluded that the Spanish-dominant participants perceived these minimal pairs as homophones, thus suggesting that phonological ambiguity entailed lexical ambiguity. Using a lexical decision task, Sebastián-Gallés, Echeverría, and Bosch (2005) demonstrated that the

Spanish-dominant Spanish-Catalan bilinguals tended to accept nonwords created by substituting the Catalan /e/-/ɛ/ contrast as words significantly more often than the Catalan-dominant bilinguals and significantly more often than nonwords with the substitution of the control vowel contrast, /i/-/u/, which is common in both languages. In a subsequent study, Sebastián-Gallés and colleagues (2006) corroborated their behavioral findings with electrophysiological evidence (ERP). They found that Catalan-dominant bilinguals and Spanish-dominant bilinguals differed in terms of the elicited error-related negativity (ERN) component, which is normally associated with error detection and response conflict (Falkenstein, Hoormann, Christ, and Hohnsbein, 2000; Yeung, Botvinick, and Cohen, 2004). In particular, Catalan-dominant bilinguals showed ERN differences between their erroneous responses to nonwords and correct responses to words in the /e/-/ɛ/ condition, whereas Spanish-dominant bilinguals showed no differences between the two types of responses. In fact, their correct responses to real words and their incorrect responses to the critical nonwords showed the same degree of uncertainty. The authors argued that they simply failed to notice errors in the critical nonwords most of the time.

Similarly to that in bilinguals, phonological difficulties in late L2 learners were also demonstrated to have consequences for lexical access and retrieval. For example, a series of experiments examined recognition of L2 English spoken words by L1 Dutch listeners (Broersma, 2002; 2005; Broersma and Cutler, 2011; and Díaz, Mitterer, Broersma, and Sebastián-Gallés, 2012). Word recognition was assumed to be contingent on the listeners' ability to discriminate between confusable phonemes in the L2. To illustrate the point in question, Dutch lacks the English /æ/-/ɛ/ vowel distinction, which could lead to the word *flash* being interpreted as *flesh* by Dutch speakers of English. This is exactly what the findings demonstrated. In auditory lexical decision tasks, Dutch listeners accepted nonwords (e.g., *lemp*) as real English words (e.g., *lamp*) more often than English listeners did. In a cross-modal priming task, nonwords extracted from word or phrase contexts (e.g., *lemp* from *eviL EMPire*) led to increased activation of corresponding real words (*lamp*) for Dutch, but again not for native speakers of English. This finding is in agreement with the data from balanced bilinguals and supports the idea that the lack of perceptual discriminability of a particular L2 phonological contrast can cause ambiguity at the lexical level.

We should add a word of caution here that inability to perceive an L2 phoneme distinction is not the only source of L2 phonolexical ambiguity in the L2. In a different study, Broersma and Cutler (2008) showed that although a given L2 contrast may be present in the L1, it can still lead to lexical confusion. For example, in Dutch, consonants can be contrasted on the basis of consonant voicing, except for the cases when they occur in the word final position. It was shown that although Dutch speakers of English can discriminate voiced/voiceless consonants in English quite accurately, they nevertheless tend to accept non-words such as *groof* or *flide* as real English word counterparts *groov*e and *flight*, and show priming effects from these nonwords to real words in a cross-modal priming task. Thus, L1 phonotactic constraints seem to be exerting influence on how L2 sounds are perceived.

Importantly, L2 listeners' lexical-phonological ambiguity is not confined to just difficulties with discrimination between minimal pairs involving particular L2 phonemic contrasts. Evidence from eye-tracking studies demonstrates that words that are distinct overall, can become temporarily indistinguishable. Thus, hearing *panda* activates *pencil* for Dutch listeners (Weber and Cutler, 2004), and *rocket* activates *locker* for Japanese listeners (Cutler, Weber, and Otake, 2006). The results of the Russian-English translation priming study (Cook and Gor, 2012) indicate that phonolexical ambiguity can occur even at a more global level and involve lexical items which differ from each other in more than one phoneme but share a substantial amount of form-related information (e.g., *забота—задание*, /zaˈbota/—/zaˈdanʲija/, “care”—“assignment”; *повесть—новость*, /ˈpovʲistʲ/—/ˈnovastʲ/, “story”—“news”). This finding shows that, because L2 lexical access is compromised by the uncertainty about the form-meaning mappings of the two words with partially overlapping phonology, the difference between the two lexical items can become blurry and one word can easily be substituted for another based on the similarity of the phonological form.

Taken together, evidence from studies of L2 spoken word recognition provides support for the close interaction of phonology and lexical knowledge at different stages of the word recognition process, and illustrates the point that phonological ambiguity created by less precise perception of L2 phonological distinctions can result in lexical ambiguity. It should be borne in mind, however, that distinguishing phonologically ambiguous minimal pairs is probably not the greatest challenge L2 listeners face. After all, there are not so many of them. But having to cope with spurious competition and overall fuzziness of lexical representations as a result of such phonological ambiguity will make word recognition much more difficult, time-costly, and less efficient. For example, when native speakers of English hear the word *rock*, they are able to extract the correct meaning of the word through mapping each sound segment in the word onto a corresponding phoneme and through matching up the resulting phoneme string with the correct lexical candidate. Thus, robust phonological information can effectively contribute to the lexical search and narrow down the list of the possible lexical candidates at early stages of processing (e.g., *rock* will be selected over *lock*, *mock, rack, rob*, etc.). Because L2 phonological representations are very often vague, approximate and lack granularity, this leads to a lot of confusion in the system and fuzziness in bottom-up activation, which in turn, cause spurious activation of irrelevant lexical candidates. In the above example, the word *rock* may not only activate the right candidate *rock*, but also a competing word *lock*, along with the other competitors from their corresponding cohort sets. So even though the L2 listener knows fewer words than the native listener, activation of multiple lexical candidates will increase the overall competition over and above the competition that L1 listeners have to deal with. We have seen that even small perceptual confusions at the phonological level may lead to the activation of "phantom" competitors—the words that are not actually present in the speech signal but are similar to parts of the words that are present (Broersma and Cutler, 2008). In short, a good part of the notorious difficulty and slowness of L2 speech processing could be due to the increased competition of lexical candidates, because, admittedly, the more competition there is and the longer it persists, the more slowly words are recognized (Norris, McQueen, and Cutler, 1995; Vroomen and De Gelder, 1995). This is not to say that increased competition due to "low-resolution" phonological representations necessarily has to disrupt L2 listeners' speech comprehension. After all, in natural speech situations, L2 listeners can exploit other cues to meaning resolution, such as prosodic, pragmatic, contextual, and even visual and gestural. In the next section, we will review the role of sentential context in L2 spoken word comprehension.

2.2.2 Context effects in L2 spoken word recognition

Like in the L1, word recognition in the L2 is never solely determined by phoneme recognition because the auditory signal is weakened and reduced in natural speech (especially in less formal registers) and can sometimes be contaminated with noise. However, unlike in the L1, L2 listeners also have to cope with the consequences of imperfections in phonetic-phonological processing. It is believed that, in the face of insufficient phonetic information, L2 speakers use the same mechanisms for resolving uncertainties in speech comprehension as L1 speakers, e.g., using contextual cues. It should be noted, however, that the role of context has received relatively little attention in the SLA research. Available evidence points to two opposing views on whether L2 speakers favor bottom-up or top-down strategies in speech comprehension. The first type of evidence suggests that nonnative speakers appear to be predominantly bottom-up processors and tend to rely heavily on low-level information, while native speakers use both bottom-up and top-down processing more interactively. Presumably, L2 speakers focus too much attention on identifying sounds and words that they have no time or working memory capacity left for building higher-level units of meaning (Baker, 1985; Berne, 2004; Block, 1992; Randall, 2007). For example, Hassan (2000) writes that learners try to understand every word in a listening text in order to make sense of it, but fail to distinguish the key words that are most important for comprehension. Focusing on every word, therefore, increases processing costs, which impede the comprehender's ability to follow the overall meaning of the text. Vogely (1995) examined L2 learners' performance on listening comprehension and their self-reported use of bottom-up or top-down listening strategies. She found that the strategies the learners reported using the least were top-down strategies, that is, their ability to anticipate, guess, and infer meaning from context. She also found that, as the students' comprehension began to break down, they found themselves resorting to bottom-up listening strategies rather than top-down.

Contrary evidence also exists, which indicates that L2 comprehenders do rely on higher-order information to aid them in compensating for gaps in understanding. Field (2004) cites a classroom-based study that suggests that L2 learners tend to construct a schema relating to the topic of a listening text and to use it to guide their processing of incomplete bottom-up information. The study showed that whenever the learners did not understand certain words in the text, they tended to replace them in their oral productions with similar-sounding words that fit their preconstructed schema. For example, in a text about travel one student substituted the word *map* for *mat* and another one used the word *bridge* instead of *ledge.* Motivated by similar observations, Field conducted an experiment where he intentionally substituted highly predictable words at the end of semantically constraining sentences with similar sounding words (e.g., "*We arrived at the airport on time, then we had to wait two hours for the train,*" where *train* replaced the better-fitting *plane*). The substitute word was always less predictable but nonetheless acceptable in the context. The learners were asked to write down the last word in each sentence. He found that the recovery of the original, better-fitting word ranged from 15% to 62% of the responses, which provided positive evidence for L2 speakers' reliance on higher-order, contextual information for meaning resolution.

Hu (2009) and Hu and Jiang (2011) conducted a series of cross-modal priming tasks to examine the effect of semantic context on word recognition in Chinese L2 speakers of English. Participants were asked to perform a lexical decision task on contextually congruent, neutral or incongruent visual targets preceded by auditorily presented context, e.g., "*The girl mailed the letter without a stamp/sticker/stone*". L2 participants demonstrated facilitation in recognition of context-congruent targets compared to incongruent and neutral targets.

Additional evidence in favor of L2 learners' use of context comes from studies on listening to speech in noise (SPIN), in which researchers consistently find that words in predictable sentence contexts are identified more easily than words spoken in isolation or embedded in unpredictable sentence contexts in all noise levels (Bradlow and Bent, 2002; Cutler, Garcia Lecumberri, and Cooke, 2008; Gor and Lukyanchenko, 2012; Kalikow, Stevens, and Elliott, 1977; Mayo, Florentine, and Buus, 1997). We should bear in mind, however, that although L2 listeners seem to benefit from sentential context in identification of words across all noise conditions, they nevertheless show a smaller contextual advantage compared to native speakers of the language in high noise conditions (e.g., Gor and Lukyanchenko, 2012).

Context effects in L2 processing have also been reported in studies using electrophysiological methods. For example, L2 speakers have been systematically shown to produce an N400 effect in response to semantically incongruent words, although it is qualitatively and quantitatively different from the N400 response observed in native speakers both in reading (e.g., Ardal, Donald, Neuter, Muldrew, and Luce, 1990; Moreno and Kutas, 2005; Ojima, Nakata, and Kakigi, 2005; Proverbio, Cok, and Zani, 2002;

Weber-Fox, Davis, and Cuadrado, 2003) and in listening (e.g., FitzPatrick and Indefrey, 2007; Hahne, 2001; Hahne and Friederici, 2001; Mueller, Hahne, Fujii, and Friederici, 2005). First, L2 groups usually display a delayed peak latency of the N400 response (approximately 20-40 ms) as well as its longer extension (about 400 ms longer) compared to native speakers. Second, the amplitude of the N400 response in L2 speakers is usually smaller, although not always statistically smaller. A delayed peak of the N400 response may reflect a slowdown or decrease in efficiency of semantic processing mechanisms and automatic word identification in the L2, while longer extension of the N400 effect may be indicative of the information integration costs, i.e., L2 participants take longer to integrate the word with context than native listeners. This might be the result of their uncertainty with respect to vocabulary knowledge and use (Hahne, 2001). Additionally, a tendency of the N400 to peak on the left side of the scalp has been observed in bilinguals compared to monolinguals, suggesting that different neural generators might be involved in response to semantic errors in monolinguals and bilinguals. Despite this, most ERP researchers agree that there are more similarities than differences between L1 and L2 speakers in terms of the underlying mechanisms for lexical–semantic processing (for review, see Moreno, Rodríguez-Fornell, and Laine, 2008; Mueller, 2005; 2006).

In contrast, studies that investigate integration of morphosyntactic information in L2 populations provide more controversial evidence. For instance, Hahne (2001) and Hahne and Friederici (2001) tested L1 Russian and L1 Japanese learners of L2 German on syntactic and semantic violations in German sentences. Although both L2 groups exhibited an N400 response to semantic violations, only native speakers of German showed both an early left anterior negativity (ELAN) and a P600 response for syntactic violations involving word category substitutions. Neither L2 group showed an ELAN response; the P600 was reduced for the Russian learners of German, and was absent in the Japanese learners of German. Instead, Japanese speakers showed a greater P600 for the correct sentences compared to the native speakers. Similarly, Weber-Fox and Neville (1996) also included a syntactic condition (phrase structure, specificity and subjacency constraint violations) in their ERP study of Chinese-English bilinguals and found a reduced asymmetry in the early components (ELAN and N400) and an absence of the P600 effect for some of their bilinguals, although their responses to semantic violations differed from those by native speakers only quantitatively. It has been suggested that the lack of early negativity components, such as ELAN and LAN, in L2 speakers as opposed to native speakers indicates that L2 speakers do not employ the same early, highly automatic syntactic processing mechanisms as native speakers. The absence of the P600 effect, on the other hand, is indicative of the L2 speakers' difficulty with late syntactic repair processes for the syntactically incorrect sentences.

Electrophysiological responses in L2 populations have been shown to be strongly modulated by L2 speakers' proficiency and familiarity with the L2 structure (Ardal et al., 1990; Kutas and Kluender, 1994; Moreno and Kutas, 2005; Weber-Fox and Neville, 1996). For example, Hahne (2001) and Hahne and Friederici (2001) point out that their Russian participants who participated in the ERP study provided a larger number of correct grammaticality judgments in the syntactic condition than the Japanese group on the behavioral measure, perhaps because they had greater language proficiency. Moreover, while the syntactic structures like those used in the German test sentences (prepositional phrases) in the study are familiar to native Russian speakers, they do not occur in Japanese. More recent studies on L2 morphosyntactic processing provide corroborating evidence that proficiency plays a crucial role, and that an (E)LAN can be elicited in non-native speakers provided they are very fluent. In the study by Rossi, Gugler, Friederici, and Hahne (2006), high-proficiency L2 learners of German and L2 learners of Italian showed the same ERP components as native speakers of those languages for all syntactic violations. For the word category violation, they displayed both an early anterior negativity (ELAN) and a late P600. For morphosyntactic violations, they showed an anterior negativity (LAN) and a P600. In their comprehensive review of ERP studies on syntactic processing in L2, Steinhauer, White, and Drury (2008) argue that although absence of (E)LANs is, indeed, a typical pattern for late L2 learners (in line with Weber-Fox and Neville, 1996; Hahne, 2001; and Hahne and Friederici, 2001), this pattern holds only for less proficient L2 learners. At low levels of proficiency, the morphosyntactic violations are not yet recognized as such by L2 learners, and so the anomalies are perceived as a lexical problem, hence an N400 effect is observed (Osterhout et al., 2006). With the beginning of grammaticalization and proceduralization of L2 knowledge, the learner begins to identify the structural nature of the problem, and attempts to integrate morphosyntactic information with other sources of information available in the input, which results in a (usually delayed) small P600 in case of the difficulty of such integration due to syntactic violations. As proficiency increases, the P600 amplitude also increases and starts to resemble that of native speakers, whereas at native-like levels of proficiency, L2 speakers display LAN-like components preceding the P600 component, very similar to native speakers.

To summarize, research on higher-order processes in L2 listening comprehension is quite sparse, and available evidence cannot be easily generalized because it comes from studies with very different research orientation (such as instruction-oriented SLA, psycholinguistic studies, studies employing cognitive neuroscience methods). In a nutshell, L2 listeners appear to be relying on the same higher-order processing mechanisms as do native speakers in listening comprehension, but these mechanisms may be not very efficient, delayed in time, or present to a varying degree depending on the learner’s proficiency and the language domain (e.g., syntax or semantics). Thus, while semantic processing suffers the most from the reduced size of L2 speakers’ lexicon and weaker semantic associations, syntactic processing seems to be affected by the lack of grammaticalization and proceduralization of morphosyntactic structures. In any case, proficiency and experience with the L2 play a critical role in determining the success and the extent to which L2 speakers will benefit from higher-level information and its integration with the lower-level information in speech comprehension.

3 The present study

3.1 Introduction

In Chapter 2, we reviewed evidence demonstrating that the phonological system of native speakers and that of L2 speakers are drastically different. Native language phonological behavior is largely driven by innate learning programs and is accomplished very early in life. It is fast, efficient, automatic, and robust. However, such early L1 phonological specialization has detrimental effects on the subsequent abilities of humans to learn the sound system of a new language. A wealth of evidence from research on L2 phonological acquisition and perception demonstrates that even highly proficient L2 speakers experience difficulties with L2 sound perception. This is not to say, however, that L2 speakers are completely “deaf” to certain phonological contrasts, or do not form L2 phonological representations of difficult contrasts at all (cf. Dupoux, Pallier, Sebastián-Gallés, and Mehler, 1997; Dupoux, Peperkamp, and Sebastián-Gallés, 2001; Dupoux, Sebastian-Galles, Navarrete, and Peperkamp, 2008). That would mean a great simplification of the problem they face. Instead, we believe that L2 speakers' phonological representations are “fuzzy” and unstable (i.e., lacking detail and specification) (Cook, 2012; Cutler et al., 2006; Weber and Cutler, 2004). That is why they demonstrate a great individual variability in perceptual sensitivity under different tasks and listening conditions: although they may score within the native range on some perceptual tasks (e.g., categorization tasks), other tasks seem to exert more demands on the perceptual system (e.g., tasks tapping into processes of lexical access and selection) (Díaz et al., 2012). L2 speakers may also have a problem with specific instantiations of a phonological contrast (e.g., in certain word positions) while demonstrating good perceptual discriminability of the same contrast in other phonological environment (Broersma and Cutler, 2008; Lukyanchenko and Gor, 2011). Besides difficulties with certain phonological contrasts, inefficient interpretation of phonological processes, such as assimilation, neutralization, reduction, segmentation, etc., may add to the overall perceptual "fuzziness" and make listening to an L2 a particularly fragile component of language competence. Even highly functional L2 users who experience little effort with reading, writing, or speaking acknowledge that auditory comprehension is cognitively more difficult, less automatic, and is prone to break-downs (Broersma and Cutler, 2011).

3.2 Motivating questions

Because difficulties in L2 auditory comprehension are pervasive but insufficiently accounted for in the SLA literature, it seems necessary to explain when and under what circumstances L2 speech comprehension breaks down and how L2 comprehenders recover from breakdowns. To this end, the following objectives need to be addressed:

1. Identify and describe difficult aspects of L2 phonology, such as contrastive features;
2. Examine what implications they have for the lexical level (word recognition);
3. Quantify how and when phonological difficulties are resolved (or are not resolved) at the sentence level.

3.3 Phonological feature under examination

Although numerous SLA studies have addressed the question of L2 phonological acquisition and have described the difficulties that L2 learners face in acquiring L2 phonologies (see Section 2.2.1 for review), they have focused on a limited number of phonemic contrasts (/r/-/l/, /æ/-/e/, etc.) and language pairings (e.g., L1 Japanese—L2 English, L1 Dutch—L2 English, etc.), with the L2-English being the focal L2-language in many of these studies. Besides, while some of the L2 phonological difficulties have been exhaustively described (e.g., the /r/-/l/ confusion for the Japanese speakers of English), others have been under-examined and many remain unidentified.

The novelty and the main contribution of this dissertation study is that it presents new experimental data on the acquisition and processing of an L2 phonological feature rather than an individual phonological contrast. Although the question about representational primitives has not been uncontroversially resolved, there are linguistic data indicating that phonological features (e.g., distinctive features, Jakobson, Fant, and Halle, 1951) are the smallest blocks of language. These features specify a number of properties of a phoneme, such as place and manner of articulation, voicing, nasality, lip rounding, etc. From a linguistic point of view, a set of abstract hierarchically organized features allows the identification of acoustically variable exemplars of natural classes of speech sounds, and is sufficient to explain the robustness of speech recognition across different conditions like accent/dialect variation, variability in the rate of speech, different contexts and levels of environmental noise (Lahiri and Reetz, 2002). Because phonological features are defined in terms of both articulatory (Halle, 2002) and acoustic properties (Stevens, 2002), they provide the fundamental link between action (articulatory, motoric gestures) and perception (auditory patterns).

This dissertation examines a phonological feature of consonantal palatalization ([± soft]) in the Russian language. Russian presents an almost unique case where the opposition of hard (unpalatalized) and soft (palatalized) consonants permeates almost the entire consonantal system and is used contrastively, e.g., *вес* ("weight", /vʲes/)—*весь* ("the whole of", /vʲesʲ/)[2]. Usually, soft consonants are interpreted as the ones having a secondary articulation (the raising of the middle part of the tongue towards the hard palate) as compared to the corresponding hard ones. Hard consonants can also have additional articulation—velarization (or the raising of the tongue towards the roof of the mouth). In addition, in CV syllables, the feature "flows" at the syllable level as well as at the segmental level, i.e., the process of consonant-vowel accommodation "smears" featural information and distributes it over the syllable. While in English the effect of vowels on consonants is usually greater, the opposite seems to be true for Russian: the consonants are more stable and independent of the vowels, and it is the vowels that accommodate themselves to the consonants through coarticulation (Howie, 2001). Therefore, the most salient cues to the softness of a consonant in a CV syllable in Russian are mainly contained in the formant transitions of the following vowel (Bondarko, 2005; Kochetov, 2002)

Several studies have shown that the distinction between hard and soft consonants in Russian poses perceptual difficulty for English-speaking learners of Russian (Bondarko, 2005; Diehm, 1998). According to the predictions of the speech perception models reviewed in Section 2.2.1, such difficulty can be explained by the fact that English speakers of Russian assimilate Russian hard and soft consonants into a single category because English does not make such a distinction in consonants (with the exception of some cases of allophonic palatalization as a result of consonant

---

[2] "ь" is a special letter that marks softness of word-final consonants in writing; "j" in superscript corresponds to phonological softness in transcription.

accommodation to front vowels and the light variant of the English /l/). Such perceptual difficulty is exacerbated by the fact that the realization of hardness or softness in consonants does not have one single articulatory or acoustic correlate for all consonants in Russian; rather, it depends on the properties of every particular consonant, both place and manner of articulation (Kochetov, 2002).

3.4 Preliminary evidence

In order to measure the level of L2 perceptual difficulty of the hard/soft consonantal contrasts and identify which contrasts and which positions present most difficulty, two AX phonetic discrimination experiments with L2 Russian speakers with intermediate to advanced proficiency (n=10) (reported in Lukyanchenko and Gor, 2011) and advanced to superior proficiency (n=32) (reported in Chrabaszcz and Gor, in press) were previously conducted. Experimental manipulations included the type of consonant, word position, and phonetic environment.

The results demonstrated a significantly reduced sensitivity to the [± soft] consonant contrasts in L2 listeners for word-final positions (relative to the L1 data) (Figure 1). Moreover, perceptual difficulties were not instantiated equally for all contrasts; discrimination of an L2 contrast presented a gradient difficulty, which depended on the phonetic properties of individual consonants (such as place and manner of articulation), phonetic environment, and word position.

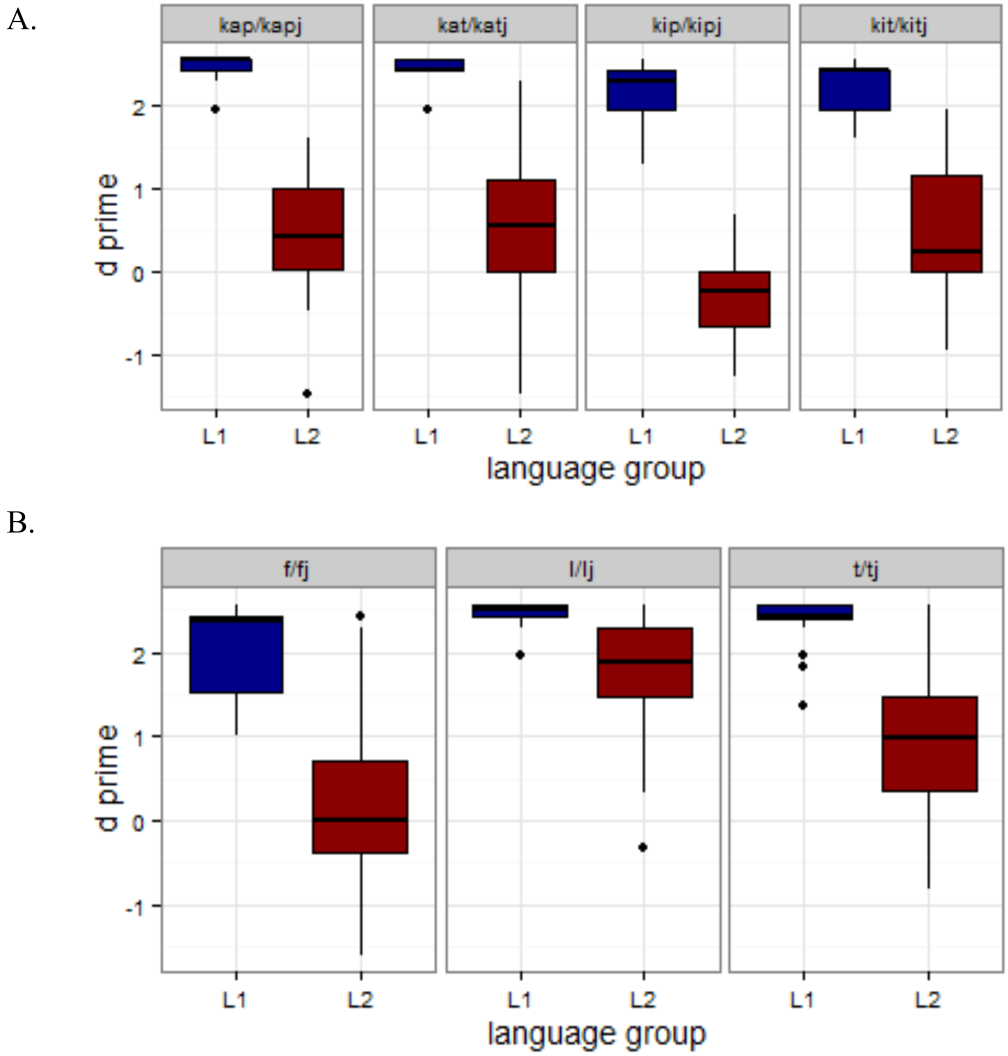


Figure 1. The results of the two pilot AX discrimination tasks with intermediate (A) and advanced (B) L2 speakers of Russian.

After we established under what circumstances [± soft] consonants presented perceptual difficulty for L2 listeners, we set out to examine whether it affected their word discrimination in sentences. In our second pilot experiment (reported in Chrabaszcz and Gor, in press), contextually congruent or incongruent target words differing on the basis of consonantal hardness/softness were embedded in semantically, syntactically, and morphologically constraining sentences, e.g., *Нам сообщили, что поезд* *прибы**л**/*прибы**ль*** (/ˈprʲibyl/—/ˈprʲibylʲ/) *на станцию с большим опозданием* (“We were told that the train has *arrived/*profit* to the station with a big delay”). After each sentence, both congruent and incongruent words (*arrived* and *profit)* appeared on the computer screen, and listeners selected which word they heard in the sentence. The task therefore tested whether L2 listeners could identify the word correctly based on the phonological form of the word, even when it was incongruent with the context, or they were biased by the context to choose the wrong, but contextually appropriate, word. The results demonstrated that, under the circumstances of unfaithful, unstable phonological perception, L2 listeners utilized contextual information for meaning disambiguation, but to a different degree. Morphological and syntactic cues appeared to be more effective than semantic cues in constraining the choice between two phonolexically ambiguous words (Figure 2).

While the findings from the pilot experiment provide some new insights into the problem of phonolexical ambiguity resolution and add to our understanding of which contextual information is most useable in L2 speech comprehension, they do not tell us much about the processing underlying such ambiguity resolution and about the temporal aspects of meaning resolution. Because the interpretation of the accuracy data was based on inferences from the measures taken at the endpoint of processing (i.e., a button press after sentence presentation) rather than continuously, it is difficult to reveal the more dynamic aspects of L2 sentence processing under the given test conditions. It is not possible to tell from the given data whether contextual effects took place during listening, or at the stage of word selection. In the latter case, the listener’s word choice may reflect post-hoc word identification strategies and post-sentence analysis rather than real-time contextual bias.

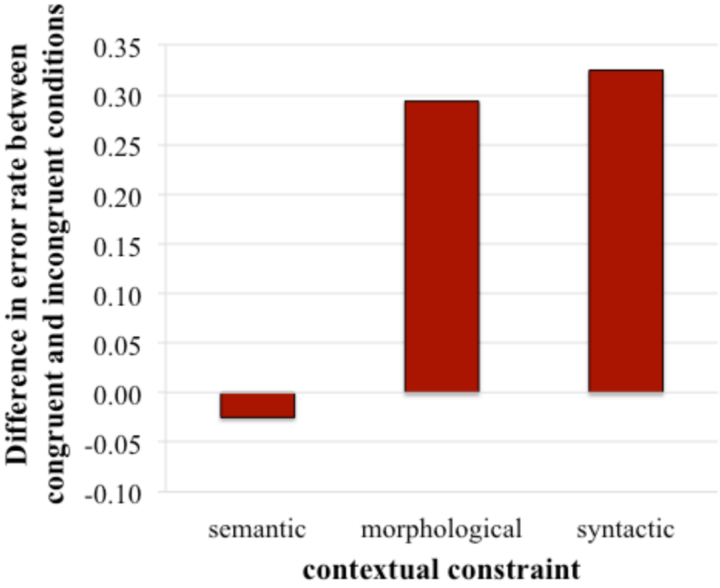


Figure 2. Context bias effect for the L2 group in the pilot listening discrimination task.

3.5 Research questions

Finer-grained, more implicit, online measures capable of capturing the ongoing processes of listening behavior are needed to address the issue of how L2 listeners resolve phonolexical ambiguities in speech comprehension, if they do. Besides quantifying the relative effects of different types of contextual information on phonolexical ambiguity resolution, we need to understand how and when bottom-up information interacts with different types of contextual information to establish an interpretation of the utterance. Several studies have examined how phonological, semantic and syntactic information interact in online listening comprehension (e.g., Connolly and Phillips, 1994; van den Brink et al., 2001; VanPetten et al., 1999), but those studies mainly focused on the population of L1 speakers who listened to sentences in optimal listening conditions. As we have argued in Section 2, data from native speakers do not always generalize to populations of L2 speakers because certain aspects of their processing may be different. For example, L2 learning shows well-attested age and proficiency effects, while higher-order processing mechanisms in the L2 are slower, less efficient, less automatic and more cognitively taxing. These constraints might fundamentally alter or limit the types of information available to L2 speakers during comprehension. On the other hand, those studies that have looked at the use of contextual cues in L2 processing, examined them in isolation from each other, focusing on either syntactic processes or semantic processes, making it very difficult to generalize the findings across studies and different methodological paradigms.

A primary objective of this dissertation research is to understand how information gleaned from the phonetic-acoustic level is brought together with that derived from the larger linguistic context—and in particular, how different kinds of contextual information (semantic, morphological and syntactic) are processed, both neurally and cognitively, and how they shape the interpretation of individual words, especially if they are perceptually ambiguous. In relation to the main research objective, several research questions are proposed:

RQ 1. Does difficulty with discrimination of phonological contrasts lead to phonolexical ambiguity in the L2?

RQ 2. What are the consequences of L2 phonolexical ambiguity for auditory sentence comprehension?

RQ 3. Do L2 listeners utilize contextual information for meaning resolution in online auditory sentence comprehension?

RQ 4. Do L2 listeners utilize different kinds of contextual information, such as semantic, morphological and syntactic, for meaning resolution to the same degree?

RQ 5. What is the time course of integration of phonological information with higher-order contextual information in L2?

RQ 6. How does auditory sentence processing compare in L1 and L2 in terms of the use of contextual information and the temporal aspects of context effects?

Three behavioral studies (Lexical decision task in context, Translation judgment task, and Self-paced listening task) and one electrophysiological (EEG) study will be administered to pinpoint the differential effects and the real-time properties of context on spoken word recognition among L2 Russian speakers in comparison to the native Russian speakers. The outcome of this work has widespread implications, including elucidating the separable mechanisms employed by L1 and L2 listeners and identifying the difficulties that L2 listeners face when processing phonologically ambiguous input. Importantly, the study has pedagogical implications. It will inform educators about the contextual cues which are routinely employed or underused by L2 learners, and this knowledge may promote the development of more effective teaching tools for improving L2 speech comprehension.

4 Sentence-level context effects in L1 and L2 auditory sentence comprehension: Behavioral evidence on processing of phonolexical ambiguity

4.1 Introduction

We espouse a viewpoint according to which spoken word comprehension proceeds in several steps: *activation, selection*, and *integration* (Marslen-Wilson, 1987). First, bottom-up phonetic-acoustic information is received incrementally, analyzed and mapped onto stored representations, activating a set of possible lexical candidates. There can be many lexical candidates competing for selection, but as the spoken word unfolds in time, lexical candidates are dropped or become less activated as soon as they no longer correspond to the unfolding auditory signal. Selection of the intended lexical candidate is said to take place when only one candidate remains that matches the acoustic signal the best, i.e., has the highest level of activation compared to other candidates. The selected word is uniquely characterized by certain phonological, morphological, syntactic and semantic attributes, which need to be integrated into the ongoing sentential or discourse context for the comprehender to arrive at the intended interpretation of the utterance.

As a rule, comprehenders make use of all available linguistic cues to build expectations for particular items or item features; thus, speech comprehension is said to be anticipatory, or predictive, to allow for the pre-activation of those items or features. For example, following a sequence of words such as ‘‘*I like my coffee with cream and…*’’ there is a high expectation for a specific lexical item, “sugar”, as well as a syntactic category of a noun and a morphological template. People are sensitive to the contextually arising expectations at each word, but there may be a difference in how different expectations operate. Semantic information is built up incrementally over the course of a sentence or an utterance to facilitate word recognition. For example, words are recognized progressively faster the later they appear in the sentence (Marslen-Wilson & Tyler, 1980), and the N400 effect is reduced over the course of the sentence (Kutas, Van Petten, & Besson, 1988; Van Petten & Kutas, 1991). Syntactic context information also affects word recognition, but its influence seems to be more localized (Gibson, 2006). For example, although word recognition is facilitated by syntactic context, it is local phrase structure and not global syntactic structure that drives these effects (MacDonald, Pearlmutter, & Seidenberg, 1994; Marslen-Wilson & Tyler, 1980; Trueswell, Tanenhaus, & Kello; 1993; Tyler & Warren, 1987).

As we have argued in previous chapters, native speakers have effective semantic, syntactic and morphological processing strategies. They also have more experience with the structural properties and distributional patterns of words, phrases, and sentences, as well as with the socio-cultural context (schema). This allows them to take advantage of the higher-order information and build linguistic predictions to speed computation of incoming words through pre-activation and preprocessing as well as to integrate discrete information derived at different levels of analysis into higher-order structures in a rapid, continuous manner, even when they have to comprehend speech in non-optimal listening conditions (i.e., noisy or ambiguous speech input).

With regard to L2 speakers, it has been proposed that they rely on similar cognitive and cortical mechanisms for speech comprehension, but these mechanisms are heavily mediated by language proficiency and are often slower, less automatic, idiosyncratic, and lacking precision even in advanced L2 speakers. Although the cognitive system mediating the semantic-conceptual level is believed to be common across L1 and L2 (Kroll and Stewart, 1994), L2 morphosyntactic knowledge acquired after puberty is represented rather differently from that of L1 (Johnson and Newport, 1989). Clahsen and Felser (2006a, 2006b) claim that even though the basic architecture of the processing system is the same in the L1 and L2, shallow morphosyntactic parsing predominates in L2 processing. According to their influential shallow structure hypothesis (SSH), the representations which adult L2 learners compute during processing contain less morphosyntactic detail than those of child and adult native speakers and lack complex hierarchical structure and abstract, configurationally determined elements (Felser, Roberts, Marinis, and Gross, 2003; Papadopoulou and Clahsen, 2003). On this assumption, L2 grammar does not provide the kind of morphosyntactic information required to process nonlocal grammatical phenomena in native-like ways.

Support for such view comes from numerous studies on morphosyntactic acquisition and processing by L2 speakers. For example, longitudinal studies show that L2 speakers continue to have difficulty in accurately using grammatical morphemes in spontaneous speech despite an extended period of language exposure and practice (e.g., Jia, 2003; Lardiere, 1998). From a psycholinguistic perspective, Jiang (2004; 2007; 2011) investigated comprehension of morphological agreement by L2 speakers using a self-paced reading task. Across several experiments, he observed that L2 speakers, unlike L1 speakers, were not sensitive to grammatical violations (e.g., plural –s marking) when a similar grammatical element was not instantiated in the learners' L1. Mecartty (2000) examined the relationship between lexical and grammatical knowledge in L2 listening comprehension and found that although both grammatical knowledge and lexical knowledge were significantly correlated with listening comprehension, only lexical knowledge explained a significant proportion of the variance. Several electrophysiological studies also provide corroborating evidence that L2 comprehenders are more sensitive to semantic as opposed to morphosyntactic expectations in sentence processing. For instance, in Hahne (2001) and Hahne and Friederici (2001), L1 Russian and L1 Japanese comprehenders of L2 German demonstrated an N400 response to semantic violations similarly to the native speakers, but neither L2 group showed an ELAN response, and only Russian group showed a reduced P600 while it was completely absent in the Japanese group, suggesting difficulty with syntactic processes.

Based on similar evidence, the prevailing view in the SLA literature holds that L2 comprehension relies primarily on semantic and pragmatic heuristics coupled with lexical semantic information and that morphosyntactic and inflectional information is generally underused. However, studies that directly compare semantic and morphosyntactic L2 processing within the same experimental set-up are not many, and the picture is far from being complete. For example, it is not clear whether L2 listeners will benefit from semantic and morphosyntactic contextual cues to a similar extent to disambiguate the identity of words during speech comprehension. If they are more sensitive to the lexical-semantic content of the utterance (as some of the literature suggests), they should be relying on semantic cues to process phonolexically ambiguous words. If, however, they pay more attention to morphosyntactic cues, those should prevail in word disambiguation. Let us consider an example of hypothetical phonolexical ambiguity created by the confusion between /r/ and /l/ sounds. Such ambiguity may be resolved at sentence level with the help of lexical-semantic context, in which the word *lock*, for example, has a very low cloze probability (1b) in comparison with the word *rock* (1a).

(1) a. I climbed a **rock** for the first time in my life.

b. I climbed a ***lock** for the first time in my life.

Similarly, syntactic and morphological information can help disambiguate meaning in phonologically ambiguous contexts, as in (2), where a verb is expected to occur after the auxiliary *didn't* (2a), but not an adverb (2b).

(2) a. He didn't **arrive** until noon the next day.

b. He didn't ***alive** until noon the next day.

Since no previous study has investigated the effects of different types of contextual information on phonolexical ambiguity resolution within the same experimental set-up, the present set of studies takes on an exploratory research goal to examine how different contextual cues (semantic, morphological, and syntactic) contribute to the identification of phonologically ambiguous words in the L2.

4.2 Experiment 1: Lexical decision task in context

The main objective of the lexical decision task (LDT) was to investigate context effects (semantic, morphological, syntactic) on spoken word recognition in L2 and L1 listeners. These effects were compared in contextually constraining and unconstraining sentences. For the L2 group, the effects were examined under two conditions: when lexical access and selection were hypothesized to be i) perceptually unimpaired, or ii) perceptually impaired due to the difficulty of discrimination of the phonological contrast involved in meaning differentiation of the two words.

The design of the experiment was based on the assumption that context effects are incremental and predictive. In contextually constraining sentences, by the time the listeners encounter the target word, they are expected to have pre-activated lexical candidates that are consistent with the context. By the time they hear the first sounds of the target word, they are expected to verify their lexical hypothesis, deactivate irrelevant competitors and make a selection in favor of the most desirable candidate. Thus, constraining context should have a facilitative effect (faster response latencies) on the recognition of the context-congruent target. If the target word is incongruent with the listener's expectations, an inhibitory effect (longer response latencies) will ensue.

For the L2 listeners, the predictions were as follows. If they do not experience perceptual difficulty with the target words, they were expected—similarly to the L1 listeners—to show a facilitative effect for context-congruent targets and an inhibitory effect of context-incongruent targets in constraining sentences. Accordingly, error rate was expected to increase for incongruent targets. On the other hand, if L2 listeners experience phonolexical ambiguity with the target words, their word recognition latencies and error rate should not differ for congruent and incongruent targets.

4.2.1 Participants

L1 group included 24 native speakers of Russian (mean age 32, range 23-58; 20 females). Most of them were college graduates, six participants were graduate students, and one had a doctorate degree. L2 group included 34 American speakers of Russian as a second language (mean age 29.5, range 21-50; 20 females). Eighteen participants had a college-level degree, the remaining participants were graduate students or had a graduate degree. For all of them English was their first and dominant language, while Russian was their second strongest language (mean age of onset of acquisition was 17.67 years old).

All L2 speakers were screened for the study based on their language proficiency. Prior to the experiment, they were asked to fill out a language background questionnaire about their language learning experience, rate their language proficiency in different linguistic domains on a scale from 1 (minimum) to 10 (maximum), and complete a 25-item proficiency cloze test. Their average score on the cloze test was 21.74 out of the maximum of 25 (Table 1). Twenty-seven out of the 34 participants reported having taken the ACTFL Oral Proficiency Interview (OPI), a widely recognized language proficiency test. Eight of these people had received a score of 2+ on the Interagency Language Roundtable (ILR) scale (Advanced High on the ACTFL scale); sixteen people—a score of 3 (Superior); three people—a score of 3+ (Superior)[3].

Table 1. Linguistic profile of L2 participants in the behavioral experiments.

	Mean	SD
Age when started learning Russian	17.67	2.79
Length of living in Russia (years)	2.72	2.28
Formal instruction in Russian (years)	6.03	1.56
Self-rated pronunciation	7.15	1.46
Self-rated oral proficiency	7.03	1.09
Self-rated listening proficiency	7.76	1.13
Self-rated reading proficiency	7.56	1.35
Self-rated writing proficiency	6.65	1.45
Self-rated knowledge of grammar	7.24	1.30
Cloze test (Proficiency measure)	21.74	1.80

4.2.2 Design and materials

The stimulus materials consisted of four 240-item sets, which were counterbalanced across four presentation lists such that no participant saw the same item more than once. Items in each set were manipulated across several parameters: *context*

---

[3] ILR scale is a standard proficiency rating scale (from 0 to 5) for language proficiency developed by the Interagency Language Roundtable.

type (constraining, neutral), *condition* (semantic, syntactic, morphological), *target* type (congruent, confusable, unrelated and nonce targets), and *block* type (critical, control, filler).

Context manipulation specified the relationship between the target and the pre-target carrier sentence, where the carrier sentence could be either contextually constraining (n=120 per list) or neutral (n=120 per list). In constraining sentences, the pre-target context created a semantic or a structural bias in favor of a certain expectation for the sentence-final word (target), e.g., *Дедушка достал старинную книгу из шкафа и аккуратно сдул с нее* *<u>пыль</u>* ("Grandfather took an ancient book from the bookcase and blew off the <u>dust</u>"), where the target word *пыль* ("dust") is semantically highly predictable. In neutral carrier sentences, no expectations for the target word were created. They always started with the sentence *Сейчас вы услышите слово…* ("Now you will hear the word…") followed by a target, e.g., *пыль* ("dust"). The neutral sentence, therefore, served as the baseline against which the effectiveness of contextual constraints on word recognition was assessed.

Besides creating a semantic contextual bias, carrier sentences could also be syntactically and morphologically constraining. For example, the target-preceding context in the following sentence biases the listener's expectations in favor of a certain syntactic category, i.e. a verb: *Взволнованный солдат сообщил, что генерал только что* *<u>прибыл</u>* ("An excited soldier announced that the general just <u>arrived</u>"), although the semantic content of the target can potentially vary (e.g., *ate, left*, etc.). Similarly, target-preceding context can constrain the target word morphologically, i.e., create an expectation for a certain morphological form. In the following sentence, *После* *вчерашней ссоры, Роман не хочет говорить* ("After the yesterday's quarrel, Roman does not want to talk"), an infinitive form and not any other form of the verb is expected[4]. As in the syntactic condition, the semantic content of the target word can vary (e.g., *to eat, leave,* etc.).

At the stage of the creation of the experimental sentences, a cloze test was administered, where 18 native speakers of Russian were asked to supply the missing last words in the test sentences. The sentences were then recalibrated based on the cloze test score such that the target words in the semantic condition were highly semantically predictable (*M* = 72.2%, *SD* = 28.8), in the morphological condition—highly morphologically predictable (*M* = 94.1%, *SD* = 15.6), but not highly semantically predictable (*M* = 19.9%, *SD* = 17.2), and in the syntactic condition—highly syntactically predictable (*M* = 93.4%, *SD* = 12.4), but not highly semantically predictable (*M* = 24.3%, *SD* = 22.7). Overall, out of a total of 120 constraining sentences there were 32 sentences in each condition (i.e., semantic, syntactic, morphological) per list (the remaining 24 items were fillers, 120-(32*3) = 24).

Target words always occurred in the word-final position and could be either real Russian words or nonce words. Real words could either constitute a logical and grammatical ending of a sentence (*congruent* targets) or be inconsistent with the preceding context (incongruent targets). Incongruent targets were of two types, *confusable* and *unrelated*. Confusable targets were phonologically similar to the congruent targets except for the realization of the word-final phoneme. The same words

---

[4] Although these types of expectations are not purely morphological in any given context, but morphosyntactic, because they also create expectations for a certain word category, here we refer to them as morphological since the target words in this condition are contrasted on the basis of the morphological form only, e.g., *говорить* (speak$_{\text{INFINITIVE}}$)—*говорит*(speak$_{\text{PRES/3rd/SING}}$).

were used as congruent and confusable targets, but the condition in which they occurred was balanced across the lists. Unrelated targets were also incongruent with the preceding sentential context but were not phonologically related to the highly predictable, congruent targets. Nonce words in the experimental conditions were created such that they had an initial phonological overlap with the congruent and confusable targets and differed only in the final consonant. For example:

Congruent target: Mike has two siblings, an older brother and a younger *sister*.
Confusable target: Mike has two siblings, an older brother and a younger *system*.
Unrelated target: Mike has two siblings, an older brother and a younger *object*.
Nonce target: Mike has two siblings, an older brother and a younger *sisteb*.

Congruent, confusable and unrelated targets were balanced according to word length, lemma and surface frequency as best as possible given the materials design constraints (see Table 2).

Importantly, because the primary goal of the experiment was to compare the effects of different contextual constraints on resolution of phonolexical ambiguity in auditory word recognition, all test items were divided into two matching experimental blocks—*critical* and *control*. An additional, filler, block (n = 48 items) was added to balance the ratio of words to nonce words. All target words in the filler block were nonce words that complied with the Russian phonotactic rules and included a root manipulation.

Table 2. Materials design and the targets' properties in the lexical decision task.

Block	Condition	Target type	Example	Translation	Number of phonemes		Lemma frequency		Surface frequency	
					M	*SD*	*M*	*SD*	*M*	*SD*
Critical	Semantic	Congr	мать	mother	3.5	*0.5*	66.8	*93.3*	31.9	*52.6*
		Conf	мат	checkmate	3.5	*0.5*	66.8	*93.3*	31.9	*52.6*
		Unrel	газ	gas	3.5	*0.5*	86.3	*46.7*	25.9	*12.6*
		Nonce	маф	na	na	*na*	na	*na*	na	*na*
	Syntactic	Congr	брать	to take	4.1	*1.1*	260.4	*428.6*	58.6	*111.9*
		Conf	брат	brother	4.1	*1.1*	260.4	*428 6*	58.6	*111.9*
		Unrel	вниз	downward	4.1	*1.1*	73.7	*69.1*	73.5	*69.1*
		Nonce	брам	na	na	*na*	na	*na*	na	*na*
	Morphological	Congr	говорить	to speak	6.2	*0.5*	361	*690.6*	361	*128.2*
		Conf	говорит	speaks	6.2	*0.5*	361	*690.6*	361	*128.2*
		Unrel	говорим	we speak	6.2	*0.5*	361	*714.8*	4.4	*7.1*
		Nonce	говорик	na	na	*na*	na	*na*	na	*na*
Control	Semantic	Congr	храм	temple	3.6	*0.5*	88	*62.5*	25.4	*19.9*
		Conf	храп	a snore	3.6	*0.5*	88	*62.5*	25.4	*19.9*
		Unrel	долг	debt	3.6	*0.5*	68.1	*28.3*	22	*9.6*
		Nonce	храк	na	na	*na*	na	*Na*	na	*na*
	Syntactic	Congr	жир	grease	3.6	*0.7*	202.9	*298*	35.6	*40.4*
		Conf	жил	lived	3.6	*0.7*	202.9	*298*	35.6	*40.4*
		Unrel	зря	in vain	3.7	*0.7*	61.1	*37.7*	61	*37.7*
		Nonce	жих	na	na	*na*	na	*na*	na	*na*
	Morphological	Congr	любим	we love	6.5	*1.5*	918.1	*686.1*	21.6	*20.3*
		Conf	любишь	you love	6.5	*1.5*	918.1	*686.1*	21.6	*20.3*
		Unrel	любит	loves	6.5	*1.5*	918.1	*710.2*	20.6	*19.7*
		Nonce	любик	na	na	*na*	na	*na*	na	*na*

Note: Congr = congruent, Conf = confusable, Unrel = unrelated

In the critical block, the congruent and confusable targets were distinguished on the basis of the consonant hardness or softness, [± soft], in the word-final position, e.g., *мат* ("checkmate", /mɑt/)—*мать* ("mother", /mɑt^{j}/). The target minimal pairs did not always share the same orthographic representation; however, they were phonologically the same phonologically except for the final consonant (e.g., *балет* ("ballet", /bɑˈlʲɛt/)—*болеть* ("to be sick", /bɑˈlʲɛt^{j}/)). As has been explained in the previous sections, the [±

soft] distinction is phonological in the Russian language, i.e., it can change the meaning of a word. Our earlier study (Lukyanchenko and Gor, 2011) showed that the [± soft] contrasts occurring in the word-final position are more difficult for L2 speakers than those occurring word-initially. That is why all congruent/confusable targets were chosen in such a way that the meaning of a word could be changed by substituting the hard consonant at the end of the word with a soft consonant, and vice versa. In the semantic condition, the [± soft] consonant contrast distinguished two nouns in the Nominative/Accusative form: *мат* ("checkmate", /mɑt/)—*мать* ("mother", /mɑtʲ/). Unrelated targets were also nouns in the Nominative/Accusative form. In the syntactic condition, the phonological distinction marked different parts of speech, e.g., a verb and a noun, as in *брат* ("brother", /brɑt/)—*брать* ("to take", /brɑtʲ/). Unrelated targets also belonged to a different (and context-incongruent) part of speech, an adverb. In the morphological condition, two verbal forms were contrasted, a verb infinitive and a 3rd person singular form in the nonpast tense: *говорит* ("speaks", /gɑvɑˈrʲit/)—*говорить* ("to speak", /gɑvɑˈrʲitʲ/). The unrelated target was a present tense form in the 2nd person plural, *говорим* ("we speak", /gɑvɑˈrʲim).

The control block was similar to the critical block in every aspect except that the congruent and the confusable targets (also minimal pairs) differed on the basis of phonological contrasts common to both Russian and English and did not pose any discrimination difficulty for L2 speakers of Russian. As in the critical block, the semantic condition contrasted nouns in the Nominative/Accusative case: *суд* ("court", /sut/)—*суп* ("soup", /sup/) (/t/ and /p/ are both voiceless stops but they are easily differentiated by the place of articulation in both Russian and English). In the syntactic condition, the minimal pairs belonged to different parts of speech, e.g., nouns vs. verbs vs. adverbs. The morphological condition included a comparison of verbs in the 2nd person plural, 2nd person singular, and 3rd person singular in the present tense. For a full list of items used in this task, see Appendix A.

All sentences were recorded by a native speaker of Russian using a normal speech rate. Target words were spliced out of the recorded sentences and pasted into congruent and incongruent sentences at the end of the sentences such that they did not differ physically and acoustically across the conditions. There were a total of 144 words and 96 nonwords per each list. In order to ensure that the participants are attending to the pre-target context, comprehension of the sentences was evaluated with occasional comprehension questions (n=24) following congruent trials. Eight practice sentences were added at the beginning for task familiarization.

4.2.3 Procedure

The experiment was delivered with the remote DMDX software (Forster and Forster, 2003). Participants were randomly assigned to one of the four presentation lists. Stimuli in each list were presented in a semi-random order such that the sentences from the same condition did not occur adjacently. The participants were instructed to listen to the sentences presented through headphones and judge whether the last word (the target) in the sentences is a real Russian word or not. The target was separated from the rest of the sentence by a 500 ms interval and was marked by an appearance of a fixation cross on the screen. Participants were asked to respond as soon as they saw the cross, but not earlier. After the response, feedback and reaction time latency were briefly displayed on the screen, after which a new sentence started playing automatically. If participants did not respond within 8 seconds, presentation moved on to the next sentence. On some trials, sentences were followed by Yes/No comprehension questions after the lexical decision was made. The total duration of the experiment was about 50-60 minutes.

4.2.4 Results

In order to make sure that the participants attended to the pre-target part of the sentence, their error rate in response to the comprehension questions was analyzed. L1 listeners made 3.4% errors (*SE* = 0.9%), L2 group made 8.2% errors (*SE* = 0.9%); accordingly, the data from all participants were retained for further analyses. Next, participants' word recognition performance was examined. It was characterized by two outcome variables: error rate (ER) and reaction time (RT). All participants' mean error rate and reaction time data are presented in Appendix B.

Only RTs to correct responses were included in the RT analysis, which resulted in a 7% data rejection. RT and ER data were fed into two mixed-design ANOVAs (for RT and for ER) with language (2 levels: L1 or L2), context (2 levels: constraining or neutral), block (2 levels: critical or control, filler block was not included in the analysis), condition (3 levels: semantic, morphological or syntactic), and target (4 levels: congruent, confusable, unrelated or nonce) as independent variables. All significant effects and interactions of the ANOVAs are presented in Table 3.

Table 3. ANOVA output for reaction time and error rate data in the lexical decision task.

		Reaction time		Error rate	
Effects and interactions	Df	*F* test	*p* value	*F* test	*p* value
language	1	6.56	< 0.05	15.62	< 0.001
context	1	15.07	< 0.001	144.579	< 0.001
block	1	4.30	< 0.05	26.335	< 0.001
condition	2	73.50	< 0.0001	74.865	< 0.001

target	3	103.01	< 0.0001	113.01	< 0.001
language:context	1	0.73	0.39	2.844	0.09
language:block	1	16.74	< 0.0001	7.368	< 0.01
context:block	1	1.42	0.23	15.469	< 0.001
language:condition	2	3.47	< 0.05	6.667	< 0.01
context:condition	2	1.91	0.15	10.052	< 0.001
block:condition	2	2.52	0.08	0.679	0.51
language:target	3	7.27	< 0.0001	5.942	< 0.0001
context:target	2	26.35	< 0.0001	43.22	< 0.0001
block:target	3	2.54	0.05	27.164	< 0.0001
condition:target	6	8.25	< 0.0001	26.84	< 0.0001
language:context:block	1	1.02	0.31	2.419	0.12
language:context:condition	2	0.08	0.93	0.976	0.38
language:block:condition	2	1.61	0.20	0.859	0.42
context:block:condition	2	9.64	< 0.0001	1.027	0.36
language:context:target	2	0.10	0.91	0.7	0.50
language:block:target	3	5.89	< 0.001	12.893	< 0.0001
context:block:target	2	2.03	0.13	1.454	0.23
language:condition:target	6	0.89	0.50	5.879	< 0.0001
context:condition:target	4	1.13	0.34	4.391	< 0.001
block:condition:target	6	1.91	0.08	6.433	< 0.0001
language:context:block:condition	2	7.26	< 0.001	0.074	0.93
language:context:block:target	2	1.32	0.27	0.173	0.84
language:context:condition:target	4	1.18	0.32	1.676	0.15
language:block:condition:target	6	0.53	0.78	3.242	< 0.001
context:block:condition:target	4	0.52	0.72	0.209	0.93
language:context:block:condition:target	4	0.99	0.41	0.688	0.60

Separate ANOVAs were run for the two experimental blocks in order to examine whether constraining sentential context facilitated word recognition in L1 and L2 listeners. The results of the omnibus F tests together with p and η^2 values are presented in Table 4. According to our predictions, context should facilitate recognition of congruent targets in contextually constraining sentences compared to incongruent targets (both confusable and unrelated) in the L1 group in both critical and control blocks. For the L2 listeners, we predicted a similar pattern of context effects in the control block, but we expected to see additional facilitation for incongruent confusable targets in the critical block. Post-hoc Tukey HSD comparisons support our predictions.

Table 4. Separate ANOVA outputs for reaction time and error rate data for the constraining condition in the lexical decision task.

ERROR RATE							
		L1: critical			L1: control		
	Df	*F* test	*p* value	η^2	*F* test	*p* value	η^2
condition	2	11.440	< 0.001	0.020	11.76	< 0.001	0.02
target	3	23.082	< 0.001	0.057	30.85	< 0.001	0.08
condition:target	6	8.355	< 0.001	0.042	7.70	< 0.001	0.04
Residuals	1140						
		L2: critical			L2: control		
	Df	*F* test	*p* value	η^2	*F* test	*p* value	η^2
condition	2	11.77	< 0.001	0.01	16.37	< 0.001	0.02
target	3	14.35	< 0.001	0.03	57.11	< 0.001	0.10
condition:target	6	7.56	< 0.001	0.03	5.31	< 0.001	0.02
Residuals	1620						

REACTION TIME							
		L1: critical			L1: control		
	Df	*F* test	*p* value	η^2	*F* test	*p* value	η^2
condition	2	2.56	0.08	0.01	24.89	< 0.001	0.05
target	3	15.34	< 0.001	0.04	14.45	< 0.001	0.04
condition:target	6	3.36	< 0.01	0.02	1.35	0.23	0.01
Residuals	1037						
		L2: critical			L2: control		
	Df	*F* test	*p* value	η^2	*F* test	*p* value	η^2
condition	2	6.77	< 0.01	0.01	9.15	< 0.001	0.01
target	3	12.88	< 0.001	0.03	20.53	< 0.001	0.04
condition:target	6	3.61	< 0.01	0.01	1.76	0.104	0.01
Residuals	1490						

For the error rate data in the control block, a higher word recognition error rate was observed for both types of incongruent (confusable and unrelated) targets compared to congruent targets for both language groups, especially for the syntactically constrained sentences (L1: $p < 0.001$, L2: $p < 0.001$). Comparisons of response latencies almost mirror the error rate results. For the L2 group, significantly longer reaction times were observed when the listeners encountered incongruent targets (confusable and unrelated) compared to when they had to judge congruent targets ($p < 0.001$) suggesting an inhibitory effect. The RT difference between confusable and unrelated targets was not significant ($p = 0.42$). L1 group showed inhibition for incongruent targets in the semantic and syntactic conditions compared to the congruent targets, but no significant RT difference was observed in the morphological condition. In the critical block, L1 participants made significantly more errors judging incongruent targets (both confusable and unrelated) compared to congruent targets ($p < 0.001$), especially in the syntactic condition. L2 participants showed an overall significant error rate difference between congruent and unrelated targets ($p < 0.001$), but not between congruent and confusable targets ($p =0.106$). In terms of reaction time, L1 listeners responded significantly more slowly to confusable and unrelated targets compared to congruent targets ($p < 0.001$), suggesting inhibition. The difference between the two types of incongruent targets did not reach significance ($p = 0.44$). In conformity with our predictions, L2 listeners did not show an RT difference between congruent and confusable targets ($p = 0.74$) across all context conditions, but showed an inhibition effect for unrelated targets compared to congruent targets ($p < 0.001$) in the morphological and the syntactic conditions. Figures 3 and 4 graphically present the ER and RT mean data, respectively.

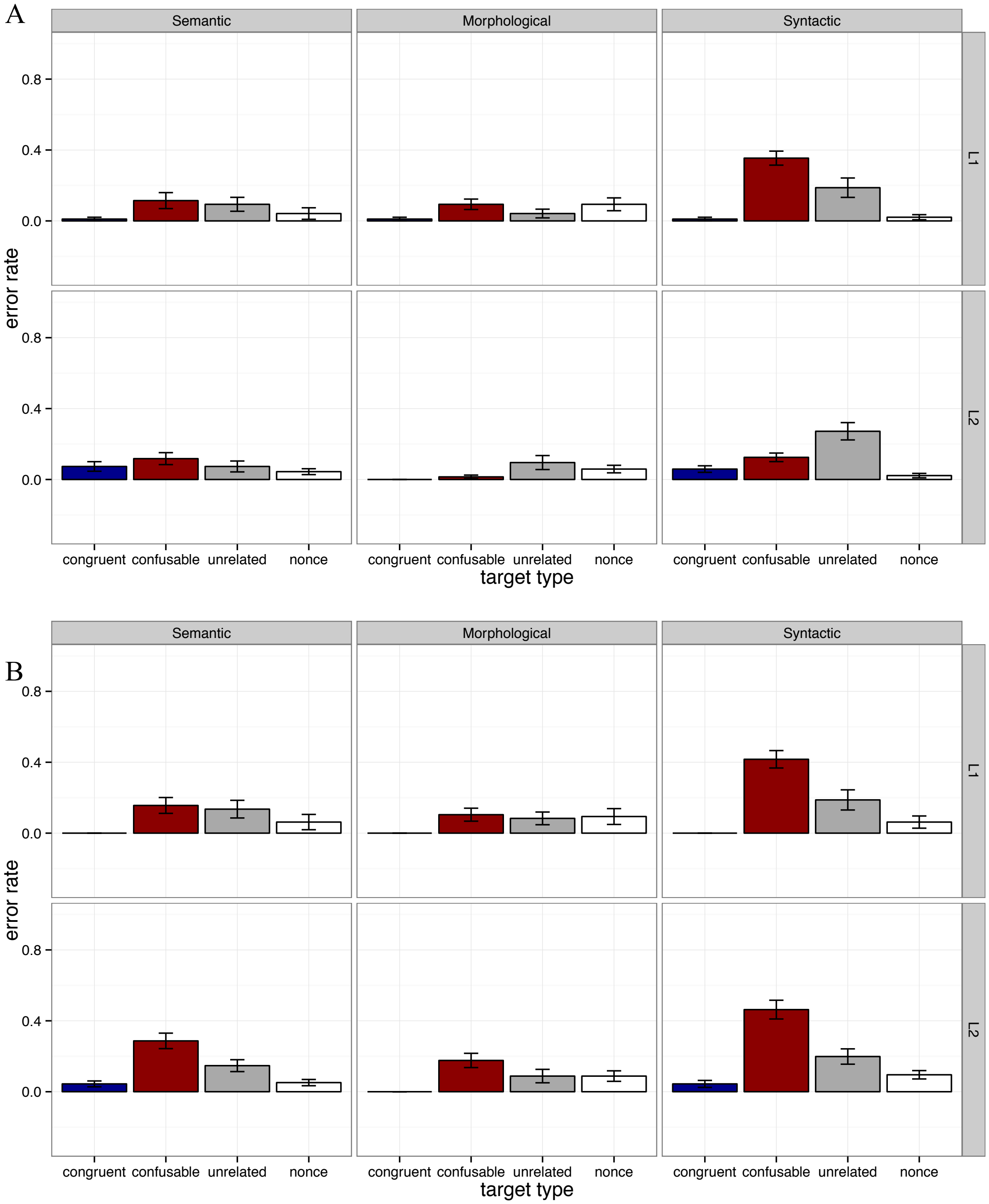


Figure 3. Participants’ word recognition error rate in contextually constraining sentences in (A) critical and (B) control conditions of the lexical decision task.

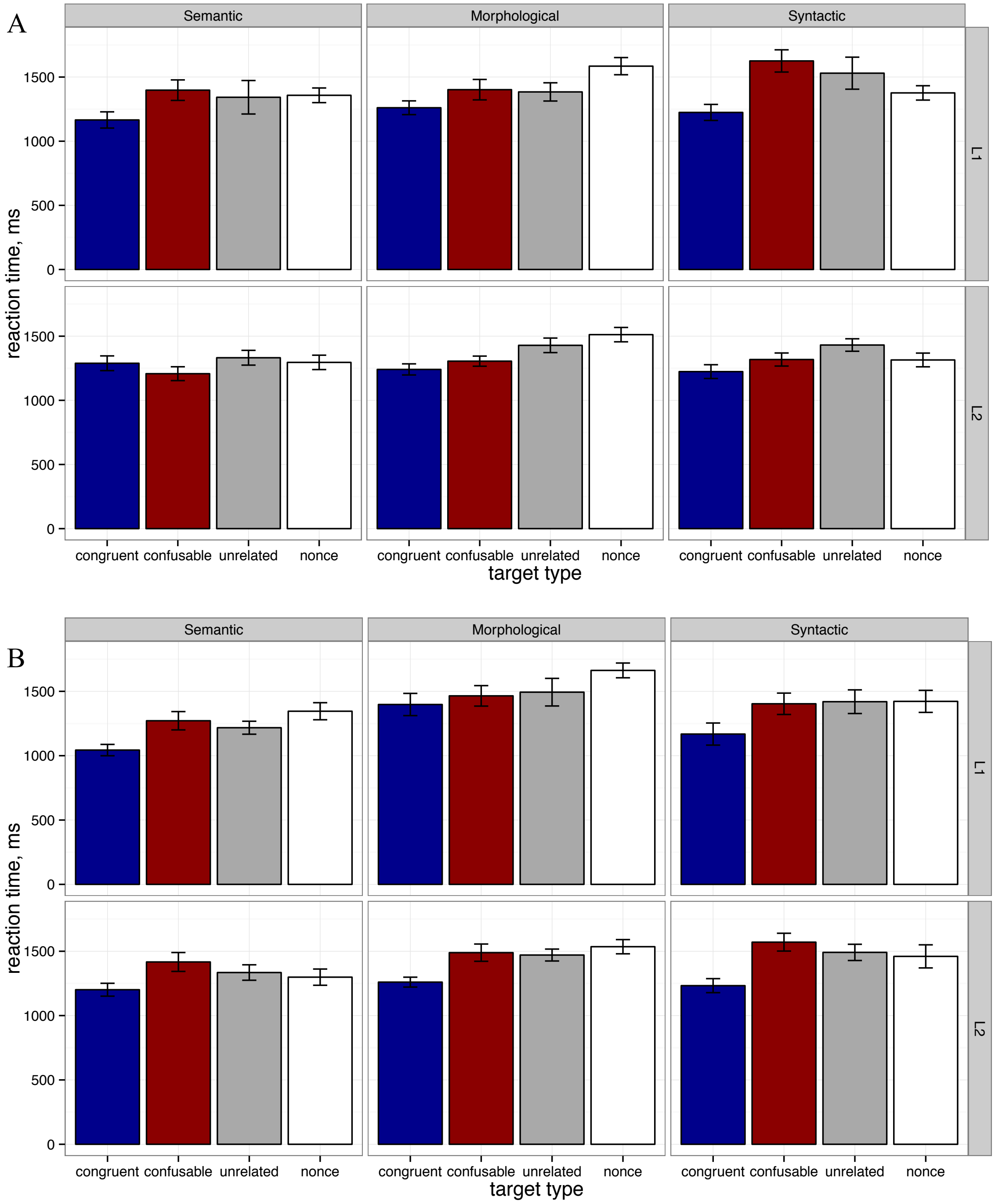


Figure 4. Participants' response latencies to word recognition in contextually constraining sentences in (A) critical and (B) control conditions of the lexical decision task.

Apart from examining whether constraining sentential context has an effect on word recognition, one of our main research questions aimed to investigate how different kinds of contextual information, such as semantic, morphological and syntactic, are utilized for meaning resolution, and whether they are utilized to the same degree by L1 and L2 comprehenders. Because ERs and RTs associated with word recognition in a particular condition (semantic, morphological, or syntactic) may be affected by item-specific properties (e.g., word frequencies, cloze probabilities) in that condition, it is not fair to compare mean differences between congruent and incongruent targets in contextually-constraining sentences across context conditions with the goal to establish the magnitude of the context effect. Instead, we evaluated context effects against the neutral condition, which served as a baseline. Context effects were calculated as a difference in mean RT and ER for the same targets when they occurred in constraining as opposed to neutral sentences for corresponding conditions. Thus, for each target in the critical and control condition, three data points were compared: when it occurred within a neutral carrier sentence, when it occurred with congruent context, and when it occurred with incongruent context. Figures 5 and 6 graphically present context effects for RTs and ERs across the two blocks for the two language groups. Positive differences in RTs and ERs suggest a facilitative role of the context on word recognition; negative RTs and ERs suggest an inhibitory role of the context. While faster RTs were expected for congruent targets occurring in constraining relative to neutral sentences (facilitation) across all conditions and bocks, longer RTs and higher ERs (inhibition) were expected for the incongruent targets.

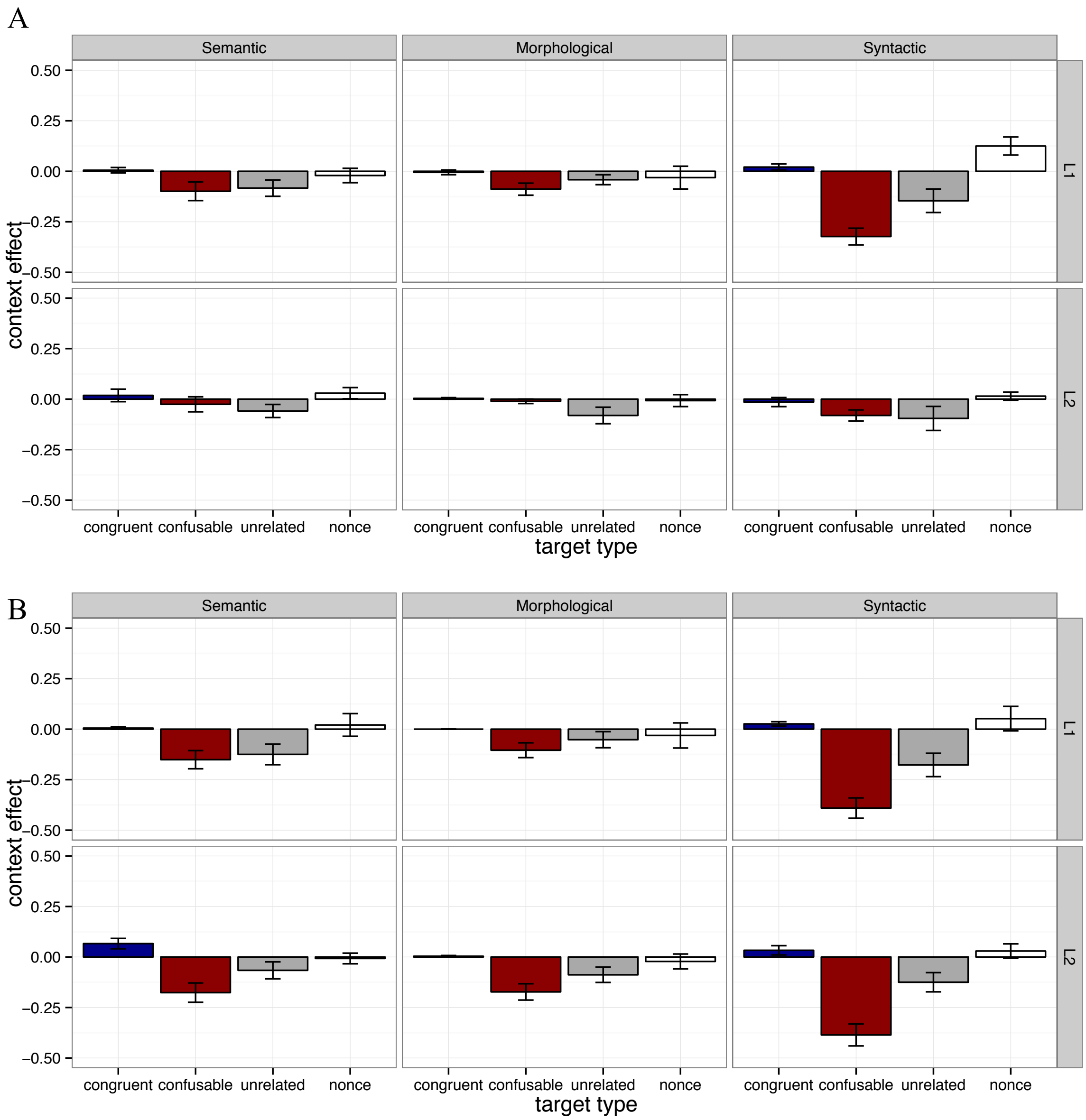


Figure 5. Context effects on word recognition error rate in (A) critical and (B) control conditions of the lexical decision task.

Note. Context effects are calculated as a difference in mean error rate (ER) between neutral and constraining sentences for corresponding conditions. Standard errors (SEs) are calculated as a square root of the sum of squares of SEs of the means to be compared: $\sqrt{se1^2 + se2^2}$. Positive ERs suggest a facilitative role of the context on word recognition; negative ERs suggest an inhibitory role of the context on word recognition.

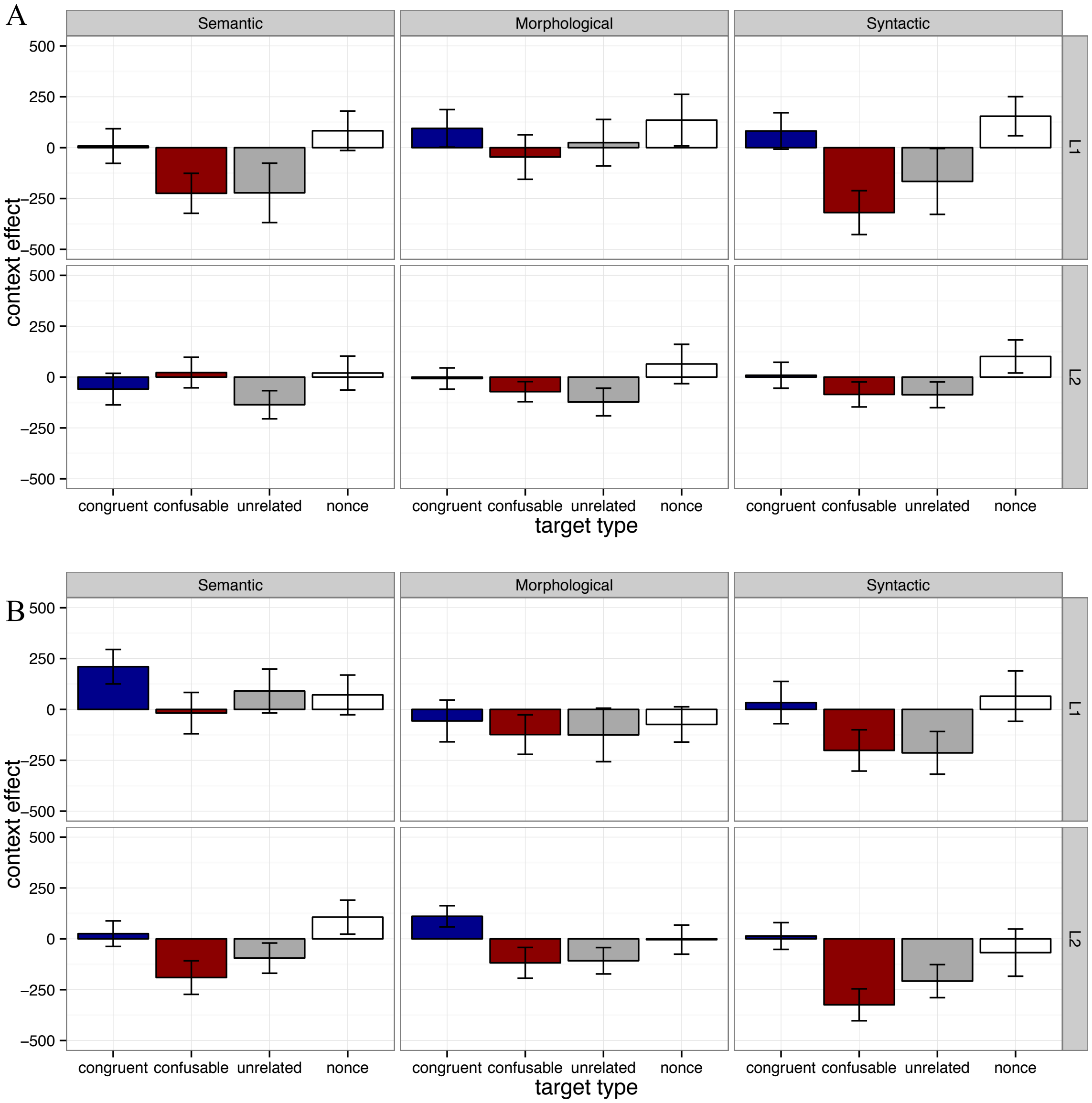


Figure 6. Context effects on word recognition time latencies in (A) critical and (B) control conditions of the lexical decision task.

Note. Context effects are calculated as a difference in mean reaction time (RT) between neutral and constraining sentences for corresponding conditions. Standard errors (SEs) are calculated as a square root of the sum of squares of SEs of the means to be compared: $\sqrt{se1^2 + se2^2}$. Positive RTs suggest a facilitative role of the context on word recognition; negative RTs suggest an inhibitory role of the context on word recognition.

Post-hoc Tukey HSD comparisons carried out for different target types provided some mixed results. For congruent targets, contextual constraints did not affect accuracy of word recognition by L1 or L2 listeners in either the critical or control blocks (ER differences are around zero in Figure 5). With regard to response latencies, facilitative context effects were observed in the morphological and syntactic conditions of the critical block, and the semantic condition of the control block for L1 speakers, and the morphological condition of the control block for the L2 speakers (Figure 6).

As far as incongruent targets are concerned, both confusable and unrelated words elicited on average more errors compared to the congruent words in both control and critical conditions in the L1 group suggesting test-takers' difficulty of overcoming incongruency and integrating the target word with the rest of the sentence. The syntactic condition created the strongest bias effect, particularly for the confusable targets ($p < 0.01$). L2 group performed similarly to the L1 group in the control condition, showing the greatest context bias effect for the confusable targets and a pronounced bias effect in the syntactic condition, but their error rate for the confusable targets was not different in the critical block.

Reaction time data for the L1 group suggest an inhibitory effect of context on the recognition of incongruent targets (both confusable and unrelated targets) in the semantic and syntactic conditions in the critical block and the morphological and syntactic condition in the control block. L2 group demonstrated an inhibitory effect of context on recognition of both confusable and unrelated targets across all conditions in the control block (semantic, $p < 0.05$; morphological, $p < 0.05$; and syntactic, $p < 0.01$). Although context bias effect was present for confusable (except for the semantic condition) and unrelated targets in the critical condition, it was diminished compared to the control block and the L1 group.

As far as nonce words are concerned, having a nonce word embedded in a meaningful sentence helped to reject it faster for L1 listeners, as evident by their RT latencies in the critical and control blocks. L2 listeners recognized nonce words faster in constraining sentences in the morphological and syntactic conditions in the critical block and semantic condition in the control block.

4.2.5 Summary of findings

The lexical decision task was designed to examine whether the difficulty with discrimination of phonological contrasts creates a phonolexical ambiguity for L2 comprehenders; whether they utilize information derived at higher levels of processing (semantic, morphological, and syntactic) to deal with such ambiguity at sentence level, and what kind of contextual information has the strongest effect on word recognition. The L2 participants' results were interpreted relative to the L1 participants' behavior.

When error rate and reaction time data for contextually congruent and incongruent words were examined, L1 speakers demonstrated an overall strong context effect on word recognition in both critical and control blocks. In other words, when sentential context constrained expectations for the upcoming word, and these expectations were not met, L1 comprehenders experienced a temporary disruption in word recognition. Previous research evidence predominantly suggests that the word recognition process is most intolerant of segmental mismatch in word-initial than in word-final positions (e.g., Allopenna, Magnuson, and Tanenhaus, 1998), but we observed longer reaction times for both the confusable targets (words with the phonological overlap in the word-initial position) and unrelated targets (words that diverged phonologically starting from the first phoneme of the word), and these reaction times were not significantly different between each other. This suggests that, in spite of the initial perfect match of a word, the mismatching sound, when it arrives, effectively disrupts comprehension flow creating a conflict in expectations. Coupled with the finding that lexical decision latencies on average did not differ between congruent words in contextually-constraining sentences and the same words in neutral sentences (contrary to the literature showing that words are recognized earlier in utterance contexts (e.g., Marslen-Wilson and Tyler, 1980)), longer reaction times for contextually incongruent words most likely reflect sentence integration costs at post-lexical stage of processing rather than at lexical access stage. This means that, by the time the listeners reached the contextually incongruent word, they have already constructed a semantic and a structural "template" of what has to come next, and when the incongruent target blocked the expected interpretation, they had to recover the intended target. Such breakdown in meaning integration is also reflected in participants' error rate data. Their task was to identify words and nonce words correctly, and although the error rate was very low for contextually congruent words, it increased for incongruent words. This could be due to the fact that the disruption of sentence processing was so strong, that having to attend to conflicting cues at the same time (i.e., having to press a "yes" button when the word does not fit the sentence) resulted in more error.

L2 participants demonstrated a similar pattern of results to the L1 speakers, but only in the control block. They also reliably showed an inhibition effect for contextually incongruent words in reaction time and error rate analyses. In line with our predictions, L2 listeners' performance was different in the critical block. While their response latencies were longer and error rate was higher for the unrelated targets in the critical block, they did not demonstrate significant differences for confusable targets compared to congruent targets. This suggests that they treated incongruent targets as congruent because they did not notice the phonological mismatch and, therefore, did not experience a disruption in word identification. Thus, L2 speakers' perceptual difficulty with the consonantal hardness and softness in Russian has consequences for lexical processing, creating spurious lexical candidates. This finding means that when L2 listeners have the necessary phonological representations in place and can differentiate between the target phonological contrasts easily, they can rely on their bottom-up strategies to extract the necessary phonological information to guide them to the correct lexical decision. In contrast, when phonological representations are fuzzy and unclear, L2 comprehenders have to rely on contextual information to compensate for the lack of perceptual resolution.

With regard to the question of which type of contextual information exerts the strongest effect on lexical expectations, our results point out that for both L1 and L2 groups, reaction time and error rate differences between congruent and incongruent words in constraining sentences on the one hand, and words in neutral sentences on the other hand, were the greatest in the syntactic condition. Context effects were the weakest in the morphological condition. We discuss possible explanations of these findings in the General Discussion section at the end of this chapter.

4.3 Experiment 2: Translation judgment task

While the findings of the previous experiment provide evidence in favor of phonolexical ambiguity and the use of contextual constraints for word recognition by L2 comprehenders, we cannot tell based on the lexical decision data whether words in some conditions caused more ambiguity than others. That is why a translation judgment task (TJT) was designed to provide additional information on the degree of ambiguity and confusability of words in each condition of the lexical decision task. In contrast to the LDT experiment, in which test-takers were required to decide whether the target word is a real word or not, the TJT experiment asked them to choose the correct translation of the target word, providing therefore more precise data about which words create more confusability. In addition, the TJT examined whether phonolexical ambiguity resulted in creating spurious lexical candidates and whether L2 speakers accepted nonwords as real Russian words as a result of such ambiguity.

4.3.1 Participants

The translation judgment task was only administered to L2 speakers. The same L2 participants who took part in the lexical decision task performed this task.

4.3.2 Design and materials

The experimental items were based on the stimuli used in the lexical decision task described in the previous chapter. They included a total of 144 items equally divided among the critical, control, and nonce blocks (48 items each). Critical and control items had to be counterbalanced in order not to expose the participants to the same translations. Nonce items were kept constant across the lists. This resulted in two 96-item presentation lists.

Similarly to the LDT, items in the critical and control blocks were phonological minimal pairs. Based on the relationship between the two members of the minimal pair, they could either mark a semantic, a syntactic, or a morphological distinction. Words in the critical block differed on the basis of consonantal hardness/softness; words in the control block differed on the basis of a phonological contrast common to both Russian and English.

Nonce words were of two types. The first type of nonwords (n=24) was created from real Russian words by replacing a word-final hard consonant with its soft counterpart, and vice versa. Such replacement resulted in two sets of nonwords, hard-to-soft, e.g., *дворь* (/dvor^{j}/) instead of *двор* (/dvor/, «yard»), and soft-to-hard, e.g., *двер* (/dver/) instead of *дверь* (/dver^{j}/, «door»), manipulation. The second type of nonce words, fillers, included Russian phonologically legal pseudowords with a "broken" stem.

All items are provided in Appendix C.

During stimuli presentation, all items were randomized. Each auditorily presented Russian word or nonce word was followed by four visually presented translation choices: a correct English translation, an English translation of a minimal pair counterpart, a distractor, and "not a real word" option. For example, for the target word *брат* ("brother"), the options were as follows: 1) brother, 2) to take, 3) jar, 4) not a word. The order of the translation choices was randomized across trials, but "not a word" choice always appeared in the fourth position.

4.3.3 Procedure

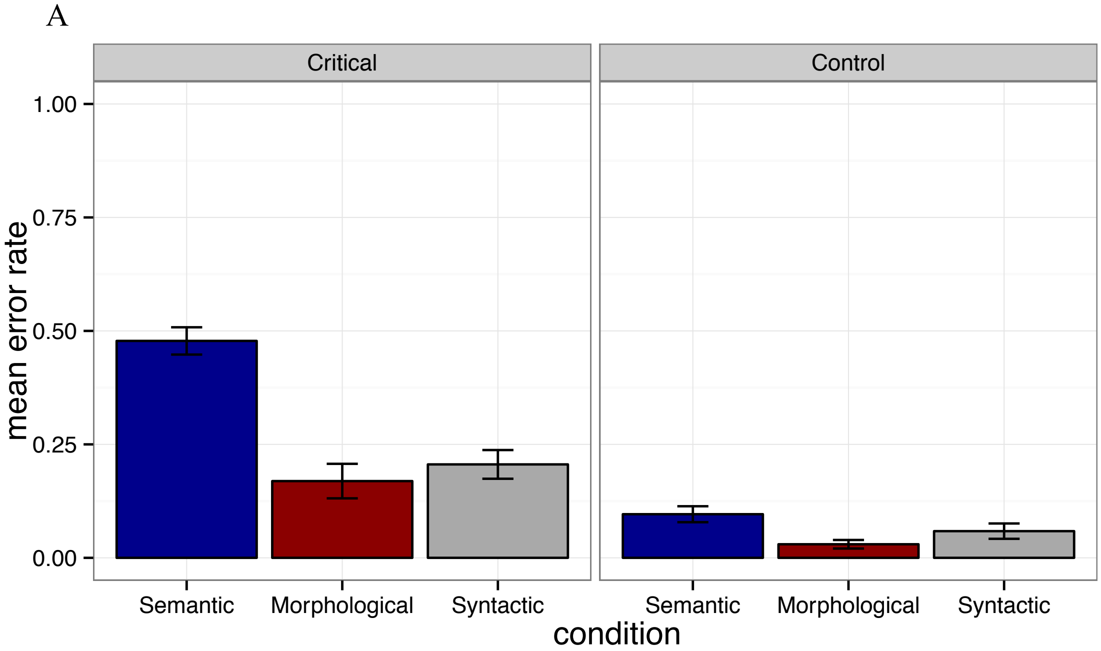
Participants listened to a list of Russian real words and sound strings that sound like real words but do not exist in the language. Each sound was followed by a 500 ms interval, after which four translation options were displayed on the computer screen. Participants were instructed to match the words with their correct English translations by pressing the corresponding button (1, 2, or 3) and identify all the non-existing words by pressing the button 4 (“not a word” response). After each button press, feedback on the response accuracy was provided. Participants were allowed to take short breaks after each 32 items. Practice items (n=6) were given at the beginning of the experiment for task familiarization purposes. The total duration of the experiment was 15-20 minutes.

4.3.4 Results

Subjects’ responses were scored as correct (error = 0) if they chose a correct translation or identified a nonword correctly, and as erroneous (error = 1) if they incorrectly chose a translation corresponding to the minimal pair counterpart (for words) or a real word instead of a nonce word (for nonwords condition). Since other types of errors were negligible, they were not included in the final analysis. Both error rate data and reaction time data were collected and analyzed. Reaction times were measured from the appearance of the English translations on the computer screen till the subject’s button press. They do not reflect real-time processing costs, and therefore should be taken with caution. Rather, they indicate a relative ease or difficulty of test-takers’ translation selection at the post-processing stage of a spoken target. Only reaction times to correct responses were included in the final analysis and subsequently trimmed at 100 ms and 10,000 ms resulting in 0.7% RT data rejection. Words and nonwords were analyzed separately.

For the analysis of words, two two-way ANOVAs (for error rate and reaction time variables) with the block (2 levels: critical or control) and the condition (3 levels: semantic, morphological, syntactic) as the within-subjects independent variables were carried out. Critically, a significant interaction between block and condition for both error rate and reaction time data was found (error rate: $F(2, 1624) = 21.28$, $p < 0.001$, $\eta^2 =$

0.03; reaction time: $F(2, 1342) = 3.7$, $p < 0.05$, $\eta^2 = 0.005$). There were also main effects of the block (error rate: $F(1, 1624) = 166.98$, $p < 0.001$, $\eta^2 = 0.09$; reaction time: $F(1, 1342) = 19.53$, $p < 0.001$, $\eta^2 = 0.013$) and condition (error rate: $F(2, 1624) = 45.04$, $p < 0.001$, $\eta^2 = 0.05$; reaction time: $F(2, 1342) = 5.01$, $p < 0.01$, $\eta^2 = 0.007$). Post-hoc Tukey HSD tests indicated that L2 listeners chose incorrect translation for the auditory target significantly more often in the critical block compared to the control block for each of the corresponding conditions ($p < 0.001$). Within blocks, error rate was not significantly different among the three conditions in the control block, but in the critical block, participants made more errors in the semantic compared to the morphological and the syntactic conditions ($p < 0.01$). Participants also took more time to choose the correct translation for target words in the morphological condition of the critical block compared to other conditions in the same block and compared to the morphological condition in the control block ($p < 0.01$). The results are graphically presented in Figure 7.



B

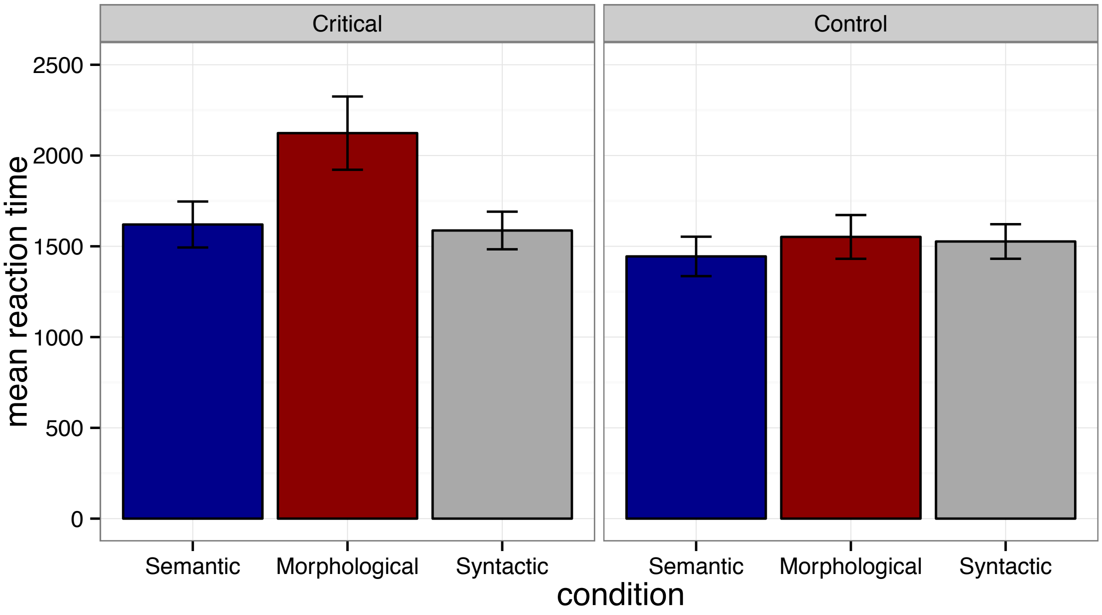


Figure 7. L2 participants' mean (A) error rate and (B) reaction time for translation of real words in the translation judgment task.

As far as identification of nonce words is concerned, a one-way ANOVA yielded a significant main effect of condition (3 levels: hard-to-soft nonwords, soft-to-hard nonwords, and fillers) for both error rate ($F(2, 1607) = 65.89$, $p < 0.001$, $\eta^2 = 0.4$) and reaction time ($F(2, 1139) = 27.3$, $p < 0.001$, $\eta^2 = 0.46$) analyses. Participants incorrectly accepted nonce words as real words in about 68% of the cases when nonce words involved a soft-to-hard consonant manipulation, 44% - when they involved a hard-to-soft consonant change, and only 1.3% in the filler condition. Participants also took less time to identify nonce words in the filler condition compared to the other two conditions. All differences were significant at $p < 0.01$. The results are graphically presented in Figure 8.

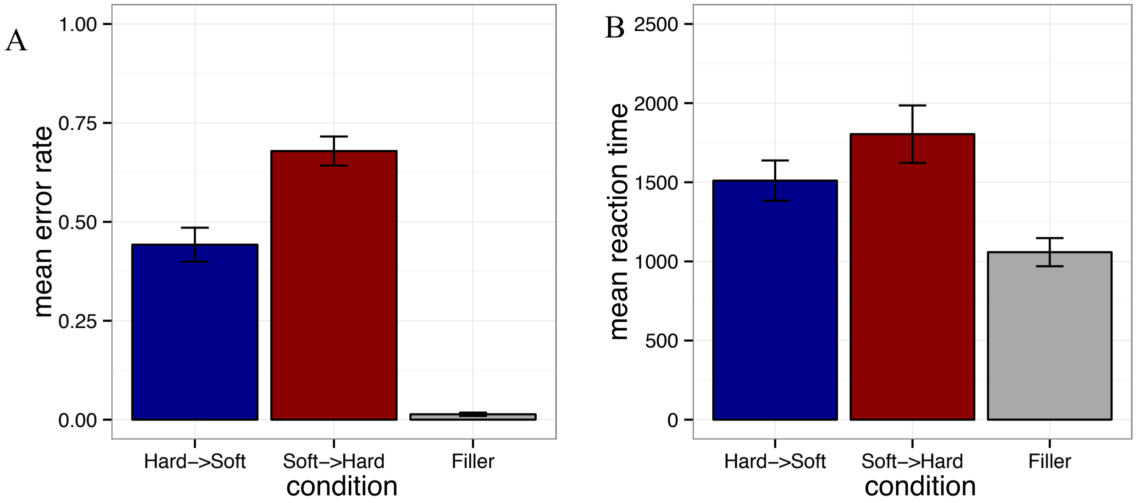


Figure 8. L2 participants' mean (A) error rate and (B) reaction times for identification of nonce words in the translation judgment task.

4.3.5 Summary of findings

The translation judgment task was designed to examine how phonolexical ambiguity affects L2 speakers' spoken word comprehension. We observed significant differences in translation accuracy for the words that presented perceptual difficulty for L2 listeners versus the words that did not. L2 listeners incorrectly chose the translation of the target's minimal pair counterpart more frequently when the words differed on the basis of consonantal hardness/softness. For example, they translated the verb *брать* (/brɑt^j/, "to take") as "brother" confusing it with the word брат (/brɑt/, "brother"). Translation accuracy was on average lower for the words in the semantic condition compared to the morphological and syntactic conditions. Such difference could be the result of the difference in the items' lemma frequency: target words in the semantic condition in both critical and control blocks had on average lower frequency (see Table 2). Notably, translation accuracy of the words in the semantic condition of the critical block was significantly lower than that of the words in the semantic condition of the control block although their lemma accuracy did not differ significantly, suggesting that L2 listeners' perceptual difficulty with the Russian hard and soft consonants extended to the lexical level.

In addition, L2 listeners appeared to accept nonce words that included substitutions of hard and soft consonants more often than they accepted nonce words in the filler condition. The acceptance rate for different types of substitutions was not the same. Nonce words with the consonantal manipulation from soft to hard (as in *двер* instead of *дверь,* «door») elicited more errors compared to the nonce words with the hard-to-soft manipulation (as in *дворь* instead of *двор,* «yard»). Such difference is also reflected in the reaction time data, with the soft-to-hard nonce words taking longer to identify than the hard-to-soft nonce words. The observed differences between the two types of nonce words are very unlikely to be due to the differences in word frequency of the corresponding real words because they were closely matched (hard-to-soft: *M* = 241.49, *SD* = 325.5; soft-to-hard: *M* = 243.64, *SD* = 396.8, according to the Russian national corpus http://www.ruscorpora.ru/index.html). The observed differences between the two types of nonce words may suggest that the effect of phonolexical ambiguity is asymmetric, and that the active category in the hard/soft consonant distinction for the L2 Russian speakers is the hard consonant (cf. Weber and Cutler, 2004). When a listener hears a nonword which is derived from a real Russian word, he or she has to match it up with the lexical representation of that word in order to be able to tell if what they hear is a real word or a made-up word. If what the listeners hear matches the representation of the word, the decision is made that it is a word; if what the listeners hear does not match any of the lexical representations stored in the long-term memory, the decision is made that it is a not a real word. The fact that soft-to-hard nonce words produced more errors and longer RTs than the hard-to-soft nonce words may mean that lexical representations for words with word-final soft consonant are less precise and more ambiguous. L2 speakers may even categorize a Russian soft consonant as a hard one because when it is substituted with a hard consonant in a word they frequently do not notice the mismatch. On the one hand, the observed perceptual asymmetry could be due to the fact that hard consonants are unmarked while soft consonants are usually interpreted as the ones having a secondary articulation (the raising of the middle part of the tongue towards the hard palate). On the other hand, a greater confusion with the soft-to-hard nonce words can also be due to the fact that, although not completely identical, the native English consonants are more proximate to the Russian hard consonants than soft consonants.

In summary, the findings of the TJT experiment provide additional evidence in favor of phonolexical ambiguity, which is routinely experienced by L2 listeners (see also Broersma and Cutler, 2011; Cook, 2012; Cook and Gor, 2012; Cutler, Weber, and Otake, 2006; Pallier et al., 2001; Sebastián-Gallés et al., 2005; Weber and Cutler, 2004). They demonstrate that successful spoken word recognition is contingent on the ability of the phonological system to encode and categorize acoustic-phonetic information efficiently and accurately, and that phonological ambiguity results in fuzzy, ambiguous lexical representations potentially creating spurious lexical competition and compromising word recognition (e.g., L2 listeners accepting nonwords like *двер* created from *дверь* ("door") as real Russian words in about 73% of the cases).

4.4 Experiment 3: Self-paced listening task

The lexical decision task in context and the translation judgment task provided evidence in favor of phonolexical ambiguity in the L2 and demonstrated that comprehenders use contextual information for word recognition in online auditory sentence comprehension. But because data were obtained from the explicit measures taken at the endpoint of processing (i.e., a button press at the end of the sentence or after word presentation) rather than continuously, they do not reveal the more dynamic aspects of spoken word comprehension. To examine the time course of integration of phonological information with higher-order contextual information under phonologically ambiguous and unambiguous conditions, a self-paced listening (SPL) task was administered. This experimental paradigm was introduced by Ferreira, Henderson, Anes, Weeks, and McFarlane (1996), who first demonstrated that it is sensitive to both lexical processes and syntactic variables in auditory sentence comprehension. As in the self-paced reading task, participants are required to press the forward button to proceed to the next region of the sentence (usually a word, but sometimes a sentence segment) while the time taken to listen to each sentence region is recorded. SPL task is also described as a useful technique for studying sentence processing at word level, because listeners' noticing of word-level violations can be tested. It is assumed that the time needed to move from one region to another reflects the relative ease or difficulty of processing the input, and, therefore, the technique can be used to examine the time course of integrative auditory comprehension.

Similarly to the LDT experiment which we described earlier, the SPL task also draws on the idea of phonological fuzziness and also makes use of the difficult L2 contrast of consonantal hardness/softness in L2 Russian. By manipulating the phonological form of the word in the critical region, we intend to examine the consequences of the phonological mismatch on sentential integration of phonological information on the one hand, and semantic, syntactic and morphological information on the other. The predictions are as follows. If phonological mismatch disrupts the comprehension flow, L1 listeners are expected to demonstrate reliable differences in listening times between contextually congruent and incongruent words in the critical region, and possibly, spillover regions. In contrast, L2 comprehenders are likely to demonstrate differential processing times for congruent versus incongruent words only in the phonologically unambiguous condition. Based on the previous literature and the findings from the LDT and TJT, substitutions of L2 phonologically ambiguous words are expected to go unnoticed. In addition, L2 listeners should demonstrate an overall slower sentence processing than L1 listeners across all conditions. The power of contextual constraints, or context effects, will be calculated as the difference in processing times for congruent versus incongruent words in the critical region as well as spillover regions.

4.4.1 Participants

The same participants who took part in the lexical decision task participated in the self-paced listening task. The presentation of the tasks was counterbalanced: half of the participants performed the self-paced listening task first, the other half performed the lexical decision task first.

4.4.2 Design and materials

The main stimulus manipulation involved the type of the target (congruent, confusable, control) and the type of the context condition (semantic, morphological, syntactic). Targets were embedded in sentences, which were divided into eight regions. A region could coincide with a word, or a phrase (sentence segment), and was presented auditorily one at a time. Target words always occurred in the fifth region of the sentence. They either fit the preceding sentential context structurally and/or semantically *(congruent* targets) or did not (*incongruent* targets). Incongruent targets were of two types, *confusable* and *control* (see Table 5 for an example sentence).

Table 5. An example of a stimulus sentence in the self-paced listening task.

Target type	Pre-target context				Target	Post-target context		
	1	2	3	4	5	6	7	8
Congruent	**Учительница** *The teacher*	**пригласила** *invited*	**на родительское собрание** *to the parents' meeting*	**отца и** *the father and the*	**мать** *mother*	**моего** *of my*	**лучшего** *best*	**друга.** *friend.*
Confusable	**Учительница** *The teacher*	**пригласила** *invited*	**на родительское собрание** *to the parents' meeting*	**отца и** *the father and the*	**мат** *checkmate*	**моего** *of my*	**лучшего** *best*	**друга.** *friend.*
Control	**Учительница** *The teacher*	**пригласила** *invited*	**на родительское собрание** *to the parents' meeting*	**отца и** *the father and the*	**газ** *gas*	**моего** *of my*	**лучшего** *best*	**друга.** *friend.*

Confusable targets and congruent targets constituted a phonological minimal pair. They differed on the basis of consonantal hardness/softness in the word-final position, e.g., *мат* (“checkmate”, /mɑt/)—*мать* (“mother”, /mɑtʲ/). The target minimal pairs did not always share the same orthography, but they were the same phonologically except for the final consonant, e.g., *балет* (“ballet”, /bɑˈlʲɛt/)—*болеть* (“to be sick”, /bɑˈlʲɛtʲ/). Each word in a minimal pair occurred both as a congruent and a confusable target, but never in the same presentation list. In total, 4 presentation lists with 144 sentences each were created such that the same listener was never exposed to the same sentence or a word from the same minimal pair (i.e., a presentation list could contain either “*мат*” or “*мать*”, but not both). Control targets were also incongruent with the preceding sentential context but were not phonologically related to the congruent targets (semantic and syntactic conditions) or did not pose a perceptual discriminability problem (morphological condition). Care was taken to match control targets with congruent and confusable targets in word length and surface frequency (see Table 6), although a limited number of available phonological minimal pairs meeting the experiment requirements posed serious design constraints.

Table 6. Stimulus materials design for the self-paced listening task.

Condition	Target type	Example	Translation	Number of phonemes		Log surface frequency	
				Mean	SD	Mean	SD
Semantic	Congruent/ Confusable	мать/ мат	mother/ checkmate	3.50	0.52	0.99	0.70
	Control	газ	gas	3.50	0.53	1.34	0.31
Syntactic	Congruent/ Confusable	брать/ брат	to take/ brother	4.13	1.09	1.21	0.79
	Control	вниз	downward	4.13	1.13	1.70	0.42
Morphological	Congruent/ Confusable	говорить/ говорит	to speak/ speaks	6.25	0.45	1.28	0.60
	Control	говорим	we speak	6.25	0.46	0.27	0.60

The second experimental manipulation involved the relationship between the pre-target context and the target itself. The congruent and confusable targets used in this experiment were phonological minimal pairs of three different kinds. The first kind included minimal pairs in which a change in the word-final consonant affected word meaning without affecting other word properties, as in *мат* ("checkmate", /mɑt/)—*мать* ("mother", /mɑtʲ/), where both members of the minimal pair are singular nouns in the Nominative/Accusative case. Such minimal pairs were included in the *semantic* condition, where the pre-target sentential context created a semantic bias in favor of one of the words in the minimal pair. A substitution of one word for another in such a sentence should, therefore, violate semantic expectations of the listener. Control words were also singular nouns in the Nominative/Accusative case and also created a semantic violation.

Minimal pairs could also include words marked by different syntactic properties, e.g., different parts of speech, as in *брат* (“brother”, /brɑt/)—*брать* (“to take”, /brɑtʲ/), where the first word is a noun and the second word is a verb in the infinitive form. Such minimal pairs were used in the *syntactic* condition, in which a target-preceding context did not only create semantic expectations, but also biased the listener’s structural expectations in favor of different syntactic categories (noun vs. verb). Similar to confusable targets, control targets in this condition also belonged to a different syntactic category, adverb.

If the two words in the minimal pair constituted different forms of the same word, the pre-target part of the sentence could constrain the target word morphologically (*morphological* condition). For example, an infinitive form of the verb is expected after another verb in the following sentence, *Подозреваемый не хочет говорить правду* (“The suspect does not want to reveal the truth”). Thus, congruent and confusable targets in the morphological condition were minimal pairs, in which one word was a verb infinitive and another one was a 3rd person singular form in the present tense: *говорит* (“speaks”, /gɑvɑˈrʲit/)—*говорить* (“to speak”, /gɑvɑˈrʲitʲ/). The control target was a present tense form in the 2nd person plural, *говорим* (“we speak”, /gɑvɑˈrʲim). It also formed a minimal pair with the critical targets, but the phonological contrast involved in its distinction (/t/ vs. /m/) was common to both Russian and English and was not expected to create perceptual difficulty for L2 listeners. A full list of targets and experimental sentences is presented in Appendix D.

The critical sentences (n=48) constituted one third of the total number of items in each presentation list. The remaining items were filler sentences, which were added in order to 1) make the critical violations less obvious, and 2) balance the number of incongruent items such that half of the sentences in each list were well-formed, and another half included semantic, syntactic, or morphological violations. All items were randomized. Filler items were not included in the analysis.

In addition, comprehension Yes/No questions (presented visually on the computer screen) were included after congruent sentences (n=72) to ensure that the listeners were attending to sentence meaning. Eight practice sentences and four questions were presented at the beginning of the test for task familiarization purposes.

All sentences were recorded by a native speaker of Russian using a normal speech rate and digitized at a sampling rate of 44 kHz. Each recording was cut into eight segments using Praat sound editing software (Boersma and Weenink, 2010), and each segment was saved as a separate sound file. The sound files were stringed together in such a way that the pre-target segments and the post-target segments were acoustically identical across different target conditions and the targets themselves did not differ physically between presentation lists.

4.4.3 Procedure

The experiment was delivered with the DMDX software (Forster and Forster, 2003). Participants were randomly assigned to one of the four presentation lists. The participants' task was to listen to the sentences presented through headphones and to answer comprehension questions on the computer screen as accurately as possible. They were asked to press RIGHT CTRL button for the affirmative response to the question, and LEFT CTRL button for the negative response. Participants were instructed to pace through the sentence segment by segment at a comfortable speed by pressing the forward button. The beginning of each sentence was signaled by a short beep sound. Auditory presentation of each segment was accompanied by a fixation cross (+) on the screen, which disappeared as soon as the participants pressed the button or after 4 seconds if the button press timed out. Reaction time was measured from the onset of the presentation of each sentence segment. The total duration of the experiment was about 45-50 minutes.

4.4.4 Results

First, each participant's accuracy of responses to comprehension questions was evaluated. L1 listeners made 3.4% errors (*SE* = 0.7%), L2 group made 5.9% errors (*SE* = 0.6%), indicating that both groups attended to sentence meaning.

Second, participants' listening latencies computed as the time interval between the onset of the sound and the button press were analyzed. False alarms (reaction times equaling zero) and timed-out responses (reaction times greater than 4 seconds) were excluded from the analysis resulting in 0.3% data rejection. The listening latencies are graphically presented in Figure 9. As apparent from the graph, L2 listeners' overall reaction times were slower than those of L1 listeners across all conditions. Importantly, L1 listeners slowed down when they encountered incongruent targets (both confusable and control) in the critical region across all context conditions, while L2 listeners only slowed down when they encountered control targets but not confusable targets.

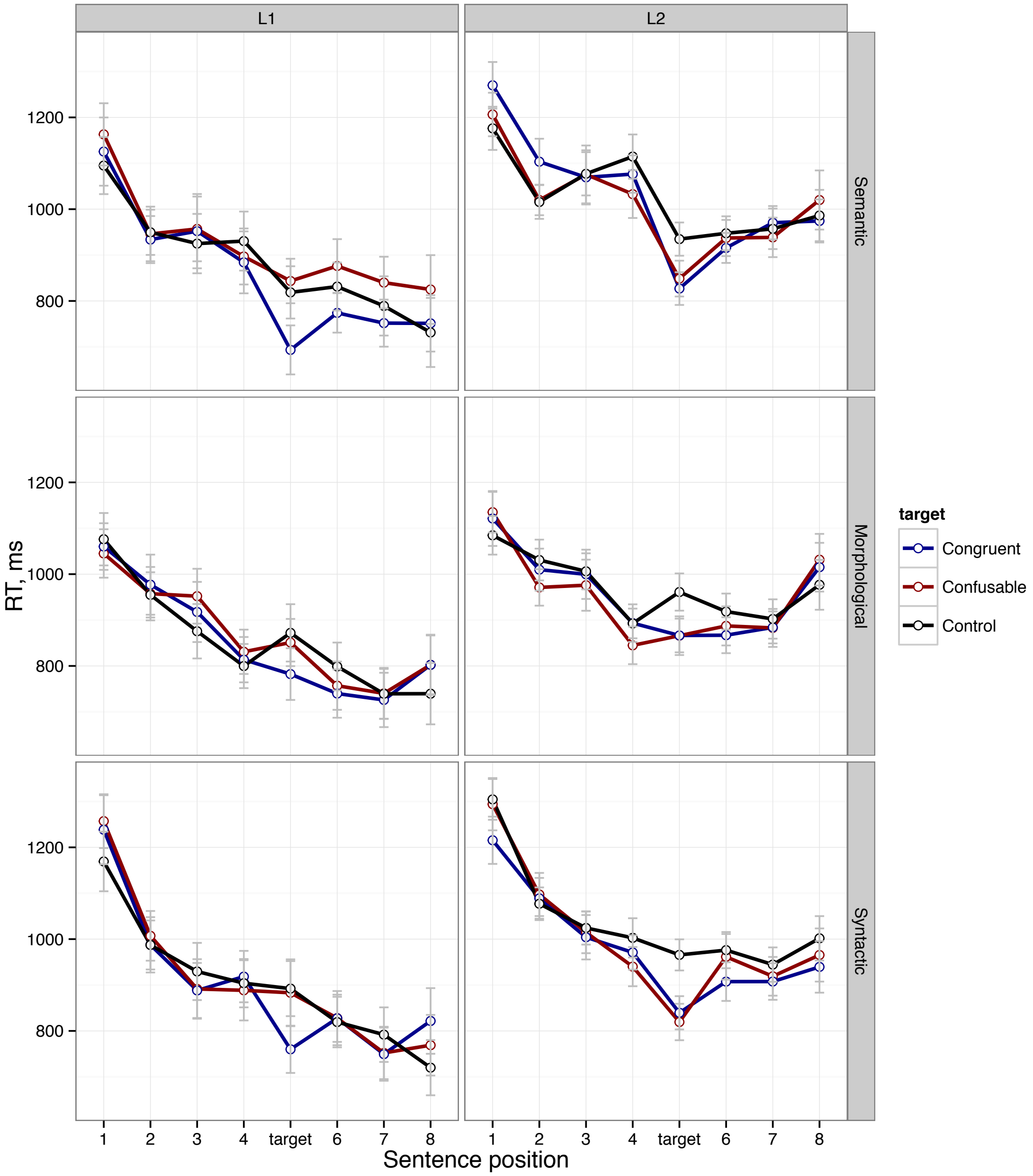


Figure 9. Participants' mean listening latencies (in milliseconds) across all experimental conditions in the self-paced listening task.

In order to account for the observed results statistically, a linear mixed-effects model was performed using the lme4 package (Bates and Maechler, 2010) in R statistical computing software (R Core Team, 2013). The mixed-effects model analysis was chosen over traditional analysis of variance (ANOVA) because it can account for possible individual differences among the participants and the variation that may exist in the stimulus materials. It also allows researchers to perform by-subject (F_1) and by-item (F_2) analyses within a single analytic framework.

Because there were no reliable effects at the positions prior to or following the target region, we concentrated our analysis on the critical region only. Language (2 levels: L1 and L2), condition (3 levels: semantic, morphological, syntactic), and target type (3 levels: congruent, confusable, control) were entered as fixed effects while subjects and items were treated as nested random effects with random intercepts. The best-fitting regression model included all main effects as well as three two-way interactions (language by condition, language by target, condition by target) and one three-way interaction (language by condition by target). Table 7 presents the model's estimated coefficients for each predictor, their standard errors, the *t* statistic, and the associated *p* values. The intercept (baseline comparison) estimated listening latency for the congruent target in the semantic condition for the L1 group. The coefficients are interpreted as the change in the reaction time brought about by the change of a predictor factor from one level to another. For example, a change of the language variable from L1 to L2 for the semantic condition and the congruent target results in the increase of reaction time of 134.02 ms.

Table 7. Estimated coefficients from a mixed-effects model for participants' listening latencies in the critical region.

Fixed effects	Estimate	SE	*t* value
(Intercept)	694.87	55.45	12.53*
(language) L2	134.02	69.12	1.94
(condition) Morphological	85.52	46.74	1.83
(condition) Syntactic	65.07	46.65	1.40
(target) Confusable	148.58	40.52	3.67*
(target) Control	123.81	35.03	3.53*

(language x condition) L2 x Morphological	-48.26	52.82	-0.91
(language x condition) L2 x Syntactic	-54.73	52.73	-1.04
(language x target) L2 x Confusable	-128.06	52.82	-2.43*
(language x target) L2 x Control	-19.65	45.71	-0.43
(condition x target) Morphological x Confusable	-77.92	57.23	-1.36
(condition x target) Syntactic x Confusable	-25.41	57.16	-0.45
(condition x target) Morphological x Control	-32.15	49.54	-0.65
(condition x target) Syntactic x Control	8.43	49.46	0.17
(language x condition x target) L2 x Morphological x Confusable	58.30	74.64	0.78
(language x condition x target) L2 x Syntactic x Confusable	-13.57	74.61	-0.18
(language x condition x target) L2 x Morphological x Control	22.58	64.64	0.35
(language x condition x target) L2 x Syntactic x Control	13.47	64.56	0.21
Random effects	Variance	SD	
Subject	47581	218.1	
Item	4339	65.87	

*Note: t-value = Coefficient/SE, with t-values over 2.0 indicating that the coefficient is significantly different from zero (Gelman & Hill, 2007). Bold indicates coefficients that are statistically significant, * p < 0.05.*

The results of the mixed-effects model yielded an overall significant effect of the target type for the L1 group: confusable and control targets took significantly longer to comprehend compared to congruent targets across all context conditions. For the L2 group, response latencies to control targets were longer than to congruent targets but not statistically different from those demonstrated by the L1 listeners, suggesting that the L2 listeners were sensitive to the violations in the sentences similarly to L1 listeners. The interaction between language and target was significant for the confusable target (*t* = -2.43, *SE* = 52.82, *p* < 0.05) in the semantic condition, and the coefficients for the confusable target in the morphological and syntactic conditions did not differ significantly from the semantic condition, suggesting that the L2 listeners did not notice word substitutions when they involved a perceptually difficult contrast across all context conditions.

One of the assumptions behind the design of the self-paced listening task was that participants' response latencies should reflect the ease or difficulty of sentence processing. If a certain word presents difficulty for integration into the sentence context, comprehending it should require more time. Based on this, context bias can be estimated as a difference in reaction times in listening to context-congruent versus context-incongruent words: the stronger the context biases the listener's expectations for the upcoming word, the more difficult it is going to be to process a word that defies such expectations and the longer it will take to move onto the next word. To compare context bias effects, reaction time differences between congruent and incongruent conditions were calculated. As evident from Table 8, L1 listeners demonstrated a context bias effect in all three context conditions (semantic, morphological, and syntactic) for both types of incongruent targets (confusable and control), but the context bias effect was greater in the semantic and the syntactic conditions compared to the morphological condition. Although it was not statistically significant, L1 participants demonstrated a spillover effect in the semantic condition suggesting that it took them longer to recover from semantic inconsistencies. For the L2 group, reaction time differences between congruent and control targets were the greatest in the syntactic condition, followed by the semantic condition, and, lastly, by the morphological condition, but only for control targets.

Table 8. Participants' mean reaction times in the critical region and mean differences between congruent and incongruent targets. SE = standard error.

Condition	Target	L1				L2			
		Mean	SE	RT difference	SE difference	Mean	SE	RT difference	SE difference
Semantic	Congruent	693.3	53.4	na	na	827	35.6	na	na
	Confusable	843.5	48.5	150.1	72.1	848.7	39	21.8	52.8
	Control	818.7	56.9	125.3	78.1	934.8	36.2	107.8	50.8
Morphological	Congruent	782.2	56.3	na	na	866.4	36.7	na	na
	Confusable	851	51.4	68.8	76.3	866	42.2	-0.4	55.9
	Control	872	62.5	89.8	84.1	961	40.5	94.6	54.7
Syntactic	Congruent	759.9	51.5	na	na	839.5	36.2	na	na
	Confusable	883.1	72.8	123.2	89.2	819.4	39.7	-20.1	53.7
	Control	892.2	60	132.2	79.1	965.6	33.9	126.1	49.6

4.4.5 Summary of findings

In online speech processing, comprehenders make use of all linguistics cues (e.g., semantic, morphological, syntactic) to build up expectations for upcoming words or word features. The SPL task was designed to examine the time course of interaction of these expectations with information derived at the perceptual level, especially when such information created ambiguity (in case of L2 listeners).

For the most part, the results of the study aligned with the predictions. L1 comprehenders showed reliable differences in listening times between contextually congruent and incongruent words suggesting difficulty of integrating incongruent words into the sentential context. Critically, they experienced the same processing difficulty integrating phonologically similar (confusable) and control (phonologically divergent) incongruent words across all context conditions. In contrast, L2 comprehenders showed a significant difference in response times to congruent compared to control targets, but not to confusable targets. This finding supports our hypotheses and suggests that although L2 speakers draw on similar (albeit slower) mechanisms during sentence comprehension and utilize contextual information to actively predict the upcoming auditory input, their vague and fuzzy phonological representations cause phonolexical ambiguity and prevent accessing phonological information of words when those are integrated with the rest of the sentence content, both semantic and structural. Such deficiency makes L2 comprehenders completely dependent on contextual information for meaning resolution if the words are phonologically ambiguous for them.

In terms of the use of specific contextual information, L1 listeners experienced the strongest context effects in the syntactic and semantic conditions followed by the morphological condition. The diminished effect in the morphological condition could be due to the fact that the phonological mismatch does not also involve a lexical mismatch, as in the syntactic and semantic conditions, so it should be relatively easier to integrate a context-incongruent form of the verb in the sentence because its meaning can still be accessed and a sentence can still be understood (e.g., "*They *<u>goes</u> to the gym every day*" instead of "*They go to the gym every day*"). In contrast, when access to a fitting lexical item is blocked, as in "*They go to the <u>tree</u> every day*" (semantic violation), or "*They go to the <u>regularly</u> every day*" (syntactic violation), listeners need more time to recover from the comprehension breakdown.

Although L2 listeners processed sentences on average more slowly than L1 listeners and the differences between their response latencies were smaller across all conditions, they demonstrated a similar pattern of results. In the control target condition, context effects were the largest in the syntactic condition, followed by the semantic and finally morphological condition. For the confusable targets, no statistically meaningful differences were observed among different context conditions in the L2 group.

4.5 General discussion

4.5.1 Phonolexical ambiguity

Discrimination of sounds in the L2 can pose perceptual difficulty. This has been demonstrated numerous times using different sounds, different languages, and different experimental paradigms. Fewer studies have looked at how such perceptual difficulties affect spoken word recognition, but those that have, reported contingency of L2 comprehenders' word recognition on their ability to discriminate between L2 phonemes (e.g., Broersma, 2002; 2005; Broersma and Cutler, 2011; Díaz et al., 2012; Pallier et al., 2001; Sebastián-Gallés et al., 2005). Unlike previous studies that investigate individual phonemic contrasts, the focal point of the present study is the processing of a phonological feature, namely, consonantal softness in the Russian language. Over a course of three experiments, we provide evidence that a lack of perceptual discriminability of words that differ on the basis of such phonological feature causes ambiguity at the lexical level.

Evidence from the translation judgment task shows that, when asked to choose a corresponding English translation for an auditorily presented Russian word, L2 participants tended to choose a translation of a similar-sounding word instead of the target word when the two words were contrasted in consonantal hardness/softness. For example, *брать* (/brɑt^{j}/, "to take") was translated as a similar-sounding *брат* (/brɑt/, "brother"). They also tended to accept nonwords created by substituting the hard and soft consonants as real Russian words, suggesting that phonological ambiguity results in lexical ambiguity and compromises word recognition. Such results extend the findings of previous studies. For example, Sebastián-Gallés and colleagues (2005) reported that Spanish-Catalan bilinguals accepted nonwords created by manipulating the Catalan /e/-/ɛ/ contrast as words significantly more often than nonwords with a control contrast. Similarly, Broersma and Cutler (2011) showed that due to a fuzzy distinction between the English /æ/-/ɛ/ vowels, Dutch listeners accept nonwords like *lemp* as real English words (i.e., *lamp*) more often than English listeners do. Using a similar translation task, Cook and Gor (2012) and Cook (2012) observed that L2 learners make the highest proportion of phonologically-related errors than errors of any other type, and that phonolexical confusion arises even when the words diverge phonologically in more than one sound and differ in the number of syllables, e.g., *костёр* (“bonfire”, /kasˈtʲor/)—*кастрюля* (“pot”, /kasˈtrʲulʲa/).

It was also found that the effects of phonolexical ambiguity are asymmetric, with the feature of consonant hardness being the dominant one. This asymmetry proves that there is no complete homophony involved. Similar effects have been reported before by Weber and Cutler (2004), who observed that Dutch L2 speakers of English mapped /æ/ and /ɛ/ inputs onto the same category /ɛ/. Such asymmetry carried through to word recognition in that *pan*, for instance, activated *pencil*, but *pen* did not activate *panda*. The observed perceptual asymmetry in our study may have several explanations. First, the dominance of the hard consonants at the phonological level can be due to the fact that they are usually thought of as ‘unmarked’ while the soft consonants are interpreted as ‘marked’, according to markedness theory. Second, it is possible that English speakers of Russian assimilated both hard and soft consonants into the same native category along the category-goodness assimilation pattern (in accordance with the predictions of the PAM-L2 model by Best and Tyler, 2007), and that hard consonants are perceived by English listeners as a more proximate category to the English consonants.

Now, what are the consequences of such phonolexical ambiguity for comprehension? If words only occurred in isolation, it would have presented an insoluble problem because of spurious competition among lexical candidates. As a result, lexical selection would have been hampered because some words would be perceived as similar sounding. Among the competing candidates, the higher-frequency word or the word that has more relevance or familiarity for the listener would win over. However, words rarely occur in isolation. Instead, in natural speech, words are strung together, and the way they are connected in sentences is mediated by complex semantic, syntactic, and morphosyntactic relationships among them. That is why recognizing spoken words in continuous speech entails not only attending to their phonological form, but also engaging higher-order processes (e.g., lexical processes, syntactic processes, compositional processes, etc.). The lexical decision task and self-paced listening task were designed in order to identify instances when contextual constraints work beneficially to facilitate word recognition and how different types of contextual information (semantic, syntactic, morphological) can potentially be used by comprehenders for meaning resolution during online sentence comprehension.

The results from both sentence processing experiments confirmed that L2 comprehenders experience effects of phonolexical ambiguity at sentence level processing. While L1 listeners exhibited reliable differences in response latencies between contextually congruent words on the one hand and both types of incongruent words on the other hand across all conditions in both experiments, L2 comprehenders demonstrated differential reaction times only for control words, but not for confusable words. Because L2 participants did not show any processing costs associated with contextual integration in the self-paced listening task, or inhibition effects in the lexical decision task for incongruent confusable words, it means that they treated these words as congruent with the context without a disruption in comprehension flow (see Figure 9, for example). This would only be possible if, despite the incompatible phonological information, they accessed the intended lexical candidates.

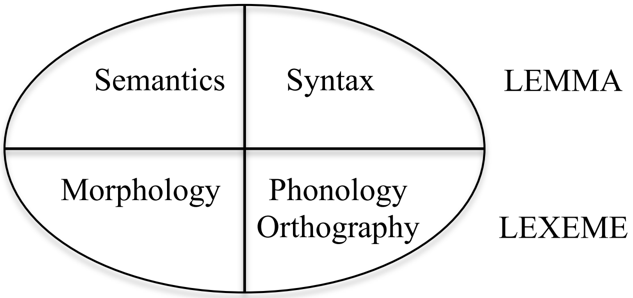


Figure 10. Structure of lexical entries (adapted from Levelt, 1993).

According to some models of organization of lexical storage (e.g., Levelt, 1993), lexical entries include two components, lemma and lexeme. Lexemes represent structural specifications of words (morphological, phonological, orthographic) while lemmas include specifications associated with meaning (semantic and syntactic) (Figure 10). These components are greatly integrated such that once a lexical entry is accessed in memory, all information becomes available. In context of this, our findings suggest that when L2 listeners do not have the robust phonological representations in place to allow them to differentiate between the words and cannot rely on the phonological properties of the word to guide them to the needed lexical candidate, they access and select the intended entry through its semantic, syntactic and morphological characteristics. Therefore, in L2 speech processing, an exact match between a lexical item and its phonological properties postulated by the Cohort model is not required. Listeners can compensate for the low-resolution phonological information by taking advantage of the information gleaned at the contextual level.

4.5.2 Context effects

When L2 comprehenders have the necessary phonological representations in place, they can combine phonetic-acoustic information coming from sensory levels of analysis with contextual information coming from higher-order processes in order to speed up and facilitate word recognition processes similarly to L1 speakers. By manipulating phonological information in the lexical decision task and the self-paced listening task, we examine which types of contextual information exert the strongest effects on lexical expectations. The results from both experiments point in the same direction. L2 listeners akin to L1 listeners experience the strongest context effects in the syntactic and semantic conditions followed by the morphological condition. Such results seem somewhat at odds with the existing SLA literature, where L2 speakers have been systematically shown to be more sensitive to semantic rather than syntactic violations (e.g., Hahne, 2001). We entertain several possible explanations.

It is possible that violation of syntactic expectations exerts the strongest influence on the parser because syntactic violations necessarily include lexical violations, e.g., *брат—брать* (brother – to take, noun – verb). According to the proposed structure of the lexical entry (Figure 10), syntactic and semantic properties of words are closely connected in the lemma component and are associated with word meaning. Thus, when listeners encounter a syntactic violation in the sentence, they have to reanalyze both the syntactic and the semantic properties of the target, which, naturally, should magnify the context effect compared to only a semantic violation. Following the same logic, context effects are the smallest in the morphological condition because the phonological mismatch does not involve a lexical mismatch as in the syntactic and the semantic conditions. When comprehenders reach a morphologically incongruent word in the sentence, they have to recheck their morphological hypotheses while the meaning of the word remains unaffected, e.g., *говорить—говорит* (to speak – speaks). That is why it should be relatively easy to re-evaluate and overwrite the formal properties of the word in order to integrate it with the context such that the sentence can still be understood (e.g., "*They go to the gym every day*" as opposed to "*They *goes to the gym every day*").

Another explanation of the weaker effect of the semantic constraints compared to the syntactic constraints could be due to the fact that semantic constraints are more specific while syntactic constraints are more general (e.g., requiring a noun and not a verb but providing little information about its specific characteristics) (Lee and Federmeier, 2009). Moreover, syntactic information is generally thought to be deterministic and definitive (and thus quite constraining) in a way that semantic information cannot be (Friederici, Pfeifer, and Hahne, 1999; Friederici, 2002). A sentence beginning with "*Mary got soaked to the skin because she forgot the …*" provides a semantically constraining context for the word *umbrella*, but it cannot rule out other options like *raincoat*. In contrast, the same sentence unambiguously and exhaustively specifies syntactic structure, i.e., a noun phrase (e.g., umbrella, new umbrella, etc.) that should follow the determiner. In case of ambiguous information, semantic cues should therefore be less effective than syntactic cues for meaning resolution. This is exactly what Folk and Morris (2003) found. They did not observe ambiguity effects when ambiguity crossed syntactic categories (e.g., a park—to park), which suggests that syntactic category information becomes available first and mediates the semantic resolution process.

It is also possible that contextual constraints operate differently for different classes of words. In our experiments, word categories of the target lexical items differed across the three critical conditions. In the semantic condition, we had noun-noun violations, the morphological condition included only verb-verb violations, and in the syntactic condition, ambiguities crossed different syntactic categories (e.g., noun-verb). Studies examining processing distinctions between nouns and verbs have observed significantly slower naming of verbs than nouns in the native and second languages (Faroqi-Shah & Waked, 2010; Szekely et al., 2005), dissociations of noun and verb retrieval in patients with aphasia (e.g., Caramazza & Hillis, 1991; Zingeser & Berndt, 1990), and different degrees of cortical activation for nouns and verbs (Yokoyama et al., 2006). Such noun-verb dissociation data are interpreted as evidence that lexical organization in the brain is governed by grammatical class (e.g., Caramazza & Hillis, 1991), with an implication that words within the same grammatical class should compete for lexical selection more than words belonging to different grammatical categories (Dell, Oppenheim, & Kittredge, 2008; Levelt, Roelofs, & Meyer, 1999; Pechmann & Zerbst, 2002). Based on these assumptions, participants in the present study may have experienced more competition and uncertainty in the semantic condition, which included violations within the same grammatical class. For the same reason, the syntactic condition was more effective in constraining word selection because it ruled out between-class competitors early in sentence comprehension. That is why minimal pairs like *мат* ("stalemate", /mɑt/)—*мать* ("mother", /mɑtʲ/) could have created more ambiguity than *брат* ("brother", /brɑt/)—*брать* ("to take", /brɑtʲ/). The morphological condition also had ambiguities within the same grammatical class (verbs). However, this condition was different from the semantic condition in that the phonological contrast marked the distinction between the two forms of the same verb rather than different verbs.

5 Sentence-level context effects in L1 and L2 auditory sentence comprehension: ERP evidence from disambiguation of morphological forms

5.1 Introduction

In Chapter 4, we examined behavioral evidence on how context can potentially help to disambiguate phonolexical ambiguity in the L2 (compared to the L1). However, speech processing occurs at extremely high rates, and very often, behavioral measures are unable to provide the desired temporal resolutions. Moreover, it is important to bear in mind that the absence of differences in behavioral measures does not necessarily mean that the underlying cognitive processing mechanisms are the same. By the same token, observed differences in behavioral measures, such as reaction times, are not necessarily the result of the involvement of different neuronal structures, even if they show qualitatively different patterns. Neurophysiological measures, such as ERPs, can complement behavioral measures and add valuable information about the nature and the time course of speech comprehension.

ERPs are summed post-synaptic electrical potentials of primarily synchronously activated pyramidal cells in the neocortex that can be triggered by an event, such as a word. These synaptic currents can be recorded at the scalp by placing electrodes on the head and amplifying the voltage difference between them (Luck, 2005). ERP is a well-suited technique for studying speech processing because it provides a temporal resolution on the order of milliseconds, which allows to observe how the process of interest unfolds in the short period of time between decoding of the acoustic signal and comprehension of the utterance such that both early and late processes can be examined. Besides, the advantage of the ERP measure is that it does not require a behavioral response, which makes it an ideal tool to study speech comprehension without confounding it with the interference from overt decision or response strategies, metalinguistic knowledge, or working memory. Most importantly, the registration of ERPs allows to tease apart lexical-semantic from syntactic processes. For example, self-paced listening data can indicate whether the listener experiences difficulty in one condition versus the other at a particular point during sentence processing. However, it is hard to tell from the difference in response times what kind of process caused that difficulty, e.g., whether a delay in RT is caused by a semantic or a syntactic problem. Using ERP method, it is possible to identify various components that are related to specific types of processes, which enables the researcher to draw inferences concerning the types of processes involved and their relation to one another (Kaan, 2007). The ERP components are typically defined by their timing, scalp distribution, sensitivity to experimental manipulations, and neural generators thereby providing useful dependent variables, such as presence/absence of a component, amplitude (size), timing, and/or the distribution over the scalp, which can reveal much information about the timing and nature of the neural and cognitive processes involved (Kutas and Federmeier, 2000).

5.2 Neurophysiological basis of morphological processing

The ERP components that are associated with morphosyntactic processes are P600 and E(LAN). The P600 component is a positive wave peaking at about 600 ms after the stimulus onset, usually distributed centro-parietally. This component is referred to as the P600 and is believed to reflect different aspects of syntactic processing. It has been repeatedly shown to be sensitive to syntactic violations (Friederici, Pfeifer, and Hahne, 1993; Neville, Nicol, Barss, Forster, and Garrett, 1991; Osterhout and Holcomb, 1992), syntactically complex structures (Kaan, Harris, Gibson, and Holcomb, 2000), the degree to which a syntactic continuation is expected, e.g., words that are ungrammatical continuations elicit a larger P600 than ones that are grammatical, but non-preferred (Osterhout, Holcomb, and Swinney, 1994). Thus, it has been interpreted as reflecting processes of reanalysis and/or syntactic repair (Osterhout et al., 1994) or as a more general index of the complexity of syntactic integration (Kaan et al., 2000).

Another component which is associated with syntactic domain is LAN. It represents a negatively going wave, which is primarily picked up at anterior or left anterior electrodes (hence the name), but its laterality and anterior location are not consistent across studies. Two types of LAN have been identified based on their timing: an early LAN (ELAN), typically occurring 100–200 ms after the onset of the critical stimulus, and a later LAN, typically peaking between 300 and 500 ms (i.e., in the same time window as the N400). LAN has been frequently found for morphosyntactic violations in the use of tense, number or gender agreement (Coulson, King, and Kutas, 1998; Gunter, Friederici, and Schriefers, 2000; Weyerts, Penke, Dohrn, Clahsen, and Münte, 1997) as well as in response to function words as compared to content words in grammatical sentences (Brown, Hagoort, and Ter Keurs, 1999; Neville, Mills, and Lawson, 1992). The ELAN has been associated with rapid first-pass parsing processes and automatic processing of phrase structure information. It is typically found for word category or phrase structure violations (e.g., when a passive participle rather than a noun follows a determiner) (Neville et al., 1991; Friederici et al., 1993). It is worth mentioning that the dissociation of the LAN and the ELAN components is not that clear-cut because LAN has also been found for phrase structure violations, and ELAN – for agreement violations. Kaan (2007) suggests that it is the same component, but its timing is influenced by the position of the affix that bears the agreement or word category information in the sentence: the earlier the parser encounters the information, the sooner it senses the difficulty and the earlier a LAN is elicited.

Although not of primary relevance for the present experiment, the N400 component at least deserves some brief mentioning. It was first reported by Kutas and Hillyard (1980), who compared brain responses to visually presented congruent sentences ("*He spread the warm bread with butter*") and sentences with a semantic anomaly ("*He spread the warm bread with *socks*') and found an enhanced negative-going wave peaking at around 400 ms post-stimulus onset time-locked to the semantically incongruent word. It typically has a right-central maximum distribution, but it can vary depending on the presentation mode (visual, auditory) and the nature of the stimuli (pictures, words). Since early 1980s the N400 component has been widely used as a dependent measure in studies examining the time course of the semantic aspects of sentence processing. It is believed to reflect either facilitation of lexical access due to context priming or pre-activation of the lexical candidate (Kutas and Federmeier, 2000; Federmeier, 2007), and/or the relative ease or difficulty of integration of the word with the semantic context (Osterhout & Holcomb, 1992; Hagoort, 2008). What is significant for the present study is that besides semantic effects, the N400 component has also been observed in response to morphological and syntactic violations in several studies. Münte et al. (1990) observed that a morphosyntactic violation of case marking in German is highly correlated with a negativity around 400 ms. Friederici and colleagues (1993) also found that morphological errors elicited a pronounced negativity between 300 and 600 ms, but with a smaller amplitude than the “semantic” N400, peaking earlier and merging into a late positivity around 600 ms. Osterhout and colleagues (2006) compared L1 and L2 speakers’ processing of morphosyntactic violations and found that while native speakers produced a P600 effect, L2 speakers demonstrated an N400 effect. Notably, the N400 effect evoked by morphosyntactic violations did not differ in its distribution from the N400 effect elicited by the semantically anomalous words in L2 speakers. The authors argued that at low levels of proficiency, morphosyntactic errors are not yet recognized as such by L2 learners, and so the anomalies are perceived as a lexical problem.

5.3 Experiment 4: Event-related potentials

The present experiment aims to investigate the electrophysiological aspects and temporal parameters of morpho-phonological processing in auditory sentence comprehension by L1 and L2 speakers on the example of the Russian language. While there is a huge amount of ERP literature on the effects of semantic and syntactic constraints in sentence processing, ERP studies on morphological processing are less abundant, and the link between morphology and phonology has hardly been explored except for a handful of studies. For example, Carrasco and Frenck-Mestre (2009), Frenck-Mestre, Osterhout, McLaughlin, and Foucart (2008), and Frenck-Mestre, Carrasco, McLaughlin, Osterhout, and Foucart (2010) examined covariation between phonology and morphology in a series of experiments on gender concord and subject-verb agreement in written French. These studies found that morphological forms are processed more readily when overtly realized phonetic cues are present (e.g., _Le matin je *mangez_$_{2}^{\text{nd}}{}_{/\text{PL}}$ _... “In the morning, I eat_$_{2}^{\text{nd}}{}_{/\text{PL}}$ _...”)_ compared to when they are absent (e.g., *Le matin je *manges*$_{2}^{nd}{}_{/SING}$ ... *“In the morning, I eat*$_{2}^{nd}{}_{/SING}$ …”) in morphological violations. The same result has been systematically replicated: L1 speakers of French showed that compared to grammatically correct instances (e.g., *Le matin je mange*$_{1}^{st}{}_{/SING}$ ... *“In the morning, I eat*$_{1}^{st}{}_{/SING}$ …”), morphological violations produced a robust P600 effect, which was significantly larger for the phonologically realized inflectional errors than for the errors that were silent (i.e., were only marked orthographically), suggesting that speakers have more solid representations of grammatical morphemes when they are supported by phonological differences. The effect was also found for L2 speakers from different language backgrounds, although it was systematically smaller and sometimes only observed for phonologically realized morphological violations, but not silent errors. No early negativities were elicited for either native or L2 groups.

The unique focus of the current study is that unlike previous ERP studies on morpho-phonological processing, it examines the impact of sentential morphological cues on the prediction of a certain morpho-phonological form. Of interest is a situation where morphological forms differ on the basis of one phonological segment (e.g., *sees—seen*), which can be either perceptually ambiguous or not for L2 comprehenders. The special contribution afforded by the Russian language is twofold. First, Russian has a very rich morphology, with words organized in highly structured and consistent sets of forms (paradigms) with inflections carrying grammatical meanings, which allows for the examination of complex morphological relations among words in sentence context. Normally, the stem of the word expresses its lexical meaning while the type of inflection specifies grammatical properties (in nouns—case, number, gender; in verbs—person, number, tense, etc.). For example, a regularly inflected verb like *ответить* (“to answer”, /ɑtvʲetʲitʲ/) can be decomposed into two transparent morphemic constituents: the stem *ответи-*, which encodes the content of the verb ‘‘to answer’’ (i.e., its meaning and grammatical category), and the inflection *-ть*, which denotes the inflectional feature [+infinitive]. Thus, due to the formative properties of morphological processes, manifold comparisons of different word forms within a paradigm can be made (e.g., *ответить–ответишь–ответит–ответим*, etc.).

Second, Russian possesses a phonological feature of consonantal hardness/softness that is quite conveniently involved in the generation and the juxtaposition of certain morphological forms. For example, note the *т-ть* (/t/–/tʲ/) distinction in the word-final position in *ответит*$_{FUTURE/3rd/SING}$ (“will answer”, /ɑtvʲetʲit) versus *ответить*$_{INF}$ (“to answer”, /ɑtvʲetʲitʲ/): the phonological contrast between the two minimal pairs also marks the morphological distinction between the two verbal forms (for a brief overview of the target feature, see Section 3.3). Because the phonological contrast between hard and soft consonants in word-final position presents a perceptual difficulty for English-speaking learners of Russian (Bondarko, 2005; Diehm, 1998; Lukyanchenko and Gor, 2011), this allows researchers to test for the phonolexical ambiguity at the level of morphological processing.

Overall, the above-outlined properties of the Russian language offer an optimal case for examining how and when low-level phonological details interact with higher-order contextual information (such as morphosyntactic agreement), both in L1 and L2 speech comprehension. The main goal of this experiment is to examine what kind of brain response (ERP component) is evoked by morpho-phonological violations and what the time course of the integration of phonological and morphological information is in L1 and L2 auditory sentence comprehension. Most previous ERP studies tested morphological (and morphosyntactic) violations during reading, but it is not clear whether the ERP effects (such as E(LAN), P600 and N400) observed in the studies using visual presentation will generalize to the auditory modality because visual and auditory stimuli presentations tap different representational levels of a morphologically complex word (in reading a word can be accessed as a whole whereas in listening it unfolds in time) (Clahsen, Sonnenstuhl, and Blevins, 2003; Holcomb and Neville, 1990; Holcomb, Coffey, and Neville, 1992; Lück, Hahne, and Clahsen, 2006). Importantly, we want to examine how the brain response changes depending on the phonological contrast involved in the distinction of two morphological forms. For example, one might predict a graded ERP response as a function of phonetic proximity/similarity (e.g., a larger P600 response to the incongruent *ответил* ("answered", /ɑtvʲetʲil/) compared to the incongruent *ответит* ("will answer", /ɑtvʲetʲit/) where the form *ответить* ("to answer", /ɑtvʲetʲitʲ/) is expected, because /tʲ/ and /t/ share more phonetic features than /tʲ/ and /l/). The predictions for the L2 listeners can go in different ways. If they lack the necessary morphological competence in accordance with the shallow-structure hypothesis (Clahsen & Felser, 2006a; 2006b) and are not sensitive to morphological cues during sentence comprehension, an ERP response to morphological violations may not be elicited. If, however, they are capable of extracting the necessary morphological cues during online auditory processing and use the grammatical information contained in the inflection for meaning integration and sentence comprehension, a difference in the ERP response to congruent versus incongruent conditions should be observed. On the other hand, in L2 learners, such response may be modulated by the level of perceptual difficulty of the morpho-phonological contrast. Given the evidence that the distinction between Russian hard and soft consonants is problematic for nonnative comprehenders of Russian, they may not show an ERP response (or show a reduced ERP response) to the morphological violations involving such a phonological contrast (as in *ответит*–*ответить*), suggesting a morphological context bias effect in the situation of phonological ambiguity.

5.3.1 Participants

L1 group included 21 native speakers of Russian (mean age 29.8, range 19-58; 14 females). Most of them were graduate students at the University of Maryland or recent graduates working in the Washington, DC area at the time of testing. L2 group included 15 American speakers of Russian as a second language (mean age 29.8, range 24-51; 7 females). All L2 speakers were screened for the study based on their language proficiency. Prior to the experiment, they were asked to fill out a language background questionnaire about their language learning experience, rate their language proficiency in different linguistic domains on a scale from 1 (minimum) to 10 (maximum), and complete a 25-item proficiency cloze test. Their average score on the cloze test was 22.27 out of the maximum of 25 (Table 9). Ten out of the 15 participants reported having taken the ACTFL Oral Proficiency Interview (OPI), a widely recognized language proficiency test. Two of these people had received a score of 2+ (Advanced High) on the Interagency Language Roundtable (ILR) scale; four people – a score of 3 (Superior); two people – a score of 3+ (Superior), and two participants – a score of 4 (Distinguished). All of the participants have visited or have lived in Russia at some point in their life (for an average of 2.83 years). At the time of testing they reported a frequent use of Russian on a daily basis (an average of 40%) with their Russian-speaking friends, for work-related purposes, and on the Internet.

Table 9. Linguistic profile of L2 participants in the ERP experiment.

	Mean	SD
Age when started learning Russian	17.00	2.75
Age when first traveled to Russia	18.93	5.44
Length of living in Russia (years)	2.83	1.61
Formal instruction in Russian (years)	2.77	0.67
Self-rated pronunciation	7.27	1.49
Self-rated oral proficiency	7.13	1.64
Self-rated listening proficiency	7.27	1.33
Self-rated reading proficiency	7.67	1.59
Self-rated writing proficiency	6.47	1.81
Self-rated knowledge of grammar	7.40	1.64
Cloze test (Proficiency measure)	22.27	2.40

5.3.2 Design and materials

The experimental materials consisted of a set of 180 triplets of sentences (a total of 540 sentences) for the critical (n = 90 triples) and the control (n = 90 triplets) conditions, which were counterbalanced across three presentation lists to ensure that no subject was exposed to the same sentence or critical word more than once. Additional 90 items were added as fillers resulting in three 270-item presentation lists. The sentences in each triplet were identical except for the target word, which was always embedded in about the middle of the sentence (for the critical condition: on average 3.26 words after sentence onset and 2.81 words before sentence offset; for the control condition: on average 4.28 words after sentence onset and 2.16 words before sentence offset). The target word could be either congruent or incongruent.

The *critical* condition included a three-way manipulation of the target word based on the type of the verbal form: the congruent infinitive form (V + -ть, /tʲ/), incongruent future-tense form (V + -т, /t/), and incongruent past-tense form (V + -л, /l/). In the congruent condition (n = 30 per list), morphological expectations for the target word were created with the help of the pre-target context: similarly to English, a verb following another verb (or auxiliary or modal) in Russian should take an infinitive form (e.g., *wants to read, loves to read, will read, may read*, etc.). The logic is that by the time the listener arrives at the target word, they should have their morphological expectations in place (through pre-activating those words that fit the expected morphological template, i.e., an infinitive form of the verb), even though the semantic content may still be unknown. In contrast, incongruent targets are supposed to conflict with the listener's morphological expectations and cause a temporary breakdown in the comprehension flow. Incongruent targets in the critical condition were of two types. The incongruent future-tense targets (n = 30 per list) differed from the congruent targets on the basis of the Russian-specific phonological contrast of consonantal hardness/softness (/tʲ/ vs. /t/ as in *ответить* – *ответит*), and were supposed to be phonologically ambiguous and perceptually difficult for L2 Russian listeners. Incongruent past-tense forms (n = 30 per list) differed from the congruent targets on the basis of an easy phonological contrast (/tʲ/ vs. /l/ as in *ответить* –*ответил*).

The *control* condition was included in the design of the experiment in order to assess the reliability of the ERP response in L1 and L2 listeners independently of the critical comparisons and in order to create more variability in the types of violations (nominal in addition to verbal paradigm). Control items also involved a three-way manipulation of the target word including congruent targets and two types of incongruent targets. The congruent targets (n = 30 per list) were always inanimate masculine nouns in the Accusative case (stem + -ø) used as direct objects. They were preceded by an adjectival modifier, which agreed with the target in gender, number and case, and, therefore, helped to set up morphological expectations for the necessary morphological form. In the morphologically incongruent control condition, the targets (n = 30 per list) were incorrectly used in the Dative case (stem + -у, /u/), e.g., *языку* ("to language", /jazɨku/) instead of *язык* ("language", /jazɨk/). In order not to make the participants too aware of the morphological violations in the experiment, sentences with semantic violations (n = 30 per list) were also added. Unlike in the critical condition, pre-target context in the incongruent semantic condition created a semantic bias in favor of a particular lexical candidate. Target words in the control incongruent semantic condition were matched with the control congruent targets in word length, lemma and surface frequency. Stimulus characteristics and sample sentences in the critical and the control conditions are presented in Table 10, and a full list of items is provided in Appendix E.

Finally, filler sentences (n = 90) were constructed in order to balance the number of congruent and incongruent sentences in each presentation list. Seventy-five of these sentences were congruent and 15 were incongruent. The latter involved various violations of the aspectual use in verbs (e.g., perfective in place of imperfective aspect). Thus, the ratio of congruent to incongruent sentences in each presentation list was 1:1 (30 critical congruent, 30 control congruent, 75 filler congruent = 135, and 60 critical incongruent, 60 control incongruent, 15 filler incongruent = 135; a total of 270 sentences per list). The ratio of critical to noncritical sentences per list was 1:2.

Table 10. Stimulus characteristics and example sentences in the critical and the control conditions of the ERP experiment.

Condition	Congruency	Example	Mean TW duration, ms	Mean TW lemma frequency, ipm
Critical	congruent	Личный помощник президента хочет ОТВЕТИТЬ$_{\mathrm{INF}}$ на провокационный вопрос журналиста. *President's personal assistant wants to ANSWER*$_{\mathrm{INF}}$ *the journalist's provocative question.*	0.78	60.95
	incongruent future *(phonologically ambiguous)*	Личный помощник президента хочет *ОТВЕТИТ$_{\mathrm{FUTURE}}$ на провокационный вопрос журналиста. *President's personal assistant wants to *ANSWER*$_{\mathrm{FUTURE}}$ *the journalist's provocative question.*	0.78	60.95
	incongruent past *(phonologically unambiguous)*	Личный помощник президента хочет *ОТВЕТИЛ$_{\mathrm{PAST}}$ на провокационный вопрос журналиста. *President's personal assistant wants to *ANSWER*$_{\mathrm{PAST}}$ *the journalist's provocative question.*	0.78	60.95
Control	congruent	Школьники начинают изучать иностранный ЯЗЫК$_{\mathrm{ACCUSATIVE}}$ с первого класса. *Students start learning a foreign LANGUAGE*$_{\mathrm{ACCUSATIVE}}$ *in the first grade.*	0.65	126.29
	incongruent case	Школьники начинают изучать иностранный ЯЗЫКУ$_{\mathrm{DATIVE}}$ с первого класса. *Students start learning a foreign LANGUAGE*$_{\mathrm{DATIVE}}$ *in the first grade.*	0.77	126.29
	incongruent semantic	Школьники начинают изучать иностранный ОВОЩ с первого класса. *Students start learning a foreign VEGETABLE in the first grade.*	0.65	139.46

Note: TW = target word

All experimental sentences were recorded with normal intonation at a normal speaking rate by a female native speaker of Russian and digitized at a sampling rate of 44 kHz. The sentences were recorded in triplets in a random order to eliminate any condition-specific prosodic patterns. Sound waveforms were examined and target word onsets and offsets were marked using Praat sound editing software (Boersma & Weenink, 2010). The target words were spliced across triplets (from congruent to incongruent conditions, and vice versa) to ensure that prosodic information and speaking rate are kept constant within each triplet but that there is no spurious effect of splicing itself.

5.3.3 Procedure

Participants were comfortably seated about 100 cm in front of the computer in a sound-attenuated room and instructed to move as little as possible. They were asked to listen to sentences attentively and understand them the best they could. They were warned that some sentences may sound strange. Sentences were presented through metal-free headphones at a comfortable volume for each individual. Each trial began with a beep tone lasting for 150 ms, followed by a 1000-ms silence period, then the auditorily presented sentence, and another silent period for 2000 ms, after which a question on the computer screen appeared. Each time the question asked the participants if the sentences sounded good. They indicated their response by pressing the “yes” or “no” button on the keyboard. The next trial started 3000 ms after the response was given. To ensure that subjects would not blink during and shortly after the presentation of the sentence, they were instructed to focus on the fixation point, which appeared on the computer screen simultaneously with the beep sound and remained there until the question was displayed. Participants were free to move their eyes or blink when the fixation point was not on the screen. Trials were presented in 9 blocks, between which participants could take short breaks. On average, the whole experiment lasted about 70 minutes. Prior to the experimental session, participants were given 10 practice trials with feedback to familiarize themselves with the task and were explained what constitutes a "good" and a "bad" sounding sentence. The sentences were presented through Matlab R2013a (Mathworks, USA), using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

5.3.4 EEG recordings

Raw EEG signal was recorded continuously using Neuroscan data acquisition system and SynAmps amplifier at a 1000-Hz sampling rate from 29 pure tin electrodes mounted in an electrode cap (Electro-cap International) at the following sites: midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2. Recordings were referenced online to the left mastoid and re-referenced offline to averaged mastoids. The vertical electro-oculogram (VEOG) was recorded from the electrodes placed above and below the left eye; the horizontal electro-oculogram (HEOG) was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5kΩ. The EEG and EOG recordings were amplified and digitized online at 1kHz with a bandpass filter of 0.1-100 Hz.

5.3.5 EEG data analysis

EEG data analysis was performed using EEGLAB v12 (Delorme and Makeig, 2004), an open source toolbox running under Matlab R2013a (Mathworks, USA). The data were epoched (−200 to 1400 ms) and baseline corrected (−200 ms to 0ms). An independent component analysis (ICA) with the runica Infomax algorithm provided by the EEGLAB toolbox was performed and the components corresponding to eye blinks, eye and muscle movement were removed from the EEG data. The data were further processed by an automatic peak-to-peak artifact rejection (rejection level ±100 μV) in order to remove any residual artifacts, resulting in 5.11% and 5.24% of discarded trials for L1 and L2 groups, respectively. For each participant, artifact-free trials were averaged into ERPs per each experimental condition (critical: congruent, incongruent future, incongruent past; control: congruent, incongruent case, incongruent semantic) for two time-locking points (target onset and target offset) for all electrodes. Weighted grand average ERPs for each participant group (L1 and L2) were computed. Grand averaged ERPs were filtered off-line with a 20 Hz low-pass filter for plotting purposes, but all statistical analyses were computed on unfiltered data.

5.3.6 Results

5.3.6.1 Behavioral results

Listeners' judgment of goodness of sentences in different conditions was evaluated along two parameters: error rate and reaction time (see Table 11). For the error rate analysis, a two-way ANOVA with condition (3 levels: critical, control, or filler) as a within-subjects factor, and language group (2 levels: L1 or L2) as a between-subjects factor yielded a significant interaction between condition and language group ($F(2, 9714) = 62.4$, $p < 0.001$, $\eta^2 = 0.013$), a significant main effect of condition ($F(2, 9714) = 51.96$, $p < 0.001$, $\eta^2 = 0.01$), and a significant main effect of language group ($F(1, 9714) = 1122$, $p < 0.001$, $\eta^2 = 0.10$). While L1 listeners' accuracy of judgment of sentence goodness did not differ across the three conditions, L2 listeners misjudged sentence goodness in 31.5%

of the sentences in the critical condition, 22.2% in the filler, and 14.5% in the control conditions (error differences between the conditions were significant at $p < 0.001$). L2 listeners also made significantly more errors than L1 listeners in all respective conditions ($p < 0.001$).

Table 11. Mean error rate and reaction time latencies for L1 and L2 listeners in the sentence goodness task.

Condition	Language group			
	L1		L2	
	Error rate	RT	Error rate	RT
Critical	0.031 *(0.007)*	1026.36 *(101.32)*	0.315 *(0.015)*	1516.64 *(249.10)*
Control	0.026 *(0.004)*	1038.37 *(106.62)*	0.145 *(0.018)*	1395.12 *(267.72)*
Filler	0.019 *(0.004)*	1051.1 *(108.65)*	0.222 *(0.021)*	1779.62 *(406.61)*

Note: Standard errors are presented in brackets.

For the reaction time analysis, a similar two-way ANOVA with condition (3 levels: critical, control, or filler) as a within-subjects factor and language group (2 levels: L1 or L2) as a between-subjects factor was conducted. We observed a significant interaction between condition and language group ($F(2, 9714) = 7.85$, $p < 0.001$, $\eta^2 = 0.002$), a significant main effect of condition ($F(2, 9714) = 6.89$, $p < 0.01$, $\eta^2 = 0.001$), and a significant main effect of language group ($F(1, 9714) = 183.15$, $p < 0.001$, $\eta^2 = 0.02$). Post-hoc Tukey HSD tests revealed that, while L1 listeners' response latencies to questions did not differ significantly across the three conditions, L2 listeners were significantly slower in the filler ($M = 1779.62$ ms, $SE = 406.61$ ms) condition than in the control ($M = 1395.12$ ms, $SE = 267.72$ ms) or critical ($M = 1516.64$ ms, $SE = 249.1$ ms) condition (control and critical conditions did not differ significantly between each other). L2 listeners also responded to the questions significantly more slowly than L1 listeners in all respective conditions ($p < 0.001$). Separate one-way ANOVAs for the critical and control conditions yielded a significant effect of congruency condition on both error rate and reaction time latencies of L2 listeners (critical: $F_{ER}(2, 1347) = 502.6$, $p < 0.001$, $\eta^2 = 0.43$; $F_{RT}(2, 1347) = 29.73$, $p < 0.001$, $\eta^2 = 0.042$; control: F_{ER} $(2, 1347) = 14.24$, $p < 0.001$, $\eta^2 = 0.02$; $F_{RT}(2, 1347) = 15.66$, $p < 0.001$, $\eta^2 = 0.02$) (see Figures 11 and 12).

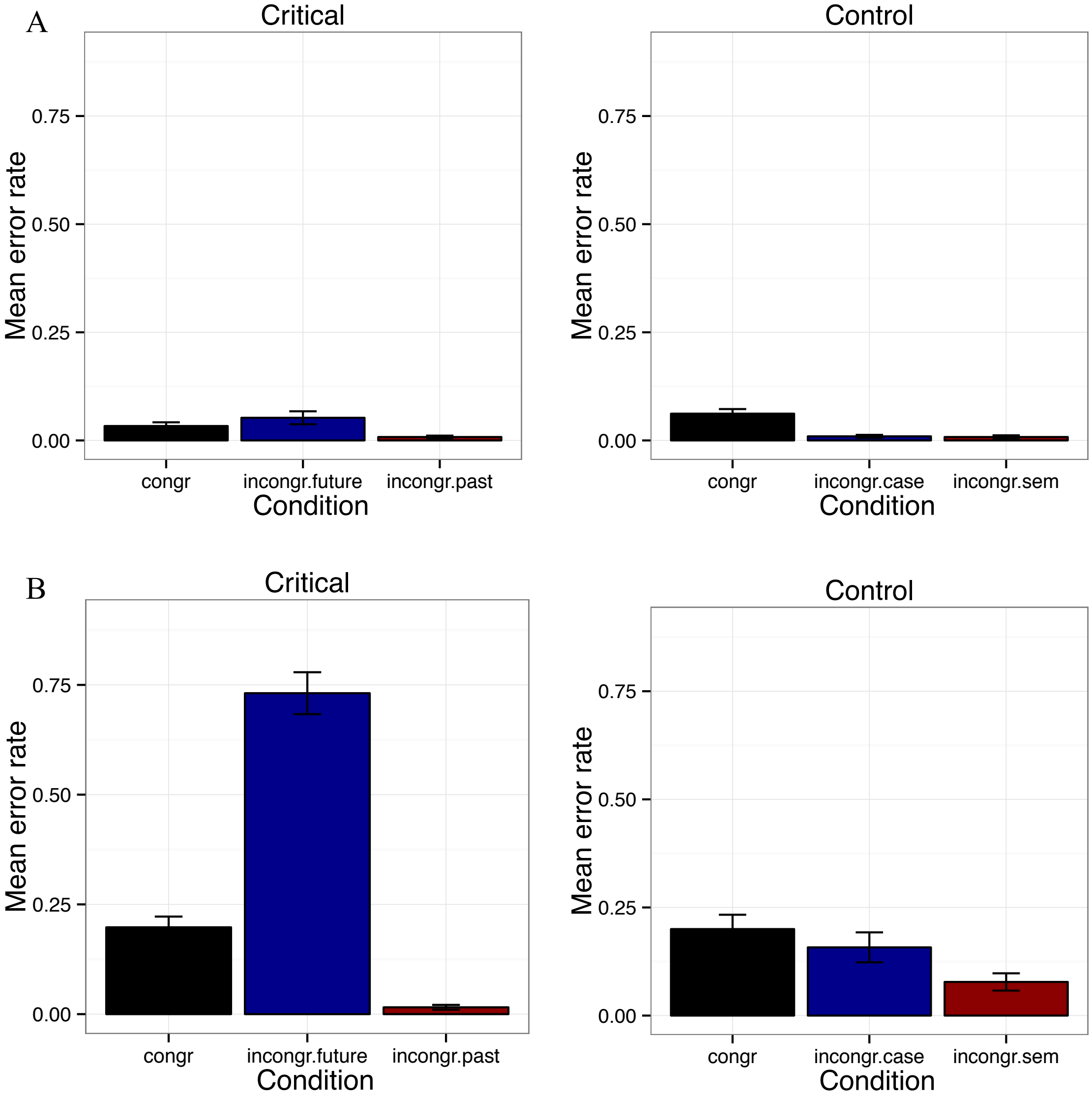


Figure 11. Mean error rate in the critical and the control conditions in (A) the L1 group and (B) the L2 group.

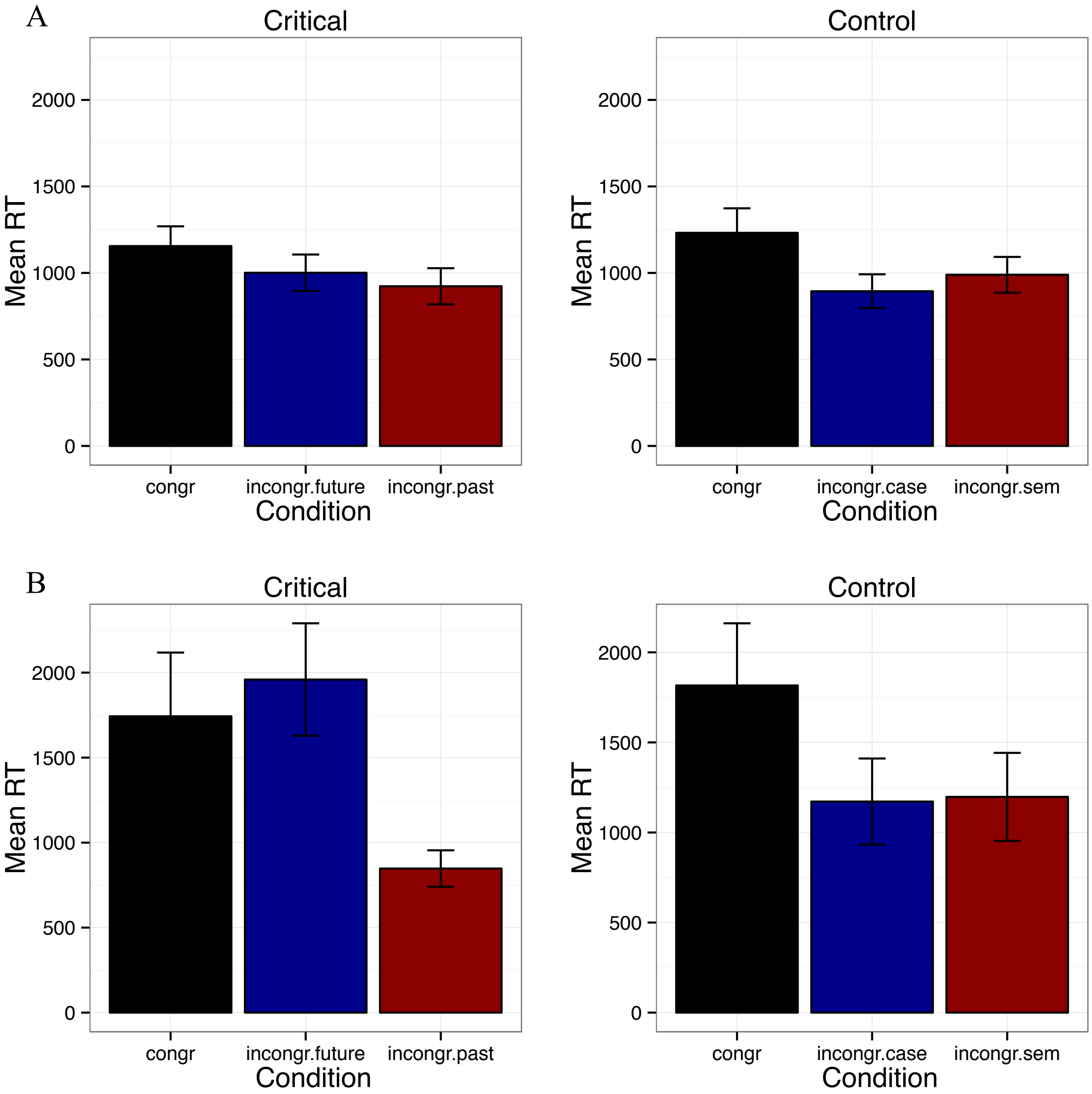


Figure 12. Mean RT in the critical and the control conditions in (A) the L1 group and (B) the L2 group.

In the control condition, post-hoc Tukey HSD comparisons showed that L2 listeners made significantly fewer errors ($p < 0.01$) in the incongruent semantic condition ($M = 0.07$, $SE = 0.02$) compared to the incongruent morphological (case mismatch) condition ($M = 0.16$, $SE = 0.03$) and congruent condition ($M = 0.2$, $SE = 0.03$) (the latter two did not differ significantly) indicating that spotting semantic violations was easier for them. Their response latencies in the control condition were significantly faster ($p < 0.001$) in both incongruent conditions (morphological: $M = 1171.63$, $SE = 239.09$, semantic: $M = 1197.55$, $SE = 244.43$) compared to the congruent condition ($M = 1816.17$, $SE = 344.87$), suggesting that by the time of the button press, the listeners have already identified a violation in the sentence.

In accordance with our predictions, the analysis of the critical condition showed that L2 listeners made significantly more errors ($p < 0.001$) in the incongruent condition involving a morphological violation of phonologically difficult forms (an infinitive form vs. a future-tense form) ($M = 0.73$, $SE = 0.05$) compared to the incongruent condition involving a morphological violation with phonologically easier forms (an infinitive form vs. a past-tense form) ($M = 0.02$, $SE = 0.005$). The congruent condition ($M = 0.2$, $SE = 0.02$) was significantly different from both incongruent conditions ($p < 0.001$). Reaction time data corroborates the observed error rate differences: L2 listeners were almost two times faster ($p < 0.001$) in identifying incongruent use of the verbs in the past tense ($M = 847.61$, $SE = 107.25$) compared to that of the verbs in the future tense ($M = 1959.32$, $SE = 330.76$).

5.3.6.2 ERP results

Statistical analyses on mean voltage amplitude were carried out on selected latency windows, which were determined after careful visual inspection of the grand average ERP waveforms for L1 and L2 groups: 200-600 ms for the N400 component (for semantic violations) and 800-1300 ms for the late P600 component (for morphological violations). Separate analyses were conducted for data from control and critical conditions since neither the target words nor the sentence structure were matched across these conditions by design. For the critical condition, repeated measures ANOVAs were performed on the 800-1300 ms window with one between-subjects variable (group: L1 vs. L2) and several within-subjects variables: congruency (3 levels: congruent, incongruent past, incongruent future), hemisphere (3 levels: left, midline, right), and anteriority (3 levels: anterior, central, posterior). For the control condition, repeated measures ANOVAs were performed on the 200-600 ms and 800-1300 ms window with one between-subjects variable (group: L1 vs. L2) and the same within-subjects variables, except for congruency, which had 2 separate levels (congruent vs. incongruent) in both semantic and morphological comparisons. A combination of the variables hemisphere and anteriority yielded 9 regions of interest (ROIs): left-anterior: F3, FC3, F7, FT7; midline-anterior: FZ, FCZ; right-anterior: F4, FC4, F8, FT8; left-central: C3, CP3, T7, TP7; midline-central: CZ, CPZ; right-central: C4, CP4, T8, TP8; left-posterior: P3, O1, P7; midline-posterior: PZ, OZ; right-posterior: P4, O2, P8). We will present the results for the control condition first, and then for the critical condition.

5.3.6.2.1 Control comparison

Semantic condition

Grand average ERPs time-locked to target word onsets demonstrate a clear negativity peaking at around 400 ms followed by a broadly distributed late positivity around 800-1300 ms in response to the semantic manipulation for both L1 and L2 listeners (Figure 13 and 14). The difference in mean amplitudes between congruent and incongruent conditions for the two latency windows yielded a topographic distribution characteristic of the N400 and P600 effects, respectively (Figure 19, top and middle).

Results from a repeated measures ANOVA on the ERP mean amplitude in the 200-600 ms latency range (N400 component) revealed a significant main effect of congruency (*F*(1, 34) = 13.427, *p* < 0.0001) and an interaction between hemisphere and anteriority (*F*(4, 136) = 3.58, *p* < 0.01). In both participant groups, the effect was bilaterally distributed over posterior and central sites (Figure 15.A). A direct comparison of L1 and L2 groups did not reveal significant differences (Table 12). Peak amplitude for the N400 response in the L1 group occurred around 361.7 ms (*SE* = 86.9 ms) from the stimulus onset and around 440.42 ms (*SE* = 77.5 ms) for the L2 group.

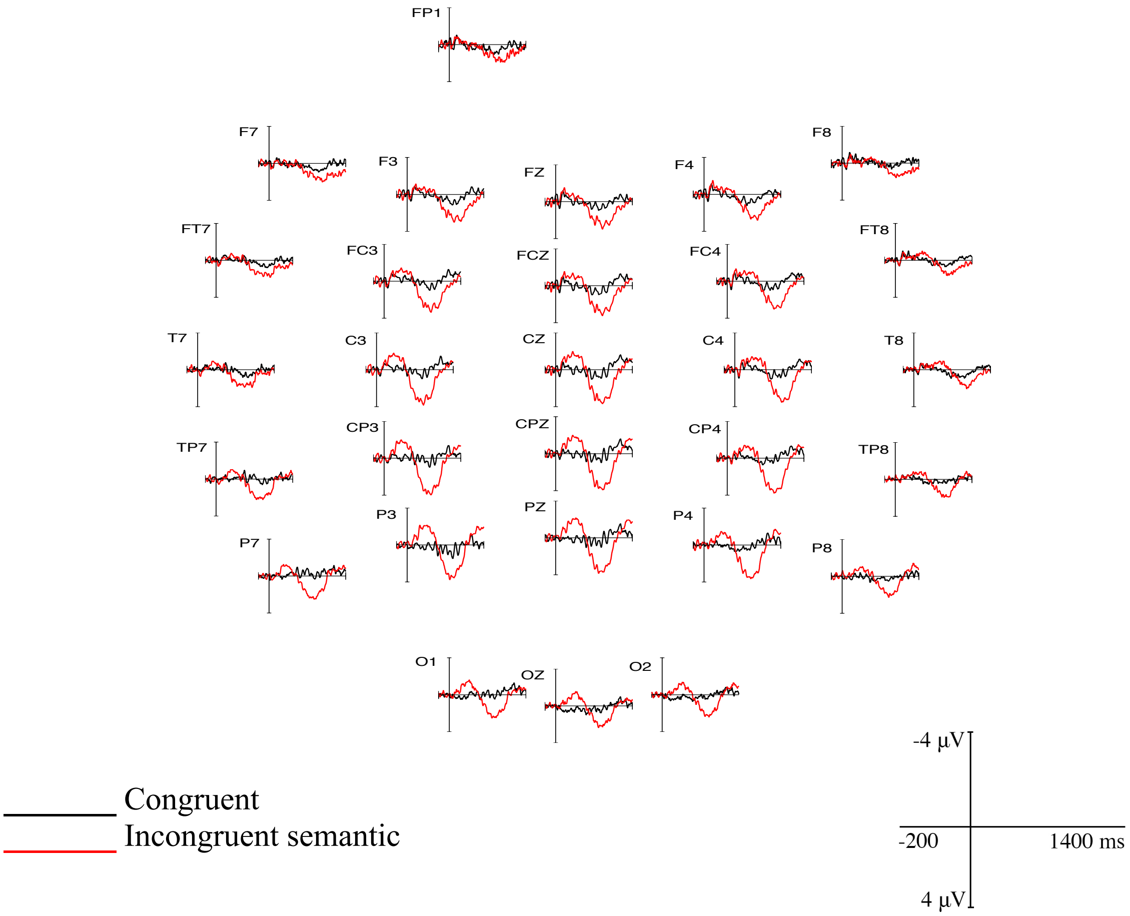


Figure 13. Grand average ERPs at the onset of the target word in congruent (black) and incongruent semantic (red) control condition for the L1 group. Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

With respect to the positivity seen in the later time window (800–1200 ms), significant effects of congruency (*F*(1, 34) = 12.86, $p < 0.01$), anteriority (*F*(2, 68) = 3.65, $p < 0.05$), as well as significant interactions for language and anteriority (*F*(2, 68) = 3.38, $p < 0.05$), congruency and hemisphere (*F*(2, 68) = 3.29, $p < 0.05$), and hemisphere and anteriority (*F*(4, 136) = 3.13, $p < 0.05$) were observed. ROI analysis revealed that the P600 effect to incongruent condition was largest at midline sites (Figure 15.B). Peak amplitude for the P600 response to incongruent sentences occurred around 955.24 ms from the stimulus onset (*SE* = 77.8 ms) in the L1 group, and around 1071 ms (*SE* = 79.3 ms) for the L2 group.

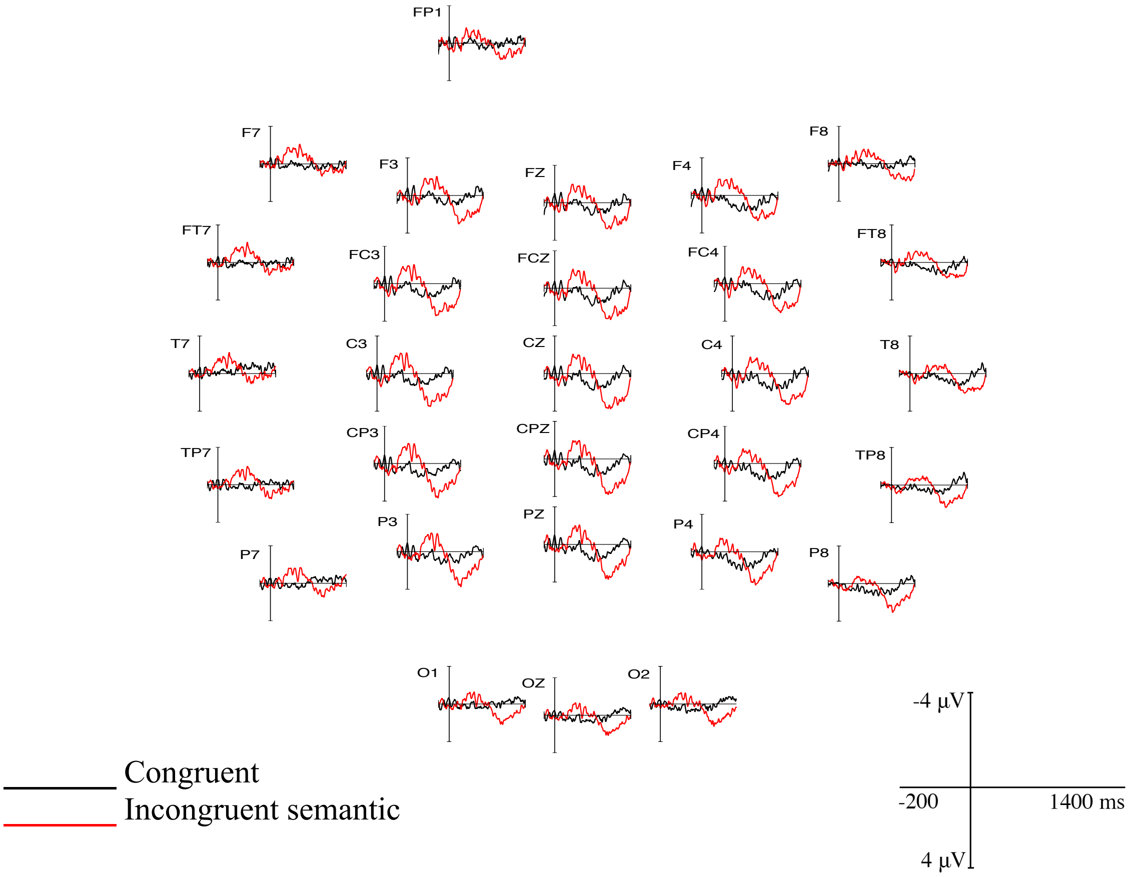


Figure 14. Grand average ERPs at the onset of the target word in congruent (black) and incongruent semantic (red) control condition for the L2 group. Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

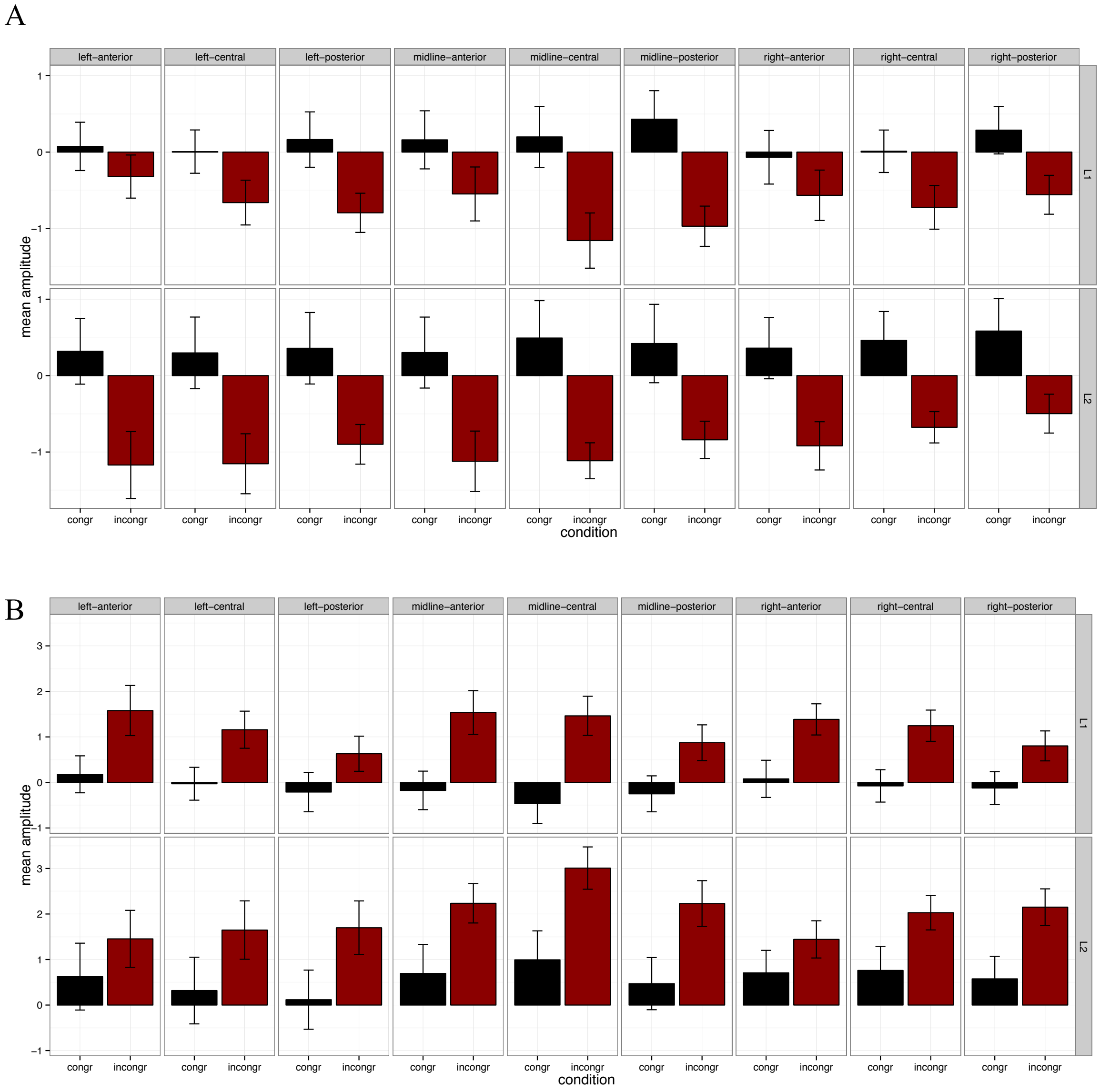


Figure 15. Average ERP amplitude for congruent (black) and incongruent (red) conditions in the control semantic condition across all regions of interest (ROIs) in the time window of (A) 200-600 ms and (B) 800-1300 ms for L1 and L2 groups.

Morphological condition

Grand average ERPs time-locked to target word onsets in the control morphological (case) condition are illustrated in Figure 16 and 17 for L1 and L2 groups, respectively. Both groups demonstrate a clear late positivity for the incongruent condition peaking at around 1040 (*SE* = 13 ms) for L1 group and 1040.5 (*SE* = 17.97 ms) for L2 group. The difference in mean amplitudes between congruent and incongruent conditions for the 800-1300 ms latency window has a topographic distribution characteristic of the P600 effects (Figure 19, bottom).

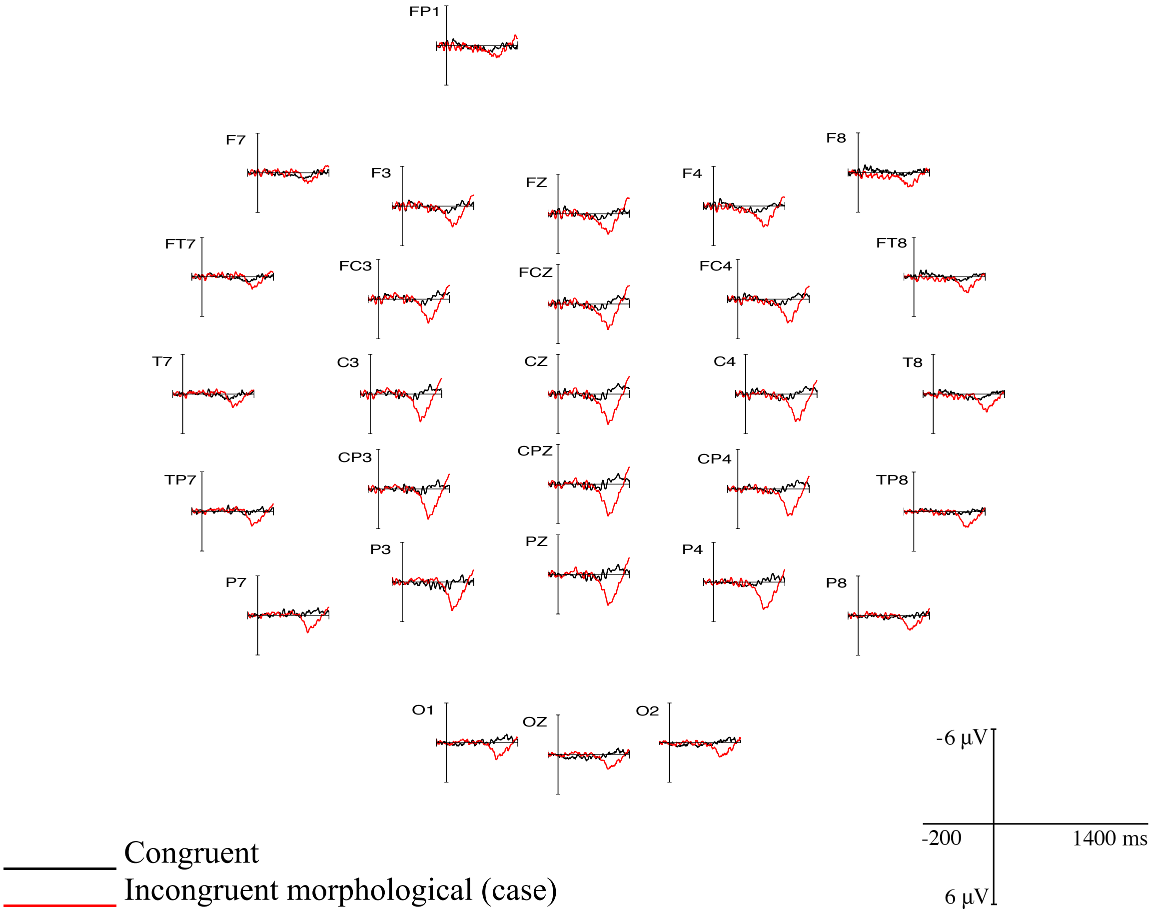


Figure 16. Grand average ERPs at the onset of the target word in congruent (black) and incongruent morphological (case) (red) control condition for the L1 group. Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

Results from a repeated measures ANOVA on the ERP mean amplitude yielded a significant main effect of congruency ($F(1, 34) = 11.88, p < 0.001$) and hemisphere ($F(2, 68) = 4.23, p < 0.05$), significant two-way interactions between congruency and hemisphere ($F(2, 68) = 4.27, p < 0.05$) and hemisphere and anteriority ($F(4, 136) = 3.01,$

$p < 0.05$), and a significant three-way interaction of language group, hemisphere and anteriority ($F(4, 136) = 3.11$, $p < 0.05$) (see Table 12). The analysis revealed that morphologically incongruent sentences elicited a particularly notable positivity at central locations (which was more pronounced in the L1 listener group) and a smaller (or absent) effect over the left than over the right hemisphere, especially in the L2 speaker group (Figure 18).

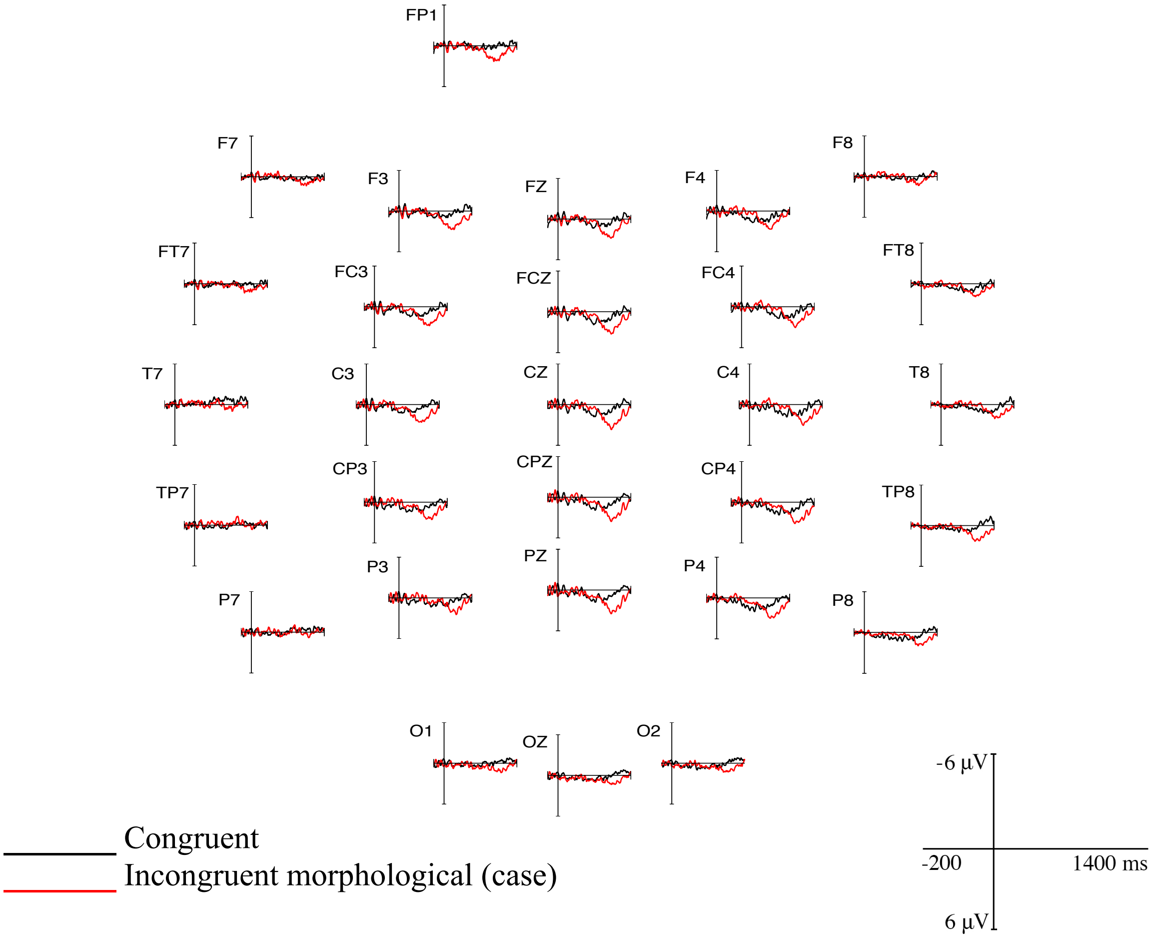


Figure 17. Grand average ERPs at the onset of the target word in congruent (black) and incongruent morphological (case) (red) control condition for the L2 group. Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

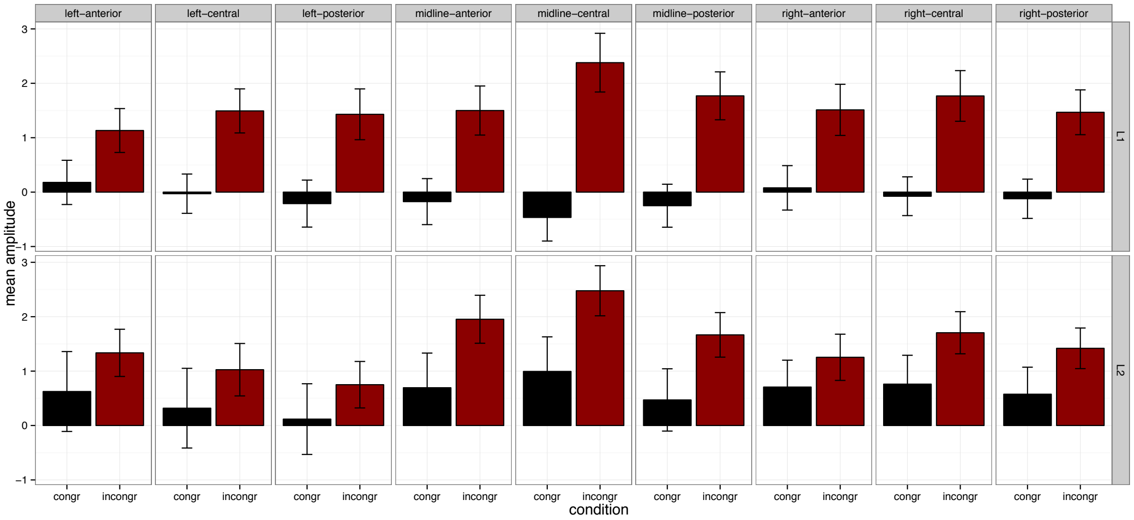


Figure 18. Average ERP amplitude for congruent (black) and incongruent (red) conditions in the control morphological (case) condition across all regions of interest (ROIs) in the time window of 800-1300 ms for L1 and L2 groups.

Table 12. *F*-tests and associated *p* values for main effects and interactions on mean ERP amplitudes in the control semantic condition for the 200-600 ms and 800-1300 ms windows and the control morphological (case) condition for the 800-1300 ms.

Effect	Df	Semantic				Morphological	
		200-600		800-1300		800-1300	
		F-test	*p* value	*F*-test	*p* value	*F*-test	*p* value
language	1, 34	0.00	0.98	2.04	0.16	0.33	0.57
congruency	1, 34	13.43	< 0.001	12.86	< 0.01	11.88	< 0.001
hemisphere	2, 68	0.81	0.45	1.09	0.34	4.23	< 0.05
anteriority	2, 68	0.87	0.42	3.65	< 0.05	1.62	0.21
language x congruency	1, 34	1.00	0.32	0.00	0.96	1.02	0.32
language x hemisphere	2, 68	1.64	0.20	2.87	0.06	2.25	0.11
language x anteriority	2, 68	0.97	0.39	3.38	< 0.05	0.41	0.67
congruency x hemisphere	2, 68	1.56	0.22	3.29	< 0.05	4.27	< 0.05
congruency x anteriority	2, 68	0.41	0.66	0.24	0.79	1.75	0.18
hemisphere x anteriority	4, 136	3.58	< 0.01	3.13	< 0.05	3.01	< 0.05
language x congruency x hemisphere	2, 68	0.63	0.54	0.16	0.85	0.13	0.88
language x congruency x anteriority	2, 68	0.94	0.40	1.36	0.26	0.41	0.67
language x hemisphere x anteriority	4, 136	0.84	0.50	2.19	0.07	3.11	< 0.02
congruency x hemisphere x anteriority	4, 136	1.04	0.39	1.63	0.17	1.47	0.22
language x congruency x hemisphere x anteriority	4, 136	0.28	0.89	0.42	0.79	1.43	0.23

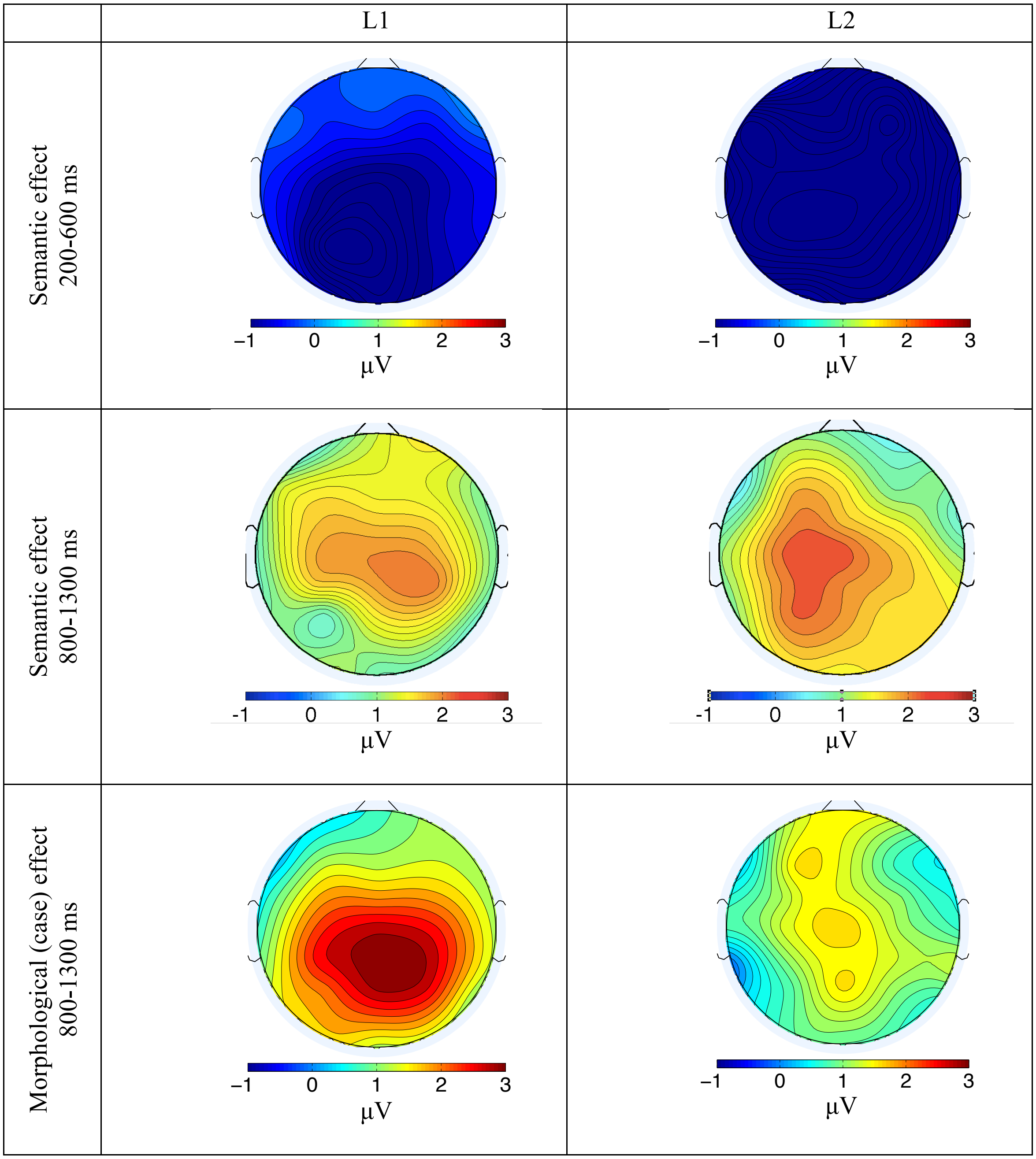


Figure 19. Topographic distribution of the ERP effects in the 200-600 ms (top) and 800-1300 ms (middle) latency windows for the control semantic condition and in the 800-1300 ms (bottom) window for the control morphological (case) condition for L1 and L2 groups.

5.3.6.2.2 Critical condition

Time-locking to word onset

Based on the visual inspection of the grand average ERPs time-locked to target word onsets in the critical condition, a clear late positivity (late P600 with a peak around 1000 ms) for both incongruent conditions (incongruent past and incongruent future) was elicited in the L1 listener group, whereas L2 group demonstrated a comparable positivity only for the incongruent past condition, but not incongruent future condition (compare Figure 20 and 21 for L1 and L2 groups, respectively). The topographic distribution of the ERP effects in the L1 and L2 groups is consistent with this observation: while in the L1 group the difference in mean amplitudes between congruent and both incongruent (past and future) conditions for the 800-1300 ms latency window has a clear positive centro-parietal distribution characteristic of the P600 component, in the L2 group it is absent for the incongruent future condition (Figure 21). A similar pattern is evident from the observation of the distribution of average amplitudes across different ROIs (Figure 23). In the L1 group, the ERP effect was bilaterally and centro-parietally distributed for both the incongruent past and the incongruent future conditions whereas the L2 group showed a similar distribution of the P600 effect only for the incongruent past condition.

The differences in the elicited ERP components in the two groups of participants were also confirmed by the statistical analyses. An omnibus repeated measures ANOVA on the ERP mean amplitude yielded significant main effects of congruency ($F(2, 68) = 16.34, p < 0.001$), hemisphere ($F(2, 68) = 15.9, p < 0.001$), anteriority ($F(2, 68) = 28.75, p < 0.001$), significant two-way interactions between congruency and hemisphere ($F(4, 136) = 5.9, p < 0.01$), hemisphere and anteriority ($F(4, 136) = 4.71, p < 0.01$), and, most importantly, between language and congruency ($F(2, 68) = 6.18$, $p < 0.01$). A three-way interaction of language, hemisphere and congruency ($F(4, 136) = 4.68$, $p < 0.01$) as well as congruency, hemisphere and anteriority also came out significant ($F(8, 272) = 2.17$, $p < 0.05$) (see Table 13).

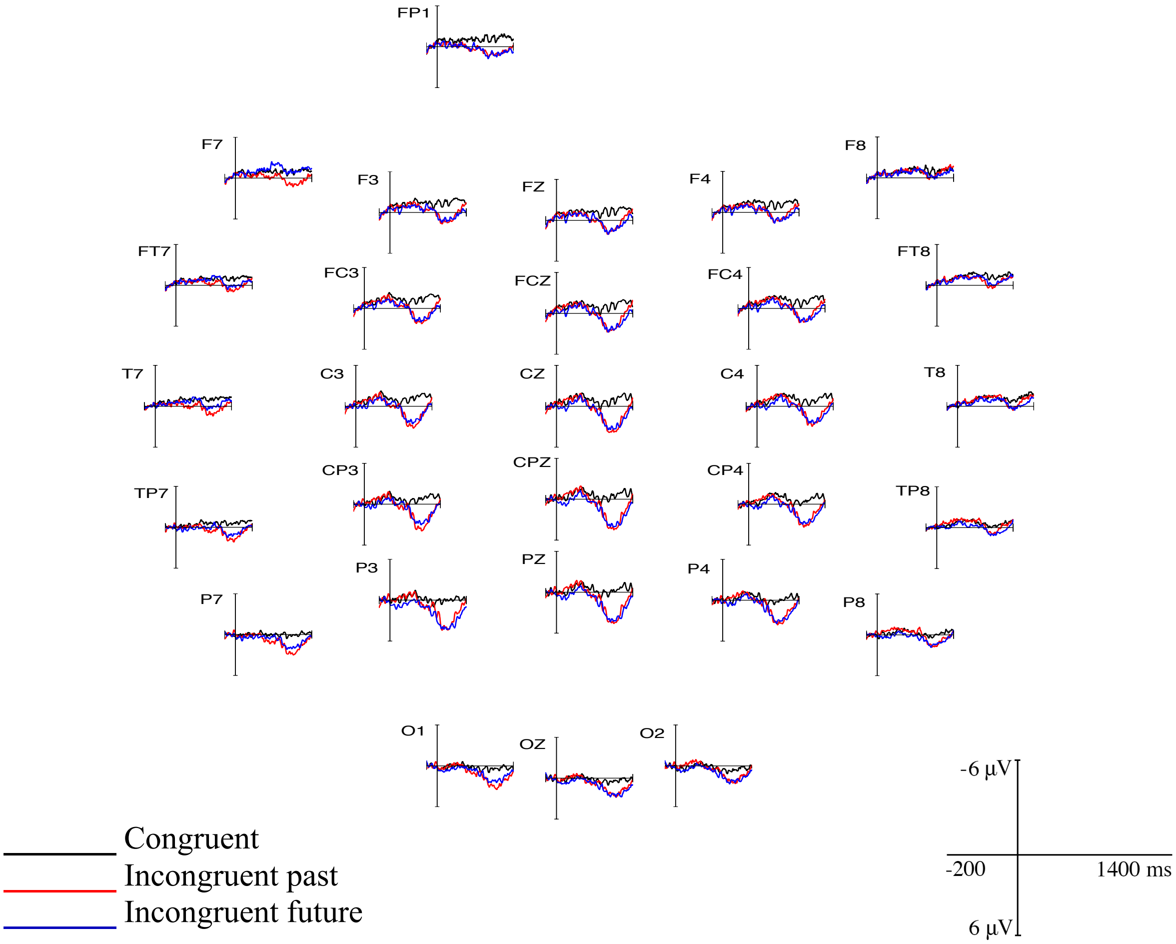


Figure 20. Grand average ERPs at the onset of the target word in congruent (black), incongruent past (red) and incongruent future (blue) conditions in the critical condition for the L1 group. Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

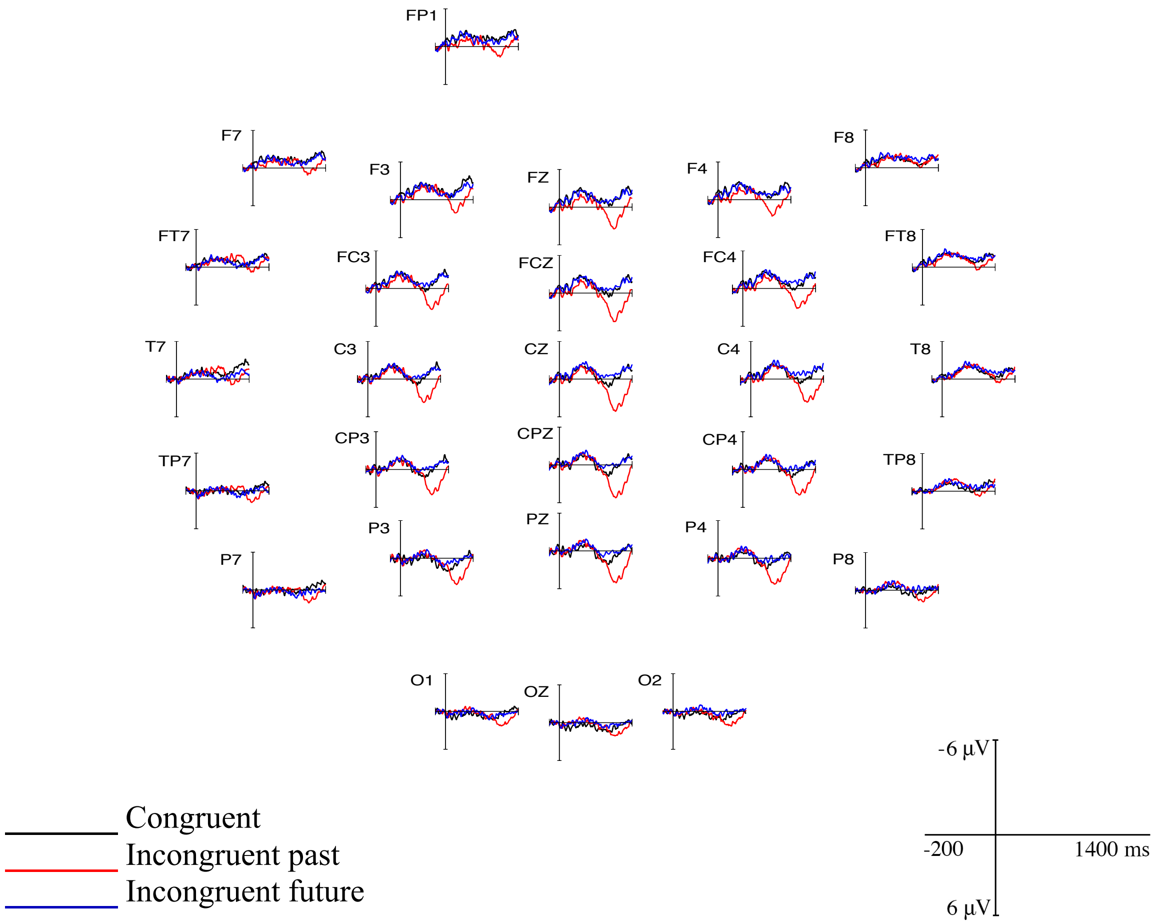


Figure 21. Grand average ERPs at the onset of the target word in congruent (black), incongruent past (red) and incongruent future (blue) conditions in the critical condition for the L2 group. Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, μV)

Separate ANOVAs with Tukey HSD post-hoc comparisons of the conditions of interest revealed that mean amplitude in response to the congruent targets was significantly smaller (*M* = -0.64, *SE* = 0.26) compared to the incongruent past-tense forms (*M* = 1.41, *SE* = 0.38) and the incongruent future-tense forms (*M* = 1.32, *SE* = 0.43) in the L1 group ($p < 0.001$), but did not differ significantly for the latter two. For the L2 comprehenders, there was a significant difference between congruent (*M* = -0.41, *SE* = 0.52) and incongruent past conditions (*M* = 1.2, *SE* = 0.59) ($p < 0.001$), but no statistical difference between congruent and incongruent future (*M* = -0.66, *SE* = 0.35) conditions. Mean amplitude for the congruent condition did not differ significantly between L1 and L2 listeners. Neither was there a statistical difference for incongruent past condition between L1 and L2 participants, but the differences in mean amplitude for the incongruent future condition were significant between the two groups ($p < 0.001$).

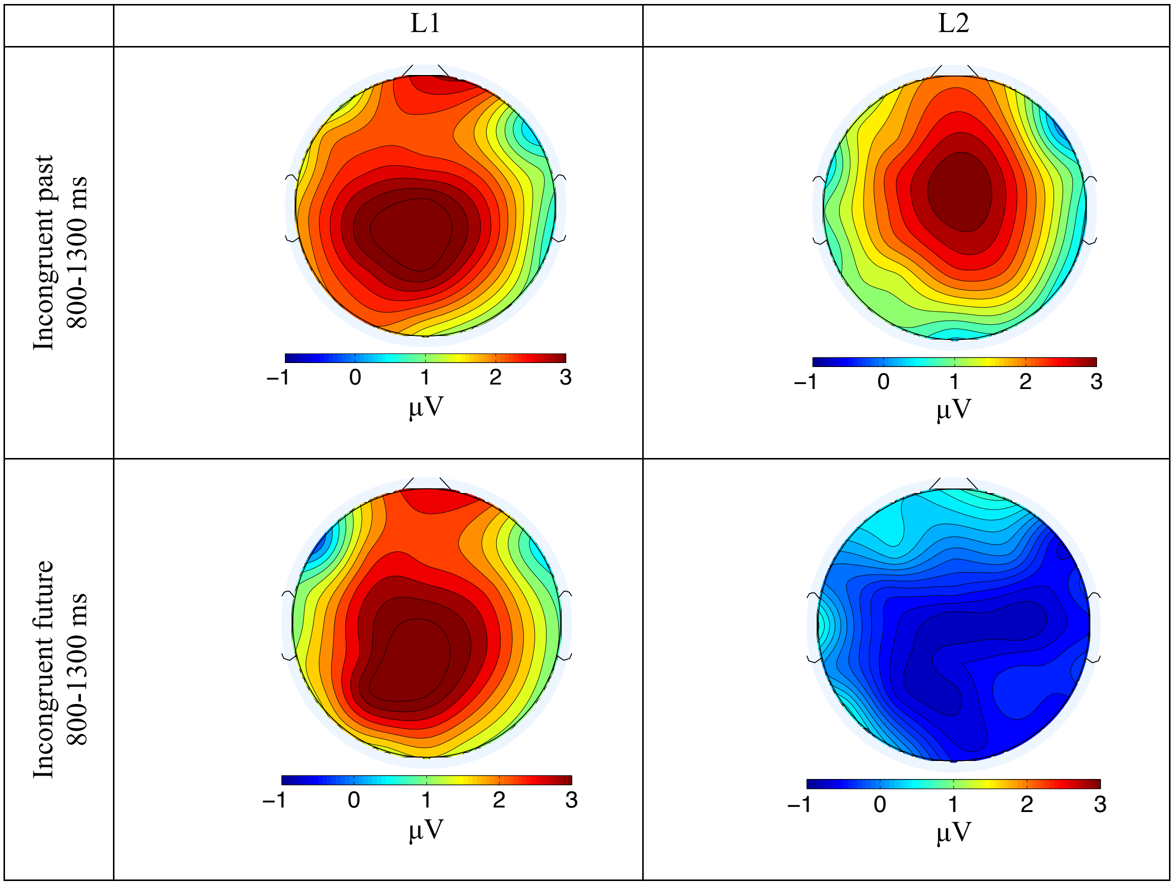


Figure 22. Topographic distribution of the ERP effects in the 800-1300 ms latency windows for the critical incongruent past (top) and critical incongruent future (bottom) conditions for L1 and L2 groups.

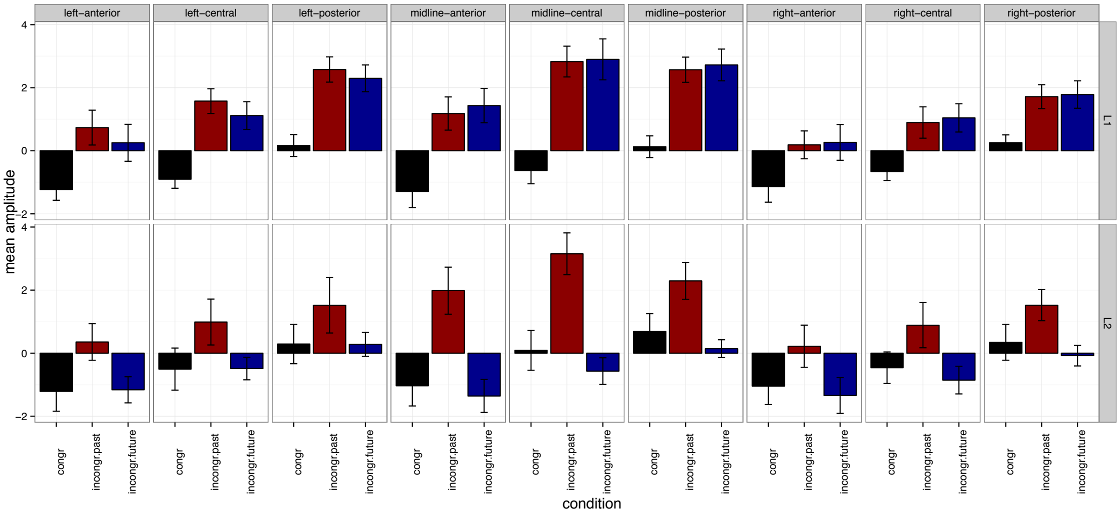


Figure 23. Average ERP (time-locked to target word onset) amplitude for congruent (black), incongruent past (red), and incongruent future (blue) conditions in the critical condition across all regions of interest (ROIs) in the time window of 800-1300 ms for L1 and L2 groups.

Time-locking to word offset

Because the disambiguation point between the three verbal forms used in the critical condition of the present study falls on the last phoneme of the word (*ответить* – *ответил* – *ответит*), some nuances of ERP components may be smeared when the waveforms are time-locked to word onsets (e.g., due to differences in word duration). Therefore, an additional analysis was performed for the ERP waveforms time-locked to target word offsets. Figures 24 and 25 display the grand average ERPs for the congruent, incongruent past and incongruent future conditions for the L1 and L2 groups, respectively. Based on the visual inspection of the grand average ERPs time-locked to word offsets, a clear early positivity in the 100-600ms latency window followed by a pronounced late negativity in the 600-1300 ms window are observed for both incongruent conditions (incongruent past and incongruent future) in the L1 group. L2 group demonstrated a similar pattern of ERP response, except for the incongruent future condition. Scalp topography of the ERP effects (Figure 26) as well as the ROI analysis (Figure 27) indicate that the ERP effects are mostly pronounced in the centro-parietal area.

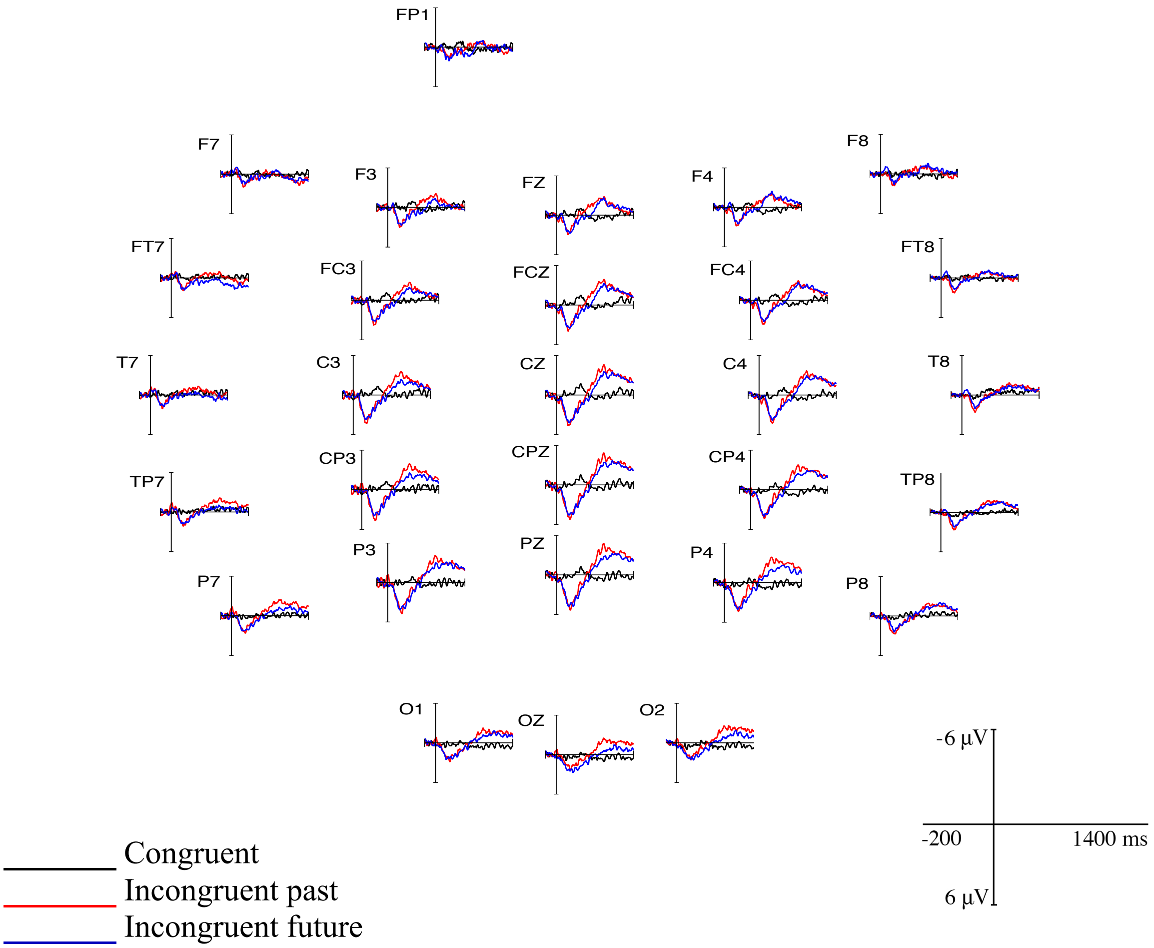


Figure 24. Grand average ERPs at the offset of the target word in congruent (black), incongruent past (red) and incongruent future (blue) conditions in the critical condition for the L1 group. Time 0 is the offset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

Again, the differences in the elicited ERP components in the two groups of participants were confirmed by the statistical analyses. In the 100-600 ms window, an omnibus repeated measures ANOVA on mean amplitudes yielded significant main effects of congruency ($F(2, 68) = 32.04$, $p < 0.001$) and hemisphere ($F(2, 68) = 6.27$, $p <$

0.01). Importantly, the interaction between language and congruency ($F(2, 68) = 13.03$, $p < 0.001$) was significant, as well as interactions between congruency and hemisphere ($F(4, 136) = 6.62$, $p < 0.001$), hemisphere and anteriority ($F(4, 136) = 5.8$, $p < 0.001$), and congruency and anteriority ($F(4, 136) = 3.59$, $p < 0.01$). Three-way interactions of language, hemisphere and congruency ($F(4, 136) = 2.94$, $p < 0.05$) as well as congruency, hemisphere and anteriority also came out significant ($F(8, 272) = 3.88$, $p < 0.01$). Finally, a four-way interaction between language, congruency, hemisphere and anteriority was found significant ($F(8, 272) = 2.55$, $p < 0.05$) (see Table 13).

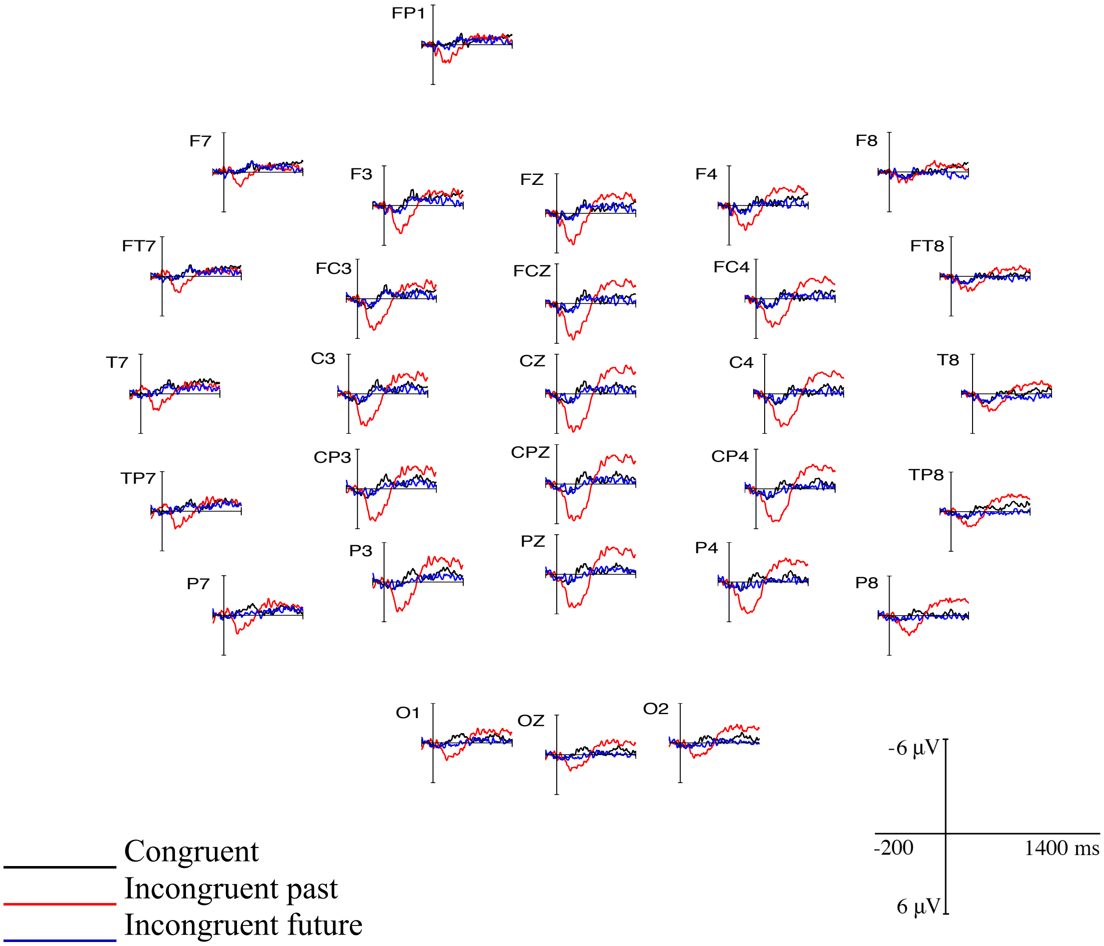
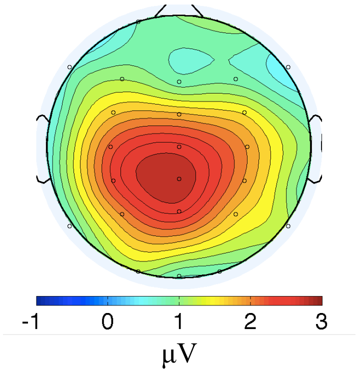
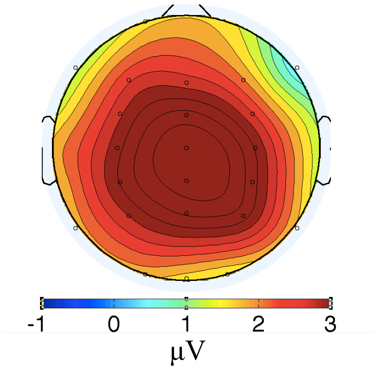
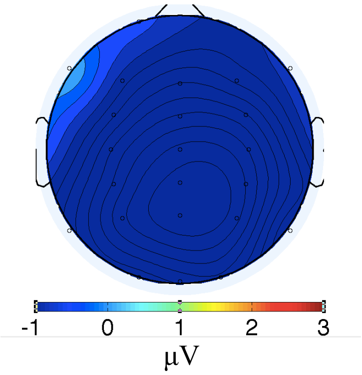
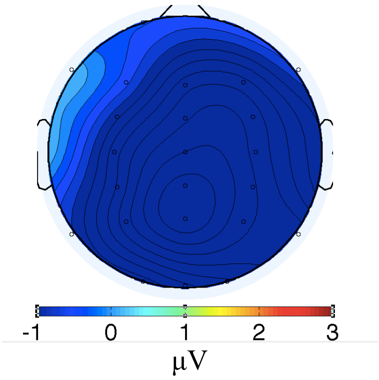
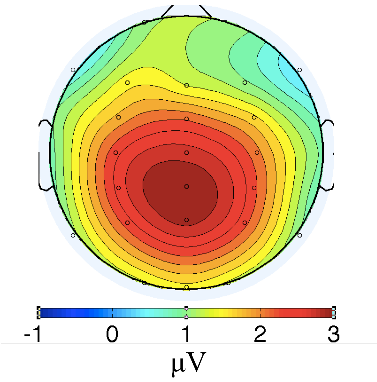
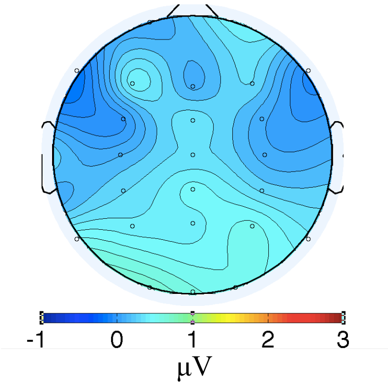


Figure 25. Grand average ERPs at the offset of the target word in congruent (black), incongruent past (red) and incongruent future (blue) conditions in the critical condition for the L2 group. Time 0 is the offset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

	L1	L2
Incongruent past 100-600 ms		
Incongruent past 600-1300 ms		
Incongruent future 100-600 ms		

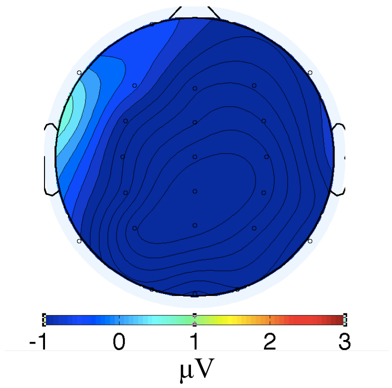
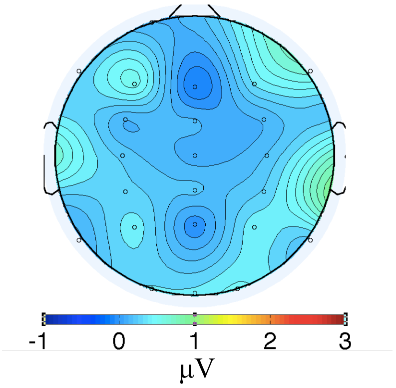
Incongruent future 600-1300 ms		

Figure 26. Topographic distribution of the ERP effects in the incongruent past and incongruent future conditions in the 100-600 ms and 600-1300 ms latency windows for L1 and L2 groups.

In the 600-1330 ms latency window, an omnibus repeated measures ANOVA also yielded significant main effects of congruency ($F(2, 68) = 13.31$, $p < 0.001$) and hemisphere ($F(2, 68) = 6.36$, $p < 0.01$). Additionally, a significant effect of anteriority was observed ($F(2, 68) = 8.92$, $p < 0.001$). Importantly, the interaction between language and congruency ($F(2, 68) = 4.14$, $p < 0.05$) was significant again, as well as interactions between congruency and hemisphere ($F(4, 136) = 7.55$, $p < 0.001$), hemisphere and anteriority ($F(4, 136) = 6.45$, $p < 0.001$), and congruency and anteriority ($F(4, 136) = 5.27$, $p < 0.01$). Three-way interactions of language, hemisphere and congruency ($F(4, 136) = 3.04$, $p < 0.05$) as well as congruency, hemisphere and anteriority also came out significant ($F(8, 272) = 3.4$, $p < 0.01$) (see Table 13).

Separate ANOVAs with Tukey HSD post-hoc comparisons of the conditions of interest revealed that, in the L1 group, mean ERP amplitude in the 100-600 ms window was significantly smaller in the congruent condition ($M = -0.09$, $SE = 0.27$) compared to the incongruent past condition ($M = 1.31$, $SE = 0.26$) and the incongruent future condition ($M = 1.51$, $SE = 0.26$) ($p < 0.001$). In the same time window L2 participants exhibited a significantly more positive amplitude in the incongruent past condition (*M* = 2.33, *SE* = 0.38) compared to the congruent (*M* = -0.03, *SE* = 0.39) and incongruent future (*M* = 0.2, *SE* = 0.21) conditions. The latter two were not significantly different. There were no significant differences between L1 and L2 groups as far as the congruent condition is concerned, but the positivity demonstrated by the L2 participants was greater in the incongruent past condition ($p < 0.01$) and smaller in the incongruent future condition ($p < 0.01$) compared to that demonstrated by L1 participants in respective conditions.

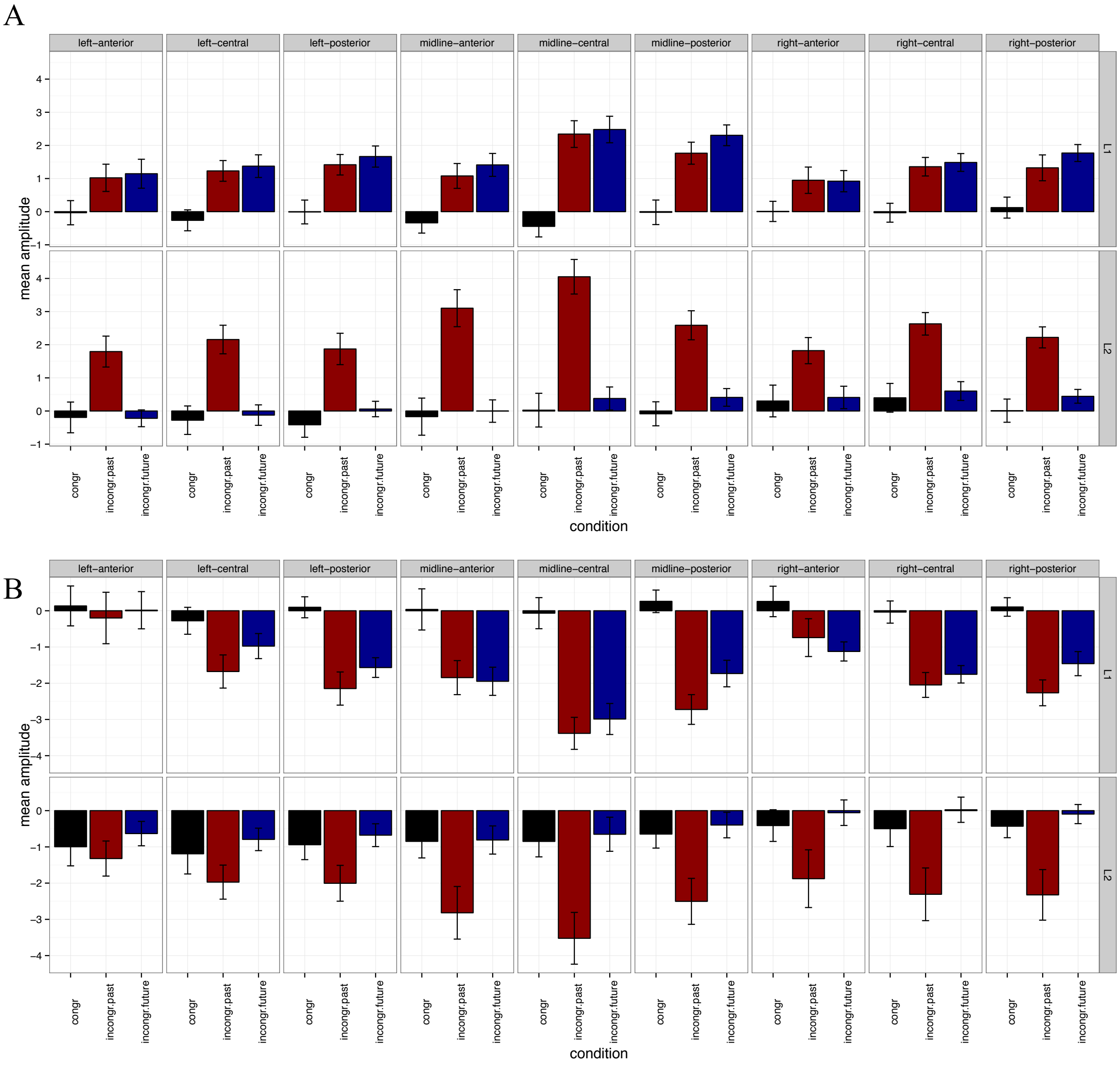


Figure 27. Average ERP (time-locked to target word offset) amplitude for congruent (black), incongruent past (red), and incongruent future (blue) conditions in the critical

condition across all regions of interest (ROIs) in the time window of (A) 100-600 ms and (B) 600-1300 ms for L1 and L2 groups.

A similar pattern of L1-L2 comparisons was obtained for the 600-1300 ms time window. L1 listeners demonstrated a more negative amplitude in the incongruent past (*M* = -1.77, *SE* = 0.33) and the incongruent future (*M* = -1.34, *SE* = 0.25) compared to the congruent (*M* = 0.05, *SE* = 0.33) conditions ($p < 0.01$). In contrast, L2 listener group showed a significantly more pronounced negativity only in the incongruent past (*M* = -2.16, *SE* = 0.57) compared to the incongruent future (*M* = -0.42, *SE* = 0.27) and the congruent (*M* = -0.76, *SE* = 0.4) conditions ($p < 0.01$). The differences in mean amplitudes between L1 and L2 groups were significant in all conditions: congruent ($p < 0.01$), incongruent past ($p < 0.05$) and incongruent future ($p < 0.001$).

Table 13. *F*-tests and associated *p* values for main effects and interactions on mean ERP amplitudes (time-locked to target words' onsets and offsets) in the critical condition for the 100-600 ms and 600-1300 ms windows.

Effect	Df	Onset		Offset			
		800-1300		100-600		600-1300	
		F-test	*p* value	*F*-test	*p* value	*F*-test	*p* value
language	1, 34	2.01	0.16	0.05	0.82	0.01	0.76
congruency	2, 68	16.35	< 0.001	32.04	< 0.001	13.31	< 0.001
hemisphere	2, 68	15.92	< 0.001	6.27	< 0.01	6.36	< 0.01
anteriority	2, 68	28.75	< 0.001	2.82	0.067	8.92	< 0.001
language x congruency	2, 68	6.18	< 0.01	13.03	< 0.001	4.14	< 0.05
language x hemisphere	2, 68	0.18	0.83	1.84	0.17	2.03	0.14
language x anteriority	2, 68	0.22	0.81	1.19	0.31	3.01	0.06
congruency x hemisphere	4, 136	5.91	< 0.001	6.62	< 0.001	7.55	< 0.001
congruency x anteriority	4, 136	0.65	0.63	3.59	< 0.01	5.27	< 0.001
hemisphere x anteriority	4, 136	4.73	< 0.01	5.81	< 0.001	6.45	< 0.001
language x congruency x	4, 136	4.68	< 0.01	2.94	< 0.05	3.04	< 0.05

hemisphere							
language x congruency x anteriority	4, 136	0.63	0.64	0.56	0.69	1.34	0.26
language x hemisphere x anteriority	4, 136	0.35	0.84	0.73	0.57	0.96	0.43
congruency x hemisphere x anteriority	8, 272	2.17	< 0.05	3.88	< 0.001	3.4	< 0.001
language x congruency x hemisphere x anteriority	8, 272	1.33	0.23	2.55	< 0.05	1.77	0.08

5.4 Discussion

The present study compared native and nonnative morpho-phonological processing during auditory sentence comprehension in Russian. Of particular interest was a situation when several morphological inflectional forms can be distinguished on the basis of a phonological contrast in the same word position (e.g., *ответить* (answer$_{\text{INF}}$)—*ответит* (answer$_{\text{FUTURE}}$)—*ответил* (answer$_{\text{PAST}}$)), where some contrasts may present a perceptual difficulty for L2 comprehenders (as in *т* – *ть* in Russian). Although the main goal of the study was to examine the type and the time-course of the ERP response evoked by these kinds of morpho-phonological violations (phonologically ambiguous or unambiguous for L2 listeners), a control condition was also included, which involved a semantic violation and a morphological violation of case marking.

5.4.1 Control condition

5.4.1.1 Semantic violation

Behavioral results showed that although L2 listeners made more errors in the control semantic condition compared to L1 subjects, their error rate was still quite low (7%). The fact that their response latencies were faster in the incongruent semantic compared to the congruent condition indicates that they were able to spot semantic incongruences before the button press.

With regard to the ERP data, semantic violations in the present study elicited an amplitude modulation of the N400 component, which was bilaterally and mostly centro-parietally distributed in both L1 and L2 listeners. The N400 mean amplitude did not differ statistically across the two groups, although the N400 peak latency was about 80 ms delayed in the L2 group. The observation of the N400 component to semantic violations is in line with some previous L1 and L2 studies, both visual (e.g., Ardal et al., 1990; Moreno and Kutas, 2005; Weber-Fox et al., 2003) and auditory (e.g., FitzPatrick and Indefrey, 2007; Hahne, 2001; Hahne and Friederici, 2001; Holcomb and Neville, 1990; Mueller et al., 2005). Since it is usually considered to serve as an index of semantic integrative and predictive mechanisms in sentence comprehension (Kutas and Federmeier, 2000), our findings suggest that there are more similarities than differences between L1 and L2 speakers in terms of the underlying mechanisms of lexical–semantic processing (see also Moreno, Rodríguez-Fornell, and Laine, 2008; Mueller, 2005; 2006).

Quite unexpectedly, the N400 component in the semantic condition was accompanied by a subsequent widely distributed centro–parietal positivity between 800 and 1300 ms post-stimulus onset in both participant groups. It was very similar to a late

P600 component in terms of its morphology and distribution, which is usually elicited in response to syntactic violations. Although it is not common, a number of previous studies have also observed an N400 followed by a P600 effect in response to semantic violations (e.g., Faustmann, Murdoch, Finnigan, and Copland, 2005; Friederici and Frisch, 2000; Gunter, Stowe, Mulder, 1997; Hoeks, Stowe, Doedens, 2004; Kuperberg, Sitnikova, Caplan, and Holcomb, 2003; Münte, Heinze, Matzke, Wieringa, and Johannes, 1998; van Herten, Kolk, and Chwilla, 2005). It has been proposed that the functional definition of the P600 as an index of purely morphosyntactic processing is too restrictive. Rather, it should reflect a more general language-related reanalysis processes based not only on syntactic, but also semantic and possibly other linguistic aspects of the sentence (Faustmann et al., 2005; Gunter et al., 1997; Münte et al., 1998), or the overall monitoring, reprocessing and repair of the initial sentence interpretation (Van Herten et al., 2005).

5.4.1.2 Morphological violation (case marking)

Similarly to the control semantic condition, L2 listeners demonstrated a higher error rate (14%) in the sentence goodness task than the L1 listeners. Their response latencies were faster in the incongruent compared to the congruent condition suggesting that they were able to spot morphological incongruences before the button press.

Examination of the ERP data revealed that morphological violations of case marking in Russian masculine singular nouns (the dative case in place of the expected accusative case, as in *язык-Ø* (language$_{\text{ACCUSATIVE}}$) – *язык-у* (language$_{\text{DATIVE}}$)) produce a clear late positivity peaking at around 1040 ms post-stimulus onset for both L1 and L2 groups. This effect was largest at centro-parietal sites and had a topographic distribution characteristic of the P600 component. No early negativities were elicited for either L1 or L2 group.

Given the functional interpretation of the P600 component in the studies examining L1 morphosyntactic processing, i.e., that it reflects secondary, more controlled morphosyntactic processes such as integration, revision, and reanalysis (e.g., Friederici, 1995; Hagoort, Brown, and Groothusen, 1993; Osterhout and Holcomb, 1992), it was expected to be elicited in response to the morphological violation of case use in the L1 group.

The predictions for the L2 group were not so obvious due to the existence of a large body of controversial empirical evidence on L2 acquisition of inflectional morphology (e.g., see Clahsen, Balkhair, Schutter, and Cunnings, 2013; Clahsen and Felser, 2006a; 2006b; Gor and Jackson, 2013). It is mostly agreed that L2 processing of inflected words is more effortful and prone to errors unlike that of L1 speakers, so it was not clear whether L2 listeners would be able to build online morphological predictions about case-inflected nominal forms. Besides, some previous ERP studies that compared acquisition of verbal and nominal agreement by L2 leaners reported that nominal number concord errors failed to produce reliable differences in the ERP trace while violations of tense use evoked a P600 effect, although it was reduced and had an atypical distribution. (Tockowitz and MacWhinney, 2005). Osterhout and his colleagues (2004; 2006) found that a P600 response to verbal agreement violations is elicited as learners' proficiency grows, but they did not observe any effect of nominal number agreement violations. The authors argued that L2 learners' differences in response to nominal versus verbal violations could be due to the similarity of features across languages. For example, for

English learners of French or Spanish, nominal agreement would be more difficult to acquire because it is not instantiated in English, as it is in French and Spanish, whereas all three languages share common features of tense and verbal agreement.

The results of the present study are especially noteworthy because they show that English speakers of L2 Russian are sensitive to violations in the nominal case use, even though English does not have a comparably complex nominal case system (except for the distinction between possessive/non-possessive nouns), as does Russian. In contrast to some previous studies on L1-L2 morphosyntactic processing (e.g., Hahne and Friederici, 2001; Weber-Fox and Neville, 1996), our results suggest that L2 listeners repair or reanalyze incorrectly inflected word forms before integrating them with the rest of the context. This finding can be interpreted as indirect evidence of L2 speakers' automatic processing of morphological decomposition into root + inflection and their sensitivity to morphological cues during sentence comprehension. They rely on the same higher-order processing mechanisms as do native speakers in listening comprehension, and are able to incorporate grammaticalized morphological knowledge into the online comprehension system.

In general, electrophysiological responses in L2 populations have been shown to be strongly modulated by learners' proficiency level (e.g., Hahne, 2011; Osterhout et al., 2006; Rossi et al., 2006; Tanner, Osterhout, and Herschensohn, 2009; Tanner, Nicol, Herschensohn, & Osterhout, 2012). Thus, a possible explanation of the differences observed in the present and some previous studies with regard to the P600 component could lie in the differences in L2 speakers' language proficiency. All L2 participants in the present study had a very high proficiency level, so it is possible that with increasing proficiency, morphological processes reflected in the P600 come into play because L2 learners are able to go beyond strictly shallow lexically based parsing strategies.

5.4.2 Critical condition and L2 phonological ambiguity

The critical experimental condition had a three-way comparison between the congruent condition on the one hand and two incongruent conditions on the other, one of which was hypothesized to be perceptually difficult for L2 listeners and the other one – perceptually easy. The former included the verbal forms (verb infinitives and future-tense verbs in the 3rd person singular) that differed on the basis of a Russian-specific phonological feature of consonantal palatalization (as in *ответить* (answerINF, /ɑtvʲetʲitʲ/)—*ответит* (answerFUTURE, /ɑtvʲetʲit/)). The latter included a juxtaposition of the verbal forms that differed on the basis of a phonological contrast common to both Russian and English (as in *ответить* (answerINF, /ɑtvʲetʲitʲ/)—*ответил* ("answered", /ɑtvʲetʲil/)).

The behavioral results from the sentence goodness judgment task demonstrated that L1 listeners performed at ceiling across the three conditions. In contrast, L2 comprehenders' perceptual difficulty with the discrimination of hard/soft consonants created a phonolexical ambiguity. While they mistakenly accepted only 2% of ungrammatical sentences as "good" ones in the incongruent past (phonologically unambiguous) condition, their incorrect acceptance rate in the incongruent future (phonologically ambiguous) condition was about 73%. Their response latencies to the questions in the incongruent future condition did not differ from the congruent condition and were significantly longer than in the incongruent past condition. This suggests that while L2 comprehenders noticed and correctly identified morphological violations involving substitutions of verb infinitive forms with past-tense forms, they were not disturbed by the violations involving substitutions with future-tense forms.

ERP results are in line with the participants' behavior in the sentence goodness task. Because morphology and syntax are both combinatorial and rule-governed systems, we expected to see a modulation of P600 (which is usually used as an index of syntactic processes) in response to morphological violations. Indeed, when time-locked to target word onsets, ERP waveforms for the L1 group showed a late positivity for both incongruent conditions peaking at around 1000 ms, which had the topography and morphology characteristic of the P600 component. Although we entertained a possibility of observing a graded ERP response in the two incongruent conditions as a function of phonetic similarity (e.g., a larger P600 response to the incongruent past compared to the incongruent future condition because the /tʲ/ and the /l/ phonemes in the inflections of the infinitive and past-tense forms, respectively, share fewer phonological features than the /tʲ/ and the /t/ phonemes in the infinitive and future-tense forms), such predictions were not borne out. Statistically, mean voltage amplitudes for the incongruent past and incongruent future conditions did not differ between each other. When time-locked to target word offsets, ERP waveforms showed a positive deflection in the 100-600 ms latency window followed by a pronounced late negativity in the 600-1300 ms window for both incongruent conditions compared to the congruent condition in the L1 group. No early negativities were present in the waveforms time-locked to either word onsets or word offsets.

With regard to the L2 listeners, we predicted that, if they use shallow processing instead of morphological parsing strategies (in accordance with the shallow-structure hypothesis), and do not decompose morphologically inflected words, no differences between the congruent and incongruent conditions will be observed in terms of P600 response. Alternatively, provided they store morphologically complex words undecomposed, some other ERP component responsible for lexical-semantic processing (e.g., N400) could reflect morphological violations. If, however, L2 comprehenders are sensitive to morphological cues during sentence comprehension, an ERP response similar to that in L1 listeners should be observed, although it is expected to be modulated by the level of perceptual difficulty of the phonological contrast involved in the distinction of the two morphological forms.

For the most part, our predictions were borne out. When time-locked to target word onsets, ERP waveforms for the L2 group showed a late positivity for the incongruent past condition (phonologically unambiguous) compared to the congruent condition, similarly to the L1 group. Scalp distribution of the ERP response was also similar to that of L1 listeners and was suggestive of the late P600 component. In accordance with our predictions, no noticeable P600 effect was observed for the incongruent future (phonologically ambiguous) condition. Mean voltage amplitudes for the congruent condition, on the one hand, and the incongruent past condition, on the other, did not differ significantly between L1 and L2 listeners, but the differences in mean amplitudes for the incongruent future condition were significant. L2 participants also showed a pattern of ERP responses similar to L1 participants when ERP waveforms were time-locked to target word offsets. In the incongruent past condition, a positive deflection in the 100-600 ms latency window followed by a late negativity in the 600-1300 ms window was observed relative to the congruent condition. The incongruent future condition did not elicit a differential ERP response relative to the congruent condition. Similarly to the L1 group, no early negativities were evoked in either the waveforms time-locked to word onsets or word offsets.

The observed differences between the incongruent conditions in the L2 group could not be due to the differences in the mastery of the two grammatical forms (past tense versus future tense). Verb conjugations in the past and future tense are usually covered in the first semester of Russian. Because all our L2 participants had a very high proficiency level, it is highly unlikely that they were familiar with the past-tense form and did not know the future-tense form.

Given evidence (Lukyanchenko and Gor, 2011; Chrabaszcz and Gor, in press) that the distinction between Russian hard and soft consonants is problematic for nonnative comprehenders of Russian, we argue that the observed differences in L2 participants' ERP traces are due to the phonolexical ambiguity created by the difficult phonological contrast. When L2 listeners have the necessary phonological representations in place and can differentiate between the target phonological contrasts easily (as in the incongruent past-tense condition), they extract the necessary phonological information as it becomes available through bottom-up processing and combine it with the constructed morphological predictions coming from top-down processing. Whenever a mismatch between the extracted phonological information and the activated, expected morpho-phonological template occurs, a break-down in comprehension happens, and the parser makes an attempt at the reanalysis and rechecking of the generated morphological predictions. In contrast, when phonological representations are fuzzy and unclear (as is the case with the consonant hard/soft distinction), L2 comprehenders cannot fully rely on the information extracted at the phonological level of processing because even after it has been extracted and processed, the output can still contain several eligible candidates (e.g., if the distinction between /t/ and /t^j/ is not accurately perceived, the output may contain both *ответит* and *ответить*). Thus, in case of ambiguous bottom-up information, L2 listeners will exhibit a morphological context bias effect and will pick the interpretation that is most compatible with the morphological predictions at no cost for the parser; hence, no P600 response which is normally associated with reanalysis and rechecking will be observed.

5.4.3 On the nature and timing of the P600

According to the neurocognitive model of auditory sentence comprehension (Friederici, 1995; 1999; 2002), online language comprehension takes place in a hierarchical manner. During the first phase (which roughly corresponds to the time window of ELAN component), word category-based phrase structure is built. This is followed by morphosyntactic and lexical-semantic processing as well as thematic role assignment in the second phase (N400 and LAN effects are observed at this stage). Finally, reanalysis, repair, and integration processes occur during phase 3 (which corresponds to the time window of the P600 component). In the context of this model, our results suggest that, similarly to syntactic processing, morphological predictions come into play during the third stage of processing and a violation of morphological prediction elicits a late P600 response. Notably, the P600 effects evoked in response to morphological violations involving nominal inflections (case marking) and verbal inflections (tense agreement) were very similar in the L1 group in terms of amplitude and timing. The L2 group also showed a comparable P600 response for violations in nominal and verbal inflections, except for the phonologically ambiguous (incongruent future) condition (Figure 28). The mean amplitude for the past-tense form violation was slightly greater than that elicited by violations in case marking. However, because nominal and verbal conditions contained different sentences, they cannot be compared directly.

With regard to the critical condition, the results suggest that the phonological violations on morphemes that are important for lexical and structural integration during sentence comprehension initiate repair processes at later stages. The average target word duration in our critical condition was 780 ms, with the phonological violation occurring in the last phoneme of the target word. For word onset time-locked waveforms, late positivity started to emerge around 800 ms and lasted for about 500 ms, peaking around 1000 ms; for word offset time-locked waveforms, it emerged as early as 100 ms but also spread over a 500-ms latency window. This suggests that as soon as the phonological information became available, it started being integrated with the morphological expectations arising from the preceding morphosyntactic context, and whenever those were not met, a morphological reanalysis was invoked, hence, the observed P600 effect. We think that no ELAN response was elicited because the kinds of morphological violations used in the present study did not involve word category violations. Rather, they represented violations within the same morphological (in this case, verbal) inflectional paradigm. It is also worth noting that, when ERP responses were time-locked to word offsets, an observed positivity in the 100-600 ms window was followed by a prolonged late negativity. This could be due to the fact that the breakdown in morphological predictions created a problem for the lexical-semantic processing of the subsequent word.

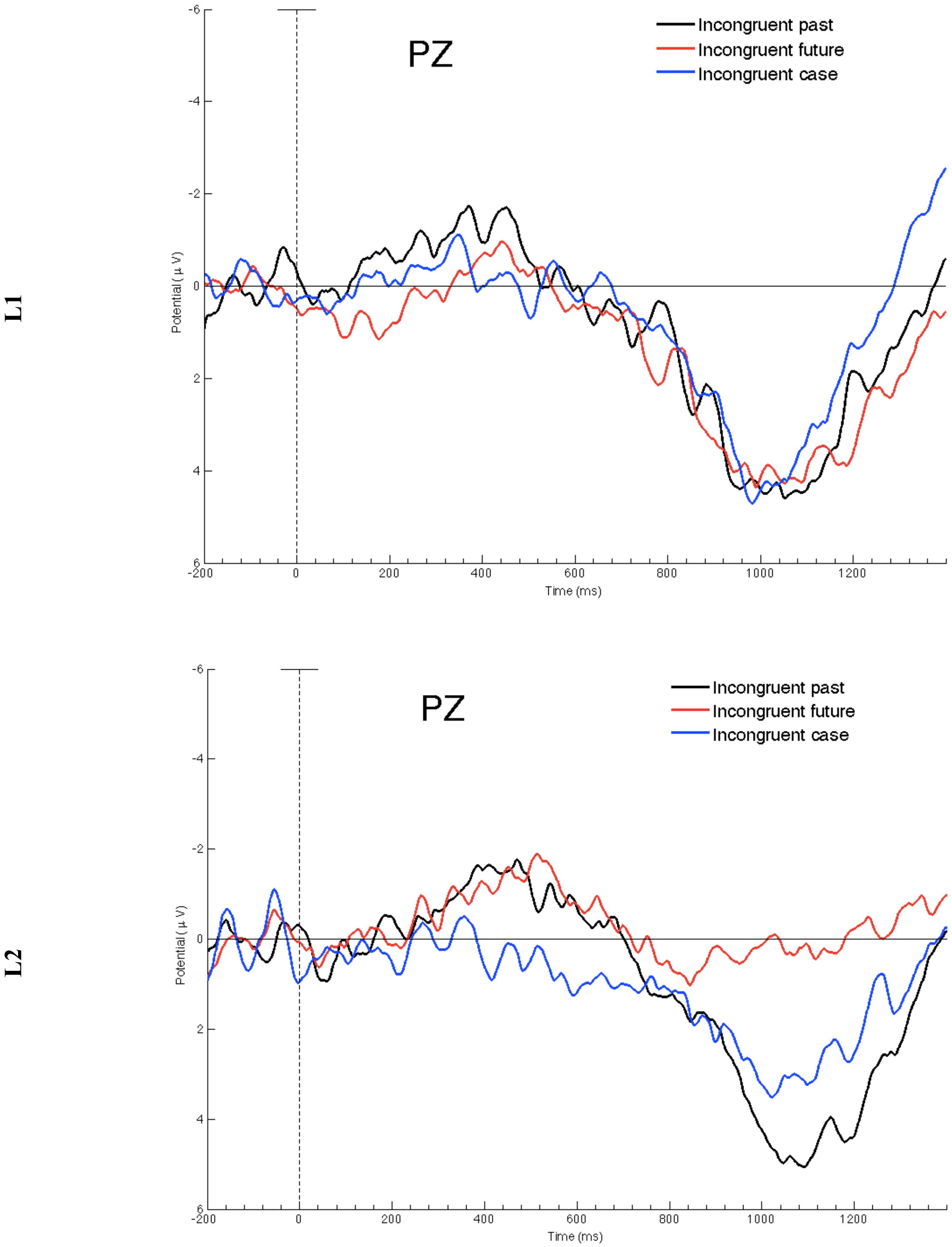


Figure 28. Grand average ERPs for a representative (PZ) electrode at the onset of the target word across all morphologically incongruent conditions in the L1 and L2 groups (incongruent past is in black, incongruent future is in red, and incongruent case is in blue). Time 0 is the onset of the stimuli. Negative polarity is plotted upwards. X axis represents time (milliseconds) and Y axis depicts voltage (microvolts, µV).

One more question that needs to be addressed is delayed timing of the P600 across different conditions and participant groups. On the one hand, this could be attributed to the auditory modality of stimulus presentation. Some auditory ERP studies on morphological processing have in fact demonstrated delayed ERP responses (Leinonen, Grönholm-Nyman, Järvenpää, Söderholm, Lappi, Laine, and Krause, 2009; Lück, Hahne, and Clahsen, 2006). In visually presented stimuli, words can be accessed instantaneously and as a whole, whereas in auditorily presented stimuli, words unfold over time such that the processor has to obey temporal and sequential dimensions of the stimuli (e.g., the stem of the word has to be processed before the inflection is encountered). Another explanation of the observed delay in the P600 component can be attributed to the nature of the stimuli used in the present study. Such delay can stem from the complexity and higher costs of the parsing process itself. Some researchers have suggested that the recognition of inflected words is more computationally complex than the recognition of monomorphemic words because of the additional procedures involved, such as verifying that the parse is exhaustive and that each morphological constituent is integrated into the prevailing linguistic context (Allen, Badecker, and Osterhout, 2003; Baayen, Dijkstra, and Schreuder, 1997; Lehtonen, Cunillera, Rodríguez-Fornells, Hultén, Tuomainen, and Laine, 2007). Thus, because Russian is an inflectionally rich language, it is possible that the processing of morphologically inflected forms requires additional processing costs.

6 Conclusion

6.1 Overview

The primary purpose of this dissertation study was to examine how phonological difficulties affect spoken word recognition in the L2, and whether L2 comprehenders can take advantage of the predictive power of contextual constraints (such as semantic, morphological, syntactic) to help them disambiguate phonolexically ambiguous words during auditory sentence comprehension. By comparing L2 listeners' performance with that of L1 listeners, we attempted to establish which of these contextual constraints are most effective in constraining word meaning in L2. In Chapters 4 and 5, we present evidence from a series of behavioral experiments and an event-related potential (ERP) experiment. Specific findings are discussed in the respective chapters. Here we summarize the main empirical findings to answer the research questions we proposed to address:

RQ 1. Does difficulty with discrimination of phonological contrasts lead to phonolexical ambiguity in the L2?

Drawing on the findings from our previous studies on L2 listeners' sensitivity to the phonological hardness/softness contrast in Russian consonants, the present study establishes across several experiments that the approximate and unstable nature of L2 phonological representations leads to phonolexical ambiguity in the L2, causing lexical confusion between the minimal pairs that differ on the basis of such phonological feature. As a result of such phonolexical ambiguity, these words become temporarily perceptually indistinguishable, potentially leading to joint activation and spurious lexical competition. These conclusions are supported by the results of the translation judgment task, where L2 listeners showed a strong tendency to provide the translation of a similar-sounding word instead of the target word, e.g., they translated the verb *брать* (/brɑtʲ/, "to take") as "brother" confusing it with the word брат (/brɑt/, "brother"). Due to the overall low-resolution phonological specifications in L2 lexical representations, it is quite possible that phonolexical ambiguity can potentially affect lexical items that are not necessarily minimal pairs but share a substantial amount of phonological overlap (see Cook, 2012).

RQ 2. What are the consequences of L2 phonolexical ambiguity for auditory sentence comprehension?

Behavioral and electrophysiological data from our three sentence-level experiments indicate that when L2 comprehenders encounter phonolexically unambiguous incongruent words during auditory speech comprehension, they experience processing difficulty trying to integrate them with the sentential context. Phonolexically ambiguous words, on the other hand, do not incur processing costs associated with contextual integration, as evidenced by i) negligible reaction time differences in the self-paced listening task, ii) a lack of inhibition effect in the lexical decision task, and iii) the absence of the P600 response in the ERP study. Such evidence suggests that L2 comprehenders treat these words as congruent with the context without meaning disruption and a breakdown of the comprehension flow. This implies that they have some other mechanisms in place that enable them to compensate for the incomplete perceptual information and to access the intended lexical candidates.

RQ 3. Do L2 listeners utilize contextual information for meaning resolution in online auditory sentence comprehension?

Phonological information plays a crucial role in spoken word recognition in that it acts as a sort of an activation code to the mental lexicon. The “bottom-up priority” principle (Marslen-Wilson and Tyler, 1980; Marslen-Wilson, 1987; 1989) postulates that phonological information contained in a word receives priority over contextual information. When the phonological form is decoded, other properties of the word (e.g., morphological, syntactic, orthographic, etc.) are also accessed. Therefore, incomplete phonological information can block lexical access. This is true for a situation when words are heard in isolation. In naturally occurring speech, however, words are embedded in sentences where they are combined with other words by means of complex semantic, syntactic, and morphosyntactic relationships. We show that L2 listeners can use knowledge about such relationships to anticipate the incoming input such that when their expectations are not met, a temporary breakdown in processing occurs, as evident by their performance in the control (perceptually unambiguous) conditions across the three sentence-level experiments. We speculate that the same kind of structural and semantic knowledge allows L2 comprehenders to process an unclear, phonolexically ambiguous word segment during sentence comprehension. Thus, provided that context has enough predictive power and that L2 listeners can take advantage of the contextual information, they should be able to access and select the intended lexical items through their semantic, syntactic and morphological characteristics despite low-resolution phonological information.

RQ 4. Do L2 listeners utilize different kinds of contextual information, such as semantic, morphological and syntactic, for meaning resolution to the same degree?

Based on the results of the lexical decision task and the self-paced listening task, L2 listeners utilize different kinds of contextual information to a different extent. L2 listeners, akin to L1 listeners, experience the strongest context effects in the syntactic and semantic conditions followed by the morphological condition, although L2 listeners can successfully generate morphological predictions and rely on them during online speech comprehension, as demonstrated by the ERP experiment. Such findings appear to be at odds with some existing SLA theories. For example, according to the shallow structure hypothesis, L2 speakers' representations lack syntactic specification and abstract configurationally determined elements (Clahsen and Felser, 2006a, 2006b; Felser et al., 2003; Papadopoulou and Clahsen, 2003), which forces them to rely on lexical-semantic and pragmatic knowledge and underuse morphosyntactic and inflectional information. Our findings cast doubt on the existing ideas about L2 listeners' use of contextual heuristics and highlight the importance of including the grammatical (morphological) level of analysis in the existing models of second language speech comprehension.

RQ 5. What is the time course of integration of phonological information with higher-order contextual information in L2?

The time course of when phonological information interacts with higher-order contextual information was examined on the example of morphologically constraining context using EEG method, which is known to have high temporal resolution. Phonetic deviation (the uniqueness point) in the target words used in the experiment corresponds to word offset, but because of the substantial initial phonological overlap in target contextually congruent and incongruent words and thanks to the predictive nature of the contextual constraints, the subjects are expected to have been able to identify target words before they reach the end of the word. In case of the phonological mismatch, the subjects realize that the actual target does not match their contextually facilitated expectations—this is the point when phonological information is in conflict with contextual information. Such conflict presumably happens after the subjects have already selected a potential lexical candidate, and should therefore reflect reanalysis, repair, or integration processes. In the ERP literature, these processes are predominantly considered to correspond to the time window of the P600 component (Friederici, 1995; Hagoort et al., 1993; Osterhout and Holcomb, 1992). The ERP data supports our speculations. When target words included a phonological mismatch with the morphological expectations, a P600 response was observed in both L1 and L2 listeners. For waveforms time-locked to word onsets, a P600 effect was delayed in time, possibly due to the modality (auditory, not visual) of stimuli presentation. The average duration of the target words in the critical condition was 780 ms, and the positivity started to emerge around 800 ms. It lasted for about 500 ms, peaking around 1000 ms. When the same waveforms were time-locked to word offsets, a positive deflection emerged as early as 100 ms and also spread over a 500-ms latency window. This suggests that as soon as the listeners reached words' uniqueness point and discovered a phonological mismatch, the parser experienced difficulty integrating the target word with the preceding context and invoked a rechecking procedure, after which meaning resolution was accomplished in about 500 ms.

RQ 6. How does auditory sentence processing compare in L1 and L2 in terms of the use of contextual information and the temporal aspects of context effects?

Based on the assumptions of the critical period hypothesis for language acquisition (Johnson and Newport, 1989), knowledge of the L2 acquired after puberty is represented rather differently from that of L1. Across several experiments, we have demonstrated that, indeed, L2 lexical representations may differ from those in L1 in that they may lack phonological specification and detail. However, we have also obtained behavioral and electrophysiological evidence showing that, despite subtle differences, the mechanisms associated with top-down processing and the use of contextual information for meaning resolution in auditory sentence comprehension are essentially the same in the L1 and the L2. Thus, our findings suggest that there are more similarities than differences between L1 and L2 auditory sentence processing.

6.2 Theoretical and practical implications

The present dissertation work provides the first comprehensive psycholinguistic analysis of how ambiguous phonological representations in the L2 affect L2 speakers' word recognition and auditory sentence comprehension, and how L2 speakers can potentially cope with such difficulties. The outcome of this work has widespread theoretical implications, including elucidating the mechanisms employed by L1 and L2 listeners during auditory speech comprehension to characterize the difficulties that L2 listeners face when processing phonologically ambiguous input. The findings challenge existing views regarding L2 speakers' ability to use contextual information in a predictive manner to resolve meaning and suggest the need to reconsider some of the common assumptions regarding L2 competence. Current models of spoken word recognition and speech comprehension, for example, should be revised to accommodate L2 data. Such attempts have already started to emerge (see, for example, the Second Language Lexical Access Model (SLLAM) in Cook (2012), which incorporates L2 specific factors, such as the underspecification of phonological representations and the proficiency-defined size of the mental lexicon). It is our hope that the results of the present study will also provide new insights into the role of phonology for speech comprehension and the ways it interacts with higher-order information coming from semantic, syntactic and morphosyntactic levels of analysis during real-time auditory processing.

This study has significant implications for pedagogical practices. Understanding speech is a critical component of communication. The importance of listening skills and the difficulty involved in listening to continuous speech have been acknowledged in all current L2 methodologies and textbooks, but listening comprehension has received relatively little attention in second language and classroom research. Our findings can inform educators about potentially difficult areas in L2 listening comprehension through identifying L1-L2 similarities and differences. More importantly, knowledge about how L2 learners can compensate for such difficulties (e.g., which contextual cues they routinely employ or underuse) can become a stepping-stone on their path to improve their linguistic competency.

6.3 Limitations and future research

We would like to acknowledge that although the present dissertation study provides some new insights into the problem of phonolexical ambiguity and context effects on meaning resolution in the L2, it is not devoid of limitations. First, there are methodological limitations associated with materials design due to a limited number of minimal pairs that exist in any given language. This imposes unavoidable restrictions on matching target words along certain parameters across different conditions (e.g., word class, word frequency, perceptual saliency).

Second, the results that we have described in this dissertation were obtained from very proficient L2 speakers. Because performance of L2 speakers is strongly mediated by their proficiency level, this raises the question of generalizability of the findings. Will L2 speakers with lower proficiency utilize contextual constraints with a similar success? Will their use of different contextual information also differ from highly proficient L2 speakers? It is possible that context effects change with changing proficiency, and that L2 speakers with lower proficiency favor semantic contextual cues over structural ones for meaning comprehension (in accordance with previous SLA literature). Examining performance of L2 speakers with different language proficiency will provide a more complete picture and a better understanding of how bottom-up and top-down mechanisms develop and change across various proficiency levels.

Finally, this dissertation work examines a selected set of contextual constraints and target features, and it remains to be seen whether the findings can be generalized to other kinds of contexts, phonological contrasts, and to L2 speakers with other L1-L2 combinations before any firm conclusions can be made.

We hope that the results of this study will not only inform current theories of speech perception and comprehension, but will also “open a window” into a new line of future research towards the study of phonology at the sentential level, which can potentially yield interesting findings and provide a more comprehensive picture of L2 auditory processing.

Appendices

APPENDIX A. Target words in the lexical decision task in context.

Block	Condition	Target	Translation	Phonemes	Lemma frequency	Surface frequency	Sound	Target type	PoS	Form
Critical	Semantic	мат	obscenity	3	11.27	3.75	t	congruent	N	Nom/Acc
		суд	court	3	301.02	90.97	t	congruent	N	Nom/Acc
		жест	gesture	4	41.68	13.77	t	congruent	N	Nom/Acc
		плод	fruit	4	54.58	11.86	t	congruent	N	Nom/Acc
		плед	plaid, throw	4	4.87	2.20	t	congruent	N	Nom/Acc
		угол	angle, corner	4	199.47	53.71	l	congruent	N	Nom/Acc
		мел	chalk	3	9.53	2.94	l	congruent	N	Nom/Acc
		пыл	heat, ardor	3	7.83	3.19	l	congruent	N	Nom/Acc
		мать	mother	3	227.84	214.90	t	confusable	N	Nom/Acc
		суть	essence, point	3	84.65	55.44	t	confusable	N	Nom/Acc
		жесть	tin	4	3.55	0.88	t	confusable	N	Nom/Acc
		плоть	flesh	4	28.56	12.38	t	confusable	N	Nom/Acc
		плеть	whip	4	9.72	1.47	t	confusable	N	Nom/Acc
		уголь	coal	4	13.35	10.39	l	confusable	N	Nom/Acc
		мель	shallow place	3	6.33	2.19	l	confusable	N	Nom/Acc
		пыль	dust	3	65.28	29.99	l	confusable	N	Nom/Acc
		газ	gas	3	77.08	22.18		unrelated	N	Nom/Acc
		нож	knife	3	62.96	23.55		unrelated	N	Nom/Acc
		обед	dinner	4	129.48	40.71		unrelated	N	Nom/Acc
		урок	lesson	4	71.86	19.59		unrelated	N	Nom/Acc
		цвет	color	4	180.70	44.90		unrelated	N	Nom/Acc
		гриб	mushroom	4	30.86	4.25		unrelated	N	Nom/Acc
		вес	weight	3	68.57	24.71		unrelated	N	Nom/Acc
		лед	ice	3	69.14	26.98		unrelated	N	Nom/Acc
		маф		3				nonce		
		сур		3				nonce		
		жерк		4				nonce		
		плор		4				nonce		

		плек		4				nonce		
		угак		4				nonce		
		меп		3				nonce		
		пык		3				nonce		
	Syntactic	брат	brother	4	317.49	126.90	t	congruent	N	Nom/Acc
		балет	ballet	5	16.06	5.68	t	congruent	N	Nom/Acc
		билет	ticket	5	60.89	23.52	t	congruent	N	Nom/Acc
		дед	grandfather	3	93.23	67.53	t	congruent	N	Nom/Acc
		ел	ate	3	196.33	21.34	l	congruent	V	Vpast
		дал	gave	3	1016.52	125.32	l	congruent	V	Vpast
		прибыл	arrived	6	94.39	28.62	l	congruent	V	Vpast
		стал	became	4	1580.42	449.48	l	congruent	V	Vpast
		брать	to take	4	228.80	45.61	t	confusable	V	Vinf
		болеть	to be sick	5	96.11	5.42	t	confusable	V	Vinf
		белеть	to whiten	5	11.78	0.33	t	confusable	V	Vinf
		деть	to put	3	10.08	1.84	t	confusable	V	Vinf
		ель	fir tree	3	30.16	2.87	l	confusable	N	Nom/Acc
		даль	distance	3	97.14	11.95	l	confusable	N	Nom/Acc
		прибыль	income	6	45.42	14.18	l	confusable	N	Nom/Acc
		сталь	steel	4	271.18	7.36	l	confusable	N	Nom/Acc
		вниз	downward	4	99.90	99.63		unrelated	Adv	
		вдали	faraway	5	29.63	29.63		unrelated	Adv	
		ничуть	not at all	5	16.62	16.60		unrelated	Adv	
		меж	between	3	27.21	27.12		unrelated	Pre	
		вне	outside of	3	64.61	64.26		unrelated	Pre	
		чуть	hardly	3	215.00	214.84		unrelated	Adv	
		вправо	to the right	6	17.14	16.78		unrelated	Adv	
		вновь	again	4	119.59	119.57		unrelated	Adv	
		брам		4				nonce		
		балер		5				nonce		
		билес		5				nonce		
		дес		3				nonce		

		ек		3				nonce		
		даф		3				nonce		
		прибыс		6				nonce		
		стах		4				nonce		
	Morphological	говорит	speaks	7	2115.36	476.28	t	congruent	V	V3sing
		готовит	cooks	7	58.99	8.60	t	congruent	V	V3sing
		ездит	goes (by a vehicle)	6	99.27	8.24	t	congruent	V	V3sing
		чистит	cleans	6	16.55	1.66	t	congruent	V	V3sing
		звонит	calls	6	66.63	10.31	t	congruent	V	V3sing
		помнит	remembers	6	322.68	20.35	t	congruent	V	V3sing
		ставит	puts	6	127.06	24.71	t	congruent	V	V3sing
		строит	builds	6	81.79	7.91	t	congruent	V	V3sing
		говорить	to speak	7	2115.36	272.10	t	confusable	V	Vinf
		готовить	to cook	7	58.99	14.45	t	confusable	V	Vinf
		ездить	to go (by a vehicle)	6	99.27	23.31	t	confusable	V	Vinf
		чистить	to clean	6	16.55	5.74	t	confusable	V	Vinf
		звонить	to call	6	66.63	12.22	t	confusable	V	Vinf
		помнить	to remember	6	322.68	23.73	t	confusable	V	Vinf
		ставить	to put	6	127.06	28.52	t	confusable	V	Vinf
		строить	to build	6	81.79	32.03	t	confusable	V	Vinf
		говорим	speak	7	2115.36	21.49	m	unrelated	V	V2pl
		готовим	cook	7	58.99	1.53	m	unrelated	V	V2pl
		ездим	go (by vehicle)	6	99.27	1.26	m	unrelated	V	V2pl
		чистим	clean	6	16.55	0.26	m	unrelated	V	V2pl
		звоним	call	6	66.63	0.41	m	unrelated	V	V2pl
		помним	remember	6	322.68	4.60	m	unrelated	V	V2pl
		ставим	put	6	127.06	3.08	m	unrelated	V	V2pl
		строим	build	6	81.79	2.48	m	unrelated	V	V2pl
		гаварик		7				nonce		
		гатовис		7				nonce		

		ездиф		6				nonce		
		чистип		6				nonce		
		звонир		6				nonce		
		помнис		6				nonce		
		ставир		6				nonce		
		строис		6				nonce		
Control	Semantic	храм	temple	4	87.91	28.69		congruent	N	Nom/Acc
		врач	doctor	4	139.63	39.90		congruent	N	Nom/Acc
		слон	elephant	4	21.39	6.09		congruent	N	Nom/Acc
		сон	dream	3	171.53	58.66		congruent	N	Nom/Acc
		пот	sweat	3	32.47	14.18		congruent	N	Nom/Acc
		стих	poem	4	162.57	11.48		congruent	N	Nom/Acc
		бок	side	3	85.96	27.34		congruent	N	Nom/Acc
		этаж	floor	4	76.13	21.13		congruent	N	Nom/Acc
		храп	snoring	4	4.45	2.74		confusable	N	Nom/Acc
		враг	enemy	4	148.75	26.77		confusable	N	Nom/Acc
		слог	syllable	4	13.91	4.20		confusable	N	Nom/Acc
		сок	juice	3	31.62	10.97		confusable	N	Nom/Acc
		пол	gender, floor	3	210.63	74.95		confusable	N	Nom/Acc
		стиль	style	4	63.95	22.29		confusable	N	Nom/Acc
		боль	pain	3	96.60	40.75		confusable	N	Nom/Acc
		этап	stage	4	61.29	16.45		confusable	N	Nom/Acc
		долг	debt	4	106.65	38.26		unrelated	N	Nom/Acc
		вкус	taste	4	79.53	29.43		unrelated	N	Nom/Acc
		ключ	key	4	70.17	30.70		unrelated	N	Nom/Acc
		шар	circle, balloon	3	50.13	17.45		unrelated	N	Nom/Acc
		конь	horse	3	97.99	17.36		unrelated	N	Nom/Acc
		пояс	belt, waist	4	41.18	17.02		unrelated	N	Nom/Acc
		луч	ray	3	75.49	16.13		unrelated	N	Nom/Acc
		плащ	raincoat	4	23.36	9.69		unrelated	N	Nom/Acc
		храк		4				nonce		
		врар		4				nonce		
		слоч		4				nonce		
		сош		3				nonce		
		пок		3				nonce		
		стип		4				nonce		
		боч		3				nonce		

		этар		4				nonce		
	Syntactic	успех	success	5	156.25	46.13		congruent	N	N
		брак	marriage, defect	4	59.51	24.41		congruent	N	N
		класс	classroom	4	177.62	40.96		congruent	N	N
		шок	shock	3	8.04	3.57		congruent	N	N
		жир	grease	3	21.62	4.76		congruent	N	N
		мышь	mouse	3	31.33	8.57		congruent	N	N
		пир	feast	3	17.96	8.07		congruent	N	N
		грех	sin	4	98.90	38.42		congruent	N	N
		успел	managed in time	5	203.30	88.52		confusable	V	Vpast
		брал	took	4	228.80	26.41		confusable	V	Vpast
		клал	put	4	45.48	4.55		confusable	V	Vpast
		шёл	walked	3	1048.95	133.69		confusable	V	Vpast
		жил	lived	3	816.81	105.99		confusable	V	Vpast
		мыл	washed	3	107.47	2.14		confusable	V	Vpast
		пил	drank	3	211.46	33.01		confusable	V	Vpast
		грел	heated	4	12.68	0.89		confusable	V	Vpast
		влево	to the left	5	17.14	16.83		unrelated	Adv	
		прочь	away	4	33.49	33.84		unrelated	Adv	
		явно	clearly	4	69.92	69.84		unrelated	Adv	
		еле	barely	3	29.70	29.65		unrelated	Adv	
		зря	in vain	3	40.24	40.02		unrelated	Adv	
		вон	there	3	97.26	97.01		unrelated	Adv	
		мимо	past	4	126.27	126.21		unrelated	Prep	
		вдоль	along	4	74.99	74.98		unrelated	Prep	
		успЕз						nonce		
		браф						nonce		
		клах						nonce		
		шомь						nonce		
		жих						nonce		
		мырь						nonce		
		пиф						nonce		

		грец						nonce		
	Morphological	любим	love	5	665.27	12.71		congruent	V	V2pl
		скажем	tell	6	2559.21	58.55		congruent	V	V2pl
		работаем	work	9	448.91	7.02		congruent	V	V2pl
		думаем	think	7	822.49	10.24		congruent	V	V2pl
		видим	see	5	1173.91	45.15		congruent	V	V2pl
		сидим	sit	5	575.83	10.91		congruent	V	V2pl
		читаем	read	7	337.83	7.94		congruent	V	V2pl
		сделаем	do	8	761.56	12.11		congruent	V	V2pl
		любишь	love	5	665.27	20.25		confusable	V	V2sing
		скажешь	tell	6	2559.21	23.49		confusable	V	V2sing
		работаешь	work	9	448.91	4.01		confusable	V	V2sing
		думаешь	think	7	822.49	46.16		confusable	V	V2sing
		видишь	see	5	1173.91	65.12		confusable	V	V2sing
		сидишь	sit	5	575.83	8.18		confusable	V	V2sing
		читаешь	read	7	337.83	5.57		confusable	V	V2sing
		сделаешь	do	8	761.56	8.25		confusable	V	V2sing
		любит	love	5	665.27	89.87		unrelated	V	V3sing
		скажет	tell	6	2559.21	52.78		unrelated	V	V3sing
		работает	work	9	448.91	66.58		unrelated	V	V3sing
		думает	think	7	822.49	56.50		unrelated	V	V3sing
		видит	see	5	1173.91	73.37		unrelated	V	V3sing
		сидит	sit	5	575.83	77.64		unrelated	V	V3sing
		читает	read	7	337.83	26.28		unrelated	V	V3sing
		сделает	do	8	761.56	24.03		unrelated	V	V3sing
		лЮбик						nonce		
		скАжер						nonce		
		рабОтаек						nonce		
		дУмаел						nonce		
		вИдир						nonce		
		сидИс						nonce		
		читАер						nonce		
		сдЕлаер						nonce		

APPENDIX B. Participants' mean error rate and reaction time in the lexical decision task.

A. Error rate

Condition	Language	Semantic				Morphological				Syntactic			
		Constraining		Neutral		Constraining		Neutral		Constraining		Neutral	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
						CRITICAL							
Congruent	L1	0.010	0.010	0.016	0.009	0.010	0.010	0.005	0.005	0.010	0.010	0.031	0.011
	L2	0.074	0.027	0.092	0.015	0.000	0	0.004	0.004	0.059	0.018	0.044	0.013
Confusable	L1	0.115	0.045	0.016	0.009	0.094	0.029	0.005	0.005	0.354	0.040	0.031	0.011
	L2	0.118	0.034	0.092	0.015	0.015	0.010	0.004	0.004	0.125	0.024	0.044	0.013
Unrelated	L1	0.094	0.039	0.010	0.010	0.042	0.025	0	0	0.188	0.055	0.042	0.019
	L2	0.074	0.031	0.015	0.010	0.096	0.040	0.015	0.010	0.272	0.049	0.176	0.034
Nonce	L1	0.042	0.033	0.021	0.014	0.094	0.036	0.063	0.043	0.021	0.014	0.146	0.042
	L2	0.044	0.017	0.074	0.022	0.059	0.021	0.051	0.021	0.022	0.012	0.037	0.015
						CONTROL							
Congruent	L1	0	0	0.005	0.005	0	0	0	0	0	0	0.026	0.011
	L2	0.044	0.017	0.110	0.020	0	0	0.004	0.004	0.044	0.020	0.077	0.012
Confusable	L1	0.156	0.045	0.005	0.005	0.104	0.037	0	0.000	0.417	0.049	0.026	0.011
	L2	0.287	0.044	0.110	0.020	0.176	0.040	0.004	0.004	0.463	0.053	0.077	0.012
Unrelated	L1	0.135	0.050	0.010	0.010	0.083	0.036	0.031	0.017	0.188	0.057	0.010	0.010
	L2	0.147	0.034	0.081	0.025	0.088	0.038	0	0	0.199	0.043	0.074	0.020
Nonce	L1	0.063	0.043	0.083	0.036	0.094	0.045	0.063	0.043	0.063	0.034	0.115	0.050
	L2	0.051	0.018	0.044	0.020	0.088	0.030	0.066	0.022	0.096	0.024	0.125	0.026

B. Reaction time

Condition	Language	Semantic				Morphological				Syntactic			
		Constraining		Neutral		Constraining		Neutral		Constraining		Neutral	
		Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
						CRITICAL							
Congruent	L1	1165.6	63.1	1173.4	57.2	1260.7	53.6	1355.5	75.1	1224.3	62.4	1306.5	64.2
	L2	1289.0	57.4	1229.9	52.2	1241.0	43.3	1233.7	29.7	1223.9	53.7	1232.8	34.4
Confusable	L1	1397.9	79.9	1173.4	57.2	1401.7	79.7	1355.5	75.1	1625.6	86.6	1306.5	64.2
	L2	1207.6	54.0	1229.9	52.2	1305.3	39.5	1233.7	29.7	1318.2	50.9	1232.8	34.4
Unrelated	L1	1342.4	130.4	1120.2	65.1	1384.1	71.0	1408.7	89.2	1530.2	124.8	1364.2	103.2
	L2	1332.0	57.5	1196.0	38.5	1428.7	56.7	1306.0	37.3	1431.1	48.6	1344.0	40.6
Nonce	L1	1357.8	57.2	1440.7	78.3	1584.9	66.9	1720.4	107.7	1376.4	56.0	1531.0	78.0
	L2	1295.8	56.1	1315.7	61.5	1512.0	55.9	1576.4	78.9	1314.7	53.4	1415.9	61.2
						CONTROL							
Congruent	L1	1043.5	44.1	1253.5	72.4	1397.7	86.3	1341.3	55.8	1168.0	86.0	1201.8	58.2
	L2	1201.1	50.1	1226.4	38.0	1260.1	38.4	1371.0	35.1	1232.6	54.5	1246.5	37.0
Confusable	L1	1271.5	71.1	1253.5	72.4	1464.9	79.5	1341.3	55.8	1403.3	83.0	1201.8	58.2
	L2	1416.7	73.3	1226.4	38.0	1489.1	67.1	1371.0	35.1	1570.6	69.2	1246.5	37.0
Unrelated	L1	1217.3	50.4	1307.7	95.2	1493.5	107.3	1368.2	76.0	1419.6	92.0	1206.0	50.6
	L2	1334.6	60.0	1239.8	43.6	1470.5	46.3	1362.8	45.5	1491.2	63.0	1283.3	51.2
Nonce	L1	1345.5	66.3	1416.9	71.6	1662.6	57.3	1588.9	65.0	1422.0	85.7	1487.2	89.4
	L2	1298.7	63.1	1405.4	55.0	1535.3	55.2	1531.2	45.2	1460.2	89.8	1392.3	73.4

Note: SE = standard error

APPENDIX C. Stimulus items in the translation judgment task.

Critical:

i) Semantic

мать—мат, суть—суд, жесть—жест, плоть—плод, плеть—плед, уголь—угол, мель—мел, пыль—пыл

ii) Morphological

говорить—говорит, готовить—готовит, ездить—ездит, чистить—чистит, звонить—звонит, помнить—помнит, ставить—ставит, строить—строит

iii) Syntactic

брать—брат, болеть—балет, белеть—билет, деть—дед, ель—ел, даль—дал, прибыль—прибыл, сталь—стал

Control:

i) Semantic

храм—храп, врач—враг, слон—слог, сон—сок, пот—пол, стих—стиль, бок—боль, этаж—этап

ii) Morphological

любишь—любим, скажешь—скажем, работаешь—работаем, думаешь—думаем, видишь—видим, сидишь—сидим, читаешь—читаем, сделаешь—сделаем

iii) Syntactic

успех—успел, брак—брал, класс—клал, шок—шёл, жир—жил, мышь—мыл, пир—пил, грех—грел

Nonce words:

i) Hard-to-soft

трудь, сынь, плань, опыть, дворь, соседь, стакань, голодь, балконь, баль, стуль, городь

ii) Soft-to-hard

ден, двер, огон, модел, парен, любов, обув, бров, груд, гел, корен, тен

iii) Fillers

страб, ноб, гук, кнИр, встрУм, цер, стет, карс, гоч, минАг, собУр, жЕндон, лЕрта, плАра, мУма, тЕкра, фИльса, рустА, врамА, клАма, линтА, стулАнта, закИла, журАна

APPENDIX D. Experimental sentences in the self-paced listening task.

Semantic condition

1. Учительница пригласила на родительское собрание отца и мать/мат/газ моего лучшего друга.
2. Шахматист сделал ход конем и поставил шах и мат/мать/газ лучшему игроку в стране.
3. Профессор объяснил нам, в чем заключается суть/суд/лед новой социальной реформы.
4. Известный адвокат советует подать жалобу в суд/суть/лед по защите прав человека.
5. Для изготовления консервных банок используется жесть/жест/гриб самого высокого качества.
6. Хозяин дома открыл дверь, улыбнулся и сделал жест/жесть/гриб рукой, приглашающий войти.
7. Буддисты укрепляют слабое тело и немощную плоть/плод/обед с помощью занятий йогой.
8. Мы посадили дерево, на котором вырос плод/плоть/обед похожий на большой апельсин.
9. Лошадь не хотела идти поэтому он взял плеть/плед/урок в руки и больно ударил ее.
10. Когда ей холодно, она закутывается в плед/плеть/урок и пьет горячий чай.
11. Шахтеры на нашей шахте добывают уголь/угол/цвет двадцать четыре часа в сутки.
12. Машина на большой скорости врезалась в угол/уголь/цвет дома на Садовой улице.
13. Российский военный корабль сел на мель/мел/вес недалеко от берегов Норвегии.
14. Ученик подошел к доске, взял мел/мель/вес в руки и написал предложение.
15. Дедушка взял книгу с полки и сдул пыль/пыл/нож со старых пожелтевших страниц.
16. Скучная игра футболистов охладила пыл/пыль/нож болельщиков обеих команд.

Syntactic condition

17. Несмотря на уговоры, полицейский не хотел брать/брат/вниз деньги арестованных преступников.
18. В маленькой квартире на втором этаже живет брат/брать/вниз великого русского писателя.
19. В последнее время люди стали болеть/балет/вдали намного чаще, чем раньше.
20. На следующих выходных друзья идут смотреть балет/болеть/вдали известного французского хореографа.
21. Пока мы показывали фотографии, его лицо начало белеть/билет/ничуть от злости и нескрываемой зависти.
22. Старик с раздражением достал из кармана билет/белеть/ничуть на автобус и протянул его контролеру.

23. Дети выросли, и родители не знают, куда деть/дед/меж старые ненужные игрушки.
24. Мы вошли в комнату, в которой спал дед/деть/меж на старом потертом диване.
25. В парке за зданием школы растет ель/ел/вне высотой пятьдесят метров.
26. Когда гости зашли в дом, хозяин ел/ель/вне пироги и запивал их молоком.
27. Капитан сидел на берегу и смотрел в даль/дал/чуть темного ночного океана.
28. Ребенок уже сломал телефон, который ты дал/даль/чуть ему вчера поиграть.
29. Банкир рассказал ребятам, как получить прибыль/прибыл/вправо от бизнеса с наименьшим риском.
30. Взволнованный солдат объявил, что генерал прибыл/прибыль/вправо на вокзал рано утром.
31. Стоимость машин будет зависеть от цены на сталь/стал/вновь в ближайшие пять лет.
32. После окончания медицинского института мой сын стал/сталь/вновь известным хирургом в городе.

Morphological condition

33. Несмотря на расспросы, мужчина не хочет говорить/говорит/говорим правду о том, что произошло.
34. Один профессор в нашем университете говорит/говорить/говорим на десяти иностранных языках.
35. Девушка моего лучшего друга любит готовить/готовит/готовим новые блюда и экспериментировать.
36. По телевизору показывают, как знаменитый актер готовит/готовить/готовим свое любимое блюдо.
37. Жена моего соседа любит ездить/ездит/ездим за покупками со своими подругами.
38. Каждое лето наша младшая дочка ездит/ездить/ездим в деревню к бабушке с дедушкой.
39. Семилетний сын моей подруги не любит чистить/чистит/чистим зубы перед сном.
40. Пока дедушка читает газету, бабушка чистит/чистить/чистим картошку для супа к обеду.
41. Моя близкая подруга из Америки любит звонить/звонит/звоним по вечерам и долго болтать.
42. Когда у брата заканчиваются деньги, он обычно звонит/звонить/звоним родителям и просит о помощи.
43. Наша страна всегда будет помнить/помнит/помним этот исторический день.
44. Известный музыкант очень ясно помнит/помнить/помним свой первый концерт.
45. Молодой преподаватель по физике не хочет ставить/ставит/ставим плохие оценки ученикам.
46. Папа приходит с работы и аккуратно ставит/ставить/ставим ботинки в дальний угол.
47. Маленький ребенок наших друзей любит строить/строит/строим замки из песка на берегу.

48. Самый богатый человек в мире строит/строить/строим дворец на берегу океана.

APPENDIX E. Target items in the ERP experiment.

A. Critical condition

#	Congruent	Incongruent future	Incongruent past
1	бросить	бросит	бросил
2	взвесить	взвесит	взвесил
3	вспомнить	вспомнит	вспомнил
4	встретить	встретит	встретил
5	вылечить	вылечит	вылечил
6	выполнить	выполнит	выполнил
7	выпустить	выпустит	выпустил
8	выразить	выразит	выразил
9	выступить	выступит	выступил
10	выучить	выучит	выучил
11	выяснить	выяснит	выяснил
12	добавить	добавит	добавил
13	доверить	доверит	доверил
14	доставить	доставит	доставил
15	закончить	закончит	закончил
16	заполнить	заполнит	заполнил
17	запомнить	запомнит	запомнил
18	запретить	запретит	запретил
19	заставить	заставит	заставил
20	защитить	защитит	защитил
21	исполнить	исполнит	исполнил
22	исправить	исправит	исправил
23	назначить	назначит	назначил
24	напомнить	напомнит	напомнил
25	обвинить	обвинит	обвинил
26	обновить	обновит	обновил
27	объединить	объединит	объединил
28	объяснить	объяснит	объяснил
29	окончить	окончит	окончил
30	оставить	оставит	оставил
31	осуществить	осуществит	осуществил
32	ответить	ответит	ответил
33	отличить	отличит	отличил
34	отметить	отметит	отметил
35	отправить	отправит	отправил
36	очистить	очистит	очистил
37	победить	победит	победил
38	поверить	поверит	поверил
39	повесить	повесит	повесил
40	повторить	повторит	повторил
41	повысить	повысит	повысил
42	поговорить	поговорит	поговорил

43	позвонить	позвонит	позвонил
44	поздравить	поздравит	поздравил
45	познакомить	познакомит	познакомил
46	покрасить	покрасит	покрасил
47	посетить	посетит	посетил
48	поставить	поставит	поставил
49	построить	построит	построил
50	потратить	потратит	потратил
51	почистить	почистит	почистил
52	представить	представит	представил
53	предупредить	предупредит	предупредил
54	пригласить	пригласит	пригласил
55	применить	применит	применил
56	примерить	примерит	примерил
57	присоединить	присоединит	присоединил
58	причинить	причинит	причинил
59	проверить	проверит	проверил
60	простить	простит	простил
61	различить	различит	различил
62	разрешить	разрешит	разрешил
63	расслабить	расслабит	расслабил
64	решить	решит	решил
65	родить	родит	родил
66	сблизить	сблизит	сблизил
67	смягчить	смягчит	смягчил
68	снизить	снизит	снизил
69	соединить	соединит	соединил
70	сократить	сократит	сократил
71	сообщить	сообщит	сообщил
72	составить	составит	составил
73	сохранить	сохранит	сохранил
74	сочинить	сочинит	сочинил
75	убедить	убедит	убедил
76	увеличить	увеличит	увеличил
77	уволить	уволит	уволил
78	угостить	угостит	угостил
79	удалить	удалит	удалил
80	ударить	ударит	ударил
81	уделить	уделит	уделил
82	удивить	удивит	удивил
83	украсить	украсит	украсил
84	уменьшить	уменьшит	уменьшил
85	уничтожить	уничтожит	уничтожил
86	усложнить	усложнит	усложнил
87	успокоить	успокоит	успокоил
88	усыновить	усыновит	усыновил

89	уточнить	уточнит	уточнил
90	ухудшить	ухудшит	ухудшил

B. Control condition

#	Congruent	Incongruent semantic	Incongruent case
1	адрес	метр	адресу
2	альбом	холод	альбому
3	багаж	омлет	багажу
4	балет	сапог	балету
5	банан	народ	банану
6	банк	песок	банку
7	бассейн	плакат	бассейну
8	берег	ковёр	берегу
9	билет	двор	билету
10	браслет	маршрут	браслету
11	воздух	момент	воздуху
12	вокзал	предел	вокзалу
13	вопрос	лагерь	вопросу
14	гараж	поиск	гаражу
15	голос	мороз	голосу
16	город	след	городу
17	десерт	каблук	десерту
18	диван	рынок	дивану
19	диск	центр	диску
20	договор	потолок	договору
21	дождь	союз	дождю
22	доклад	огурец	докладу
23	дом	сыр	дОму
24	живот	способ	животу
25	журнал	чемодан	журналу
26	завод	кулак	заводу
27	завтрак	автобус	завтраку
28	закон	ветер	закону
29	запах	совет	запаху
30	знак	день	знаку
31	зонт	клуб	зонту
32	ключ	срок	ключу
33	конверт	помидор	конверту
34	концерт	секрет	концерту
35	корабль	вариант	кораблю
36	костюм	хвост	костюму
37	крест	отдел	кресту
38	кризис	рюкзак	кризису
39	курс	кран	курсу
40	куст	крик	кусту
41	магазин	предмет	магазину

42	мозг	угол	мозгу
43	музей	бензин	музею
44	номер	вагон	номеру
45	океан	период	океану
46	опыт	смех	опыту
47	остров	приказ	острову
48	ответ	взгляд	ответу
49	отпуск	размер	отпуску
50	пакет	гвоздь	пакету
51	паспорт	самолёт	паспорту
52	пляж	цвет	пляжу
53	поезд	труд	поезду
54	пожар	запад	пожару
55	полёт	вечер	полёту
56	портрет	витамин	портрету
57	поход	образ	походу
58	пример	бокал	примеру
59	процесс	восток	процессу
60	рассказ	район	рассказу
61	ресторан	пистолет	ресторану
62	рецепт	карман	рецепту
63	роман	повод	роману
64	салат	закат	салату
65	свет	танец	свету
66	свитер	состав	свитеру
67	сериал	чайник	сериалу
68	словарь	пейзаж	словарю
69	снег	плач	снегу
70	спорт	порог	спорту
71	стакан	отдых	стакану
72	стол	лист	столу
73	стул	овощ	стулу
74	театр	фрукт	театру
75	текст	отряд	тексту
76	теннис	рукав	теннису
77	торт	круг	торту
78	туалет	космос	туалету
79	ужин	вход	ужину
80	урок	мост	уроку
81	успех	туман	успеху
82	учебник	интерес	учебнику
83	факт	рост	факту
84	фильм	удар	фильму
85	хлеб	волос	хлебу
86	цирк	рубль	цирку
87	шарф	глаз	шарфу

88	экзамен	процент	экзамену
89	этаж	орган	этажу
90	язык	конец	языку

Bibliography

Allen, M., Badecker, W., Osterhout, L. (2003). Morphological analysis in sentence processing. An ERP study. *Language and Cognitive Processes, 18*, 405–430.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–39.

Altmann, G. T. (1997). *The ascent of Babel: An exploration of language, mind, and understanding.* Oxford: Oxford University Press.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52,* 163–187.

Ardal, S., Donald, M. W., Neuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language, 39,* 187-205.

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn and R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 147-254). New York: Wiley.

Baayen, H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language, 37*, 94–117.

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon, and T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 155-205). Orlando, FL: Academic Press.

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed-effects models using S4 classes.* R package version 0.999375-37, http://CRAN.R-project.org/package=lme4

Berne, J. E. (2004). Listening comprehension strategies: A review of the literature. *Foreign Language Annals, 37*(4), 521-533.

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 in cross-language research* (pp.171-203). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn and M. J. Munro (Eds*.), Language experience and second language speech learning: In honor of James Emil Flege* (pp. 12-34). Amsterdam/Philadelphia: John Benjamins.

Block, E. L. (1992). See how they read: Comprehension monitoring of L1 and L2 readers. *TESOL, 26* (2), 319-343.

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer. computer program, version 4.5.16, www.praat.org

Bondarko, L. (2005). Phonetic and phonological aspects of the opposition of “soft” and “hard” consonants in the modern Russian language. *Speech Communication*, *47*, 7–14.

Bosch, L., Costa, A., & Sebastián- Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology, 12*, 189–222.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112,* 272–284.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spatial Vision, 10*, 443-446.

Broersma, M. (2002). Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors. In J. H. L. Hansen and B. Pellom (Eds.*), Proceedings of the 7th International Conference on Spoken Language Processing* (pp. 261– 264). Boulder, CO: University of Colorado, Center for Spoken Language Research.

Broersma, M. (2005). *Phonetic and lexical processing in a second language* (MPI Series in Psycholinguistics No.34). Doctoral dissertation, Radboud University, Nijmegen, The Netherlands.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*, 22-34.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology, 64*(1), 74-95.

Brown, C. M., Hagoort, P., & Ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience, 11*(3), 261–281.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349*, 788-790.

Carrasco, C. & Frenck-Mestre, C. (2009). Phonology helps in processing grammatical gender: ERP evidence from L1 and L2 French. *22cd Annual CUNY Conference on Human Sentence Processing,* Davis, USA, 26-28 March.

Cervera, T., & González-Alvarez, J. (2011). Test of Spanish sentences to measure speech intelligibility in noise conditions. *Behavioral Research, 43*, 459–467.

Chow, W. Y. (2013). *The temporal dimension of linguistic prediction.* Doctoral Dissertation, University of Maryland.

Chrabaszcz, A., and Gor, K. (In press). Context effects in the processing of phonolexical ambiguity in L2. *Language Learning*.

Clahsen, H., & Felser, C. (2006a) Continuity and shallow structures in language processing: a reply to our commentators. *Applied Psycholinguistics, 27*, 107-126.

Clahsen, H., & Felser, C. (2006b) Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3-42.

Clahsen, H., Balkhair, L., Schutter, J. S., & Cunnings, I. (2013). The time course of morphological processing in a second language. *Second Language Research, 29*(1), 7-31.

Clahsen, H., Sonnenstuhl, I., Blevins, J. P. (2003). Derivational morphology in the German mental lexicon: a dual mechanism account. In: Baayen, H., Schreuder, R. (Eds.), *Morphological Structure in Language Processing*. Mouton de Gruyter, Berlin, pp. 125–155.

Connine, C. M. (1987) Constraints on interactive processes in auditory word recognition: the role of sentence context. *Journal of Memory and Language, 26*, 527–38.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193–210.

Connine, C. M., Blasko, D., & Hall, M. (1991) Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234–50.

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,* 256–266.

Cook, S. (2012). *Phonological form in L2 lexical access: friend or foe?* Doctoral Dissertation, University of Maryland.

Cook, S., & Gor, K. (2012). How fine-grained are phonological representations of lexical entries in L2 mental lexicon? 31st Annual Second Language Research Forum, Pittsburg, PA.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*, 21–58.

Craig, C, Kim, B., Rhyner, P., & Chirillo, T. (1993). Effects of word predictability, child development and aging on time-gated speech recognition performance. *Journal of Speech and Hearing Research, 36*, 832-841.

Cutler, A. (1989). Auditory lexical access: Where do we start? In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 342-357), Cambridge, MA: MIT Press.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: effects of local context. *Journal of the Acoustical Society of America, 124,* 1264–1268.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second language listening. *Journal of Phonetics, 34*, 269–284.

Dahan, D.,Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes, 16,* 507–34.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8*, 919-924.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12*(4), 635-647.

Dell, G. S., Oppenheim, G. M., & Kittredge, A. (2008). Saying the right word at the right time: syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes, 23*, 583–608.

DeLong, K. A. (2009). *Electrophysiological explorations of linguistic pre- activation and its consequences during online sentence processing*. Doctoral Dissertation, UC San Diego.

Delong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117-1121.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods, 134*, 9-21.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences, 22*(6), 680-689.

Diehm, E. E. (1998). *Gestures and linguistic function in learning Russian: Production and perception studies of Russian palatalized consonants*. Ann Arbor.

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing 'deafness' in French? *Journal of Memory and Language, 36,* 406-421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America, 110*, 1606-1618.

Dupoux, E., Sebastian-Galles, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition, 106*(2), 682-706.

Elman, J. L. (1989). Connectionist approaches to acoustic/phonetic processing. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 227-261), Cambridge, MA: MIT Press.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology, 51*, 87–107.

Faroqi-Shah, Y. & Waked, A. (2010). Grammatical category dissociation in multilingual aphasia. *Cognitive Neuropsychology, 27*, 181 -203

Faustmann, A., Murdoch, B. E., Finnigan, S. P., & Copland, D. A. (2005). Event-related brain potentials elicited by semantic and syntactic anomalies during auditory sentence processing. *Journal of the American Academy of Audiology, 16*(9), 708-725.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K.D., Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *Journal of Memory and Language, 41,* 469-495.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children’s processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11*, 127–163.

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*(03), 453-489.

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(2), 324-335.

Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top-down? *System, 32*, 363–377.

FitzPatrick, I. & Indefrey, P. (2007). Effects of sentence context in L2 natural speech comprehension. *Nijmegen CNS, 1*(2), 43-56.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America, 93*, 1589-1608.

Folk, J., & Morris, R. (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye movement analysis. *Memory and Cognition*, *31*(1), 87-99.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116-124.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics, 36*, 359-368.

Frauenfelder, U. H., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes, 16*, 583–607.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 26*, 505-526.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica, 128*(3), 528-536.

Frenck-Mestre, C.; Carrasco, H., McLaughlin, J., Osterhout, L.; Foucart, A. (2010). Linguistic input factors in native and L2 processing of inflectional morphology: Evidence from ERPs. *Language, Interaction and Acquisition, 1*(2), 206-228.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78–84.

Friederici, A. D. (1999). 'The neurobiology of language comprehension. In A.D. Friederici (Ed.), *Language comprehension: A biological perspective* (pp. 263–301). Berlin/Heidelberg/New York: Springer.

Friederici, A. D., Frisch, S. (2000). Verb argument structure processing: the role of verb-specific and argument specific information. *Journal of Memory and Language, 43*, 476–507.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1,* 183–192.

Friederici, A.D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain and Language, 50*, 259–281.

Gaskell M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325–349.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes, 12*, 613–56.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 380-396.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge: University Press.

Gibson, E. (2006). The interaction of top–down and bottom–up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language, 54*(3), 363-388.

Goldinger, S., Luce, P., Pisoni, D., & Marcario, J. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1211–1238.

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: the contribution of semantics. *Bilingualism: Language and Cognition, 12*, 385–392.

Gor, K., & Jackson, S. (2013). Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes, 28*(7), 1065-1091.

Gor, K., & Lukyanchenko, A. (2012). The effect of context predictability on understanding speech in noise by native and nonnative listeners. 31st Annual Second Language Research Forum, Pittsburgh, PA.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds ‘l’ and ‘r’, *Neuropsychologi*a*, 9*, 317–323.

Gow, D.W. (2001) Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133–59.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28,* 267–83.

Grossberg, S., & Myers, C.W. (2000). The resonant dynamics of speech perception: interword integration and duration-dependent backward effects. *Psychological Review, 107*, 735–67.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12,* 556–568.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34,* 660–676.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1055-69.

Hagoort, P., Brown, C.M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*(3), 2001.

Hahne, A., & Friederici, A. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4,* 123-141.

Halle, M. (2002). *From memory to speech and back*. Berlin: Mouton de Gruyter.

Hamburger, M. B., & Slowiaczek, L. M. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review, 3*, 520-525.

Hasan, A. S. (2000). Learner's perception of listening comprehension problems. *Language, Culture and Curriculum, 13*(2), 137–153.

Hawkins, J. (2004). *On Intelligence*. New York: Henry Holt.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research,19*, 59–73.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes, 5*(4), 281-312.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology, 8*(2-3), 203-241.

Howie, S. (2001). Formant transitions of Russian palatalized and nonpalatalized syllables. *IULC Working Papers, 1,* 1-22.

Hu, G. (2009). *Cognitive mechanisms underlying second language listening comprehension*. Doctoral dissertation. Georgia State University, Atlanta, GA.

Hu, G., & Jiang, N. (2011). Semantic integration in listening comprehension in a second language: Evidence from cross-modal priming. In P. Trofimovich & K. McDonough (Eds.), *Applying Priming Research to L2 Learning and Teaching* (pp. 199-218). Philadelphia/ Amsterdam: John Benjamins.

Isenberg, D., Walker, E., & Ryder, J. (1980). A top-down effect on the identification of function words. *Journal of the Acoustical Society of America, 68,* S48-S48.

Jakobson, R., Fant, G., & Halle, M. (1951). *Preliminaries to speech analysis. The distinctive features and their correlates*.

Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese–speaking children. *Journal of Speech, Language, and Hearing Research, 46*, 1297–1311.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57*, 1-33.

Jiang, N., Novokshanova, E., Masuda, K., & Wang, Xin. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning, 61*, 940-967.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60–99.

Jusczyk, P. W. (1986). Towards a model for the development of speech perception. In J. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 1-19), Hillsdale, NJ: Erlbaum.

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass, 1*(6), 571–591.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of integration difficulty. *Language and Cognitive Processes, 15,* 159–201.

Kalikow, D. N., Stevens, K. N., & Elliot, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 5*, 1337–1360.

Kazanina, N., Phillips, C., & Idsardi, W. (2006).The Influence of Meaning on the Perception of Speech Sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103* (30), 11381-11386.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279-312.

Kochetov, A. (2002). *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization.* New York, London: Routledge.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Language and Memory, 33*, 149–174.

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50,* 93–107.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*, 831-843.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships with simple sentences. *Cognitive Brain Research, 17*, 117–129.

Kutas, M. and Dale, A. (1997). Electrical and magnetic readings of mental functions. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 197–242), Psychology Press.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*(5), 539–550.

Kutas, M., & Kluender, R. (1994). What is who violating? Are consideration of linguistic violations in light of event-related brain potentials. In H. J. Heinze, T. F. Münte, & G. R. Mangun (Eds.), *Cognitive electrophysiology: Basic and clinical applications* (pp. 183–210). Boston: Birkhäuser.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*, 218–233.

Lahiri, A., & Reetz, H. (2002). Underspecified Recognition. In C. Gussenhoven and N. Warner (Eds.), *Laboratory Phonology VII* (pp. 637-75). Berlin: Mouton.

Lardiere, D. (1998). Case and tense in the 'fossilized' steady state. *Second Language Research, 14*, 1-26.

Laszlo, S., & Federmeier, K. (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*, 326-338.

Lau, E. F. (2009). *The predictive nature of language comprehension.* Doctoral Dissertation, University of Maryland.

Lee-Ellis, S., Idsardi, W. J., & Phillips, C. (2009). Distinguishing effects of early exposure and language dominance: Speech perception by heritage speakers. 36th Boston University Conference on Language Development, Boston, MA.

Lehtonen, M., Cunillera, T., Rodríguez-Fornells, A., Hultén, A., Tuomainen, J., & Laine, M. (2007). Recognition of morphologically complex words in Finnish: Evidence from event-related potentials. *Brain research, 1148*, 123-137.

Leinonen, A., Grönholm-Nyman, P., Järvenpää, M., Söderholm, C., Lappi, O., Laine, M., & Krause, C. M. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: evidence from a morphologically rich language. *Brain research, 1275*, 54-66.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). MIT press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral Brain Science. 22*, 1–75.

Liederman, J., Gilbert, K., McGraw Fisher, J., Mathews, G., Frye, R., & Joshi, P. (2010). Are women more influenced than men by top-down semantic information when listening to disrupted speech? *Language and Speech, 54*(1), 33–48.

Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*(3), 651-662.

Liu, Y., Shu, H., & Wei, J. (2006). Spoken word recognition in context: Evidence from Chinese ERP analyses. *Brain and Language, 96,* 37–48.

Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in new perceptual categories. *Journal of the Acoustical Society of America, 94* (3), 1242–1255.

Lively, S., Pisoni, D., Yamada, R., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*(4), 2076–2087.

Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition, 27,* 385–398.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39*, 155–8.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics, 62*, 615–25.

Lück, M., Hahne, A., & Clahsen, H. (2006). Brain potentials to morphologically complex words during listening. *Brain research, 1077*(1), 144-152.

Lück, M., Hahne, A., & Clahsen, H. (2006). Brain potentials to morphologically complex words during listening. *Brain research, 1077*(1), 144-152.

Luck, S. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Lukyanchenko, A., & Gor, K. (2011). Perceptual correlates of phonological representations in heritage speakers and L2 learners. *Proceedings of the 35th Annual Boston University Conference on Language Development.* Somerville, MA: Cascadilla Press.

Lukyanchenko, A., & Gor, K. (2012). Meaning resolution in L2 speech comprehension: Top-down or bottom-up. 31st Annual Second Language Research Forum, Pittsburgh, PA.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory & Language, 32,* 692-715.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language & Cognitive Processes, 9,* 157-201.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 3-25), Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1-71.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989) Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576–85.

Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376–92.

Marslen-Wilson, W. D., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes, 10,* 285-308.

MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

MATLAB version 7.10.0 (2010). Natick, Massachusetts: The MathWorks Inc., 2010.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*(3), 686–693.

McClelland J. L., & Elman J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McClelland, J. L., Thomas, A., McCandliss, B. D., & Fiez, J. A. (1999). Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. In J. Reggia, E. Ruppin, & D. Glanzman (Eds.), *Progress in brain research: Vol. 121. Disorders of brain, behavior, and cognition: The neurocomputational perspective* (pp. 75–80). Oxford, England: Elsevier.

McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37-53). Oxford: Oxford University Press.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363–89.

Mecartty, F. H. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning, 11*(2), 323–348.

Miller, J. L., Green, K., & Schermer, T. (1984). A distinction between effects of sentential speaking rate and semantic congruity on word identification. *Perception and Psychophysics, 36,* 327-336.

Mills, D., Prat, C., Zangl, R., Stager, C., Neville, H. & Werker, J. (2004). Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. *Journal of Cognitive Neuroscience, 16,* 1452–1464.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *The Quarterly Journal of Experimental Psychology, 59*(8), 1395–1424.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1981). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception and Psychophysics, 18*, 331–340.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research, 22*, 205– 220.

Moreno, E. M., Rodríguez-Fornell, A., & Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *Journal of Neurolinguistics, 21*, 477–508.

Morris, J., & Holcomb, P. J. (2005). Event-related potentials to violations of inflectional verb morphology in English. *Cognitive Brain Research, 25*(3), 963-981.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Moss, H. E., & Marslen-Wilson, W. D. (1993). Access to word meanings during spoken language comprehension: effects of sentential semantic context. *Journal of Experimental Psychology: Learning, Memory & Cognition, 19*, 1254–76.

Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research, 21*(2), 152–174.

Mueller, J. L. (2006). L2 in a nutshell: The investigation of second language processing in the miniature language model. *Language Learning, 56*, 235-270.

Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience, 17,* 1229-1244.

Münte, T. F, Heinze H.-J., and Prevedel, H. (1990). Event-related brain potentials reflect semantic and syntactic errors during language processing. *EEG EMG Z Verwandte Gebiete, 21*(2), 75-81.

Münte, T. F., Heinze, H., Matzke, M., Wieringa, B. M., Johannes, S. (1998). Brain potentials and syntactic violations revisited: no evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*, 217–226.

Näätänen R. (1995). The mismatch negativity—A powerful tool for cognitive neuroscience. *Ear Hear, 16,* 6-18.

Näätänen, R., Lehtoskoskl, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.

Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance, 31* (5), 912-918.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex, 2*, 244–258.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes – evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–65.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191–243.

Norris, D., McQueen, J., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1209–1228.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Onifer, W., & Swinney, D. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition, 9,* 225-236.

Osterhout L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences – evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 786–803.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., McLaughlin, J., Kim, A., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking ERPs and beyond.* Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: a means for exploring the neurocognition of second language processing. *Language Learning, 56,* 199–230.

Ostrin, R., & Tyler, L. (1995). Dissociations of lexical function: Semantics, syntax, and morphology. *Journal of Cognitive Neuropsychology, 12*, 345-389.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64,* B9-B17.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science, 12,* 445-449.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition, 24*, 501–528.

Pechmann, T., & Zerbst, D. (2002). The activation of word class information during speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 233–243.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision, 10*, 437-442.

Poeppel, D., Idsardi, W., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 363* (1493), 1071-86.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague, 4*, 79–96.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 421-435.

Proverbio, A. M., Cok, B., & Zani, A. (2002). Electrophysiological measures of language processing in bilinguals. *Journal of Cognitive Neuroscience, 14*(7), 994-1017.

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Randall, M. (2007). *Memory, psychology and second language learning*. John Benjamins.

Rayner, K., Pacht, J. M., & Du, S. A. (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words: Evidence from eye fixations. *Journal of Memory and Language, 33*, 527-544.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947-9.

Rossi, S., Gugler, M., Friederici, A., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*, 2030–48.

Salasoo, A., & Pisoni, D. (1985). Interaction of knowledge sources in auditory word identification. *Journal of memory and language, 24,* 210-231.

Sebastián- Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240–255.

Sebastián- Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B. (2006). First- and Second-language Phonological Representations in the Mental Lexicon. *Journal of Cognitive Neuroscience, 18* (8), 1277–1291.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). On-line processing of native and non-native phonemic contrasts in early bilinguals. *Cognition, 72*, 111-123.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489–537.

Sereno, S. C., Pacht, J. M., & Rayner, K. (1992). The effect of meaning frequency on processing lexically ambiguous words: Evidence from eye fixations. *Psychological Science, 3,* 296-300.

Shcherba, L. V. (1939). Phonetics of the French language. *An overview of French pronunciation in comparison with Russian: A manual for students of foreign languages departments.* Leningrad.

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception & Performance, 11,* 28-39.

Slowiaczek, L. M., & Hamburger, M. B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1239-1250.

Soto-Faraco, S., Sebastián-Gallés, N., and Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412–32.

Steinhauer, K., White, E., &. Drury, J. (2008). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research, 25*(1), 13–41.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of Acoustic Society of America, 111*, 1872-1891.

Strange, W., & Shafer, V. (2008). Speech perception in second learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (eds.), *Phonology and second language acquisition* (Vol. 36) (pp.153-191). Amsterdam/Philadelphia: John Benjamins.

Streeter, L. A., and Nigro, G. N. (1979) The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America, 65,* 1533–41.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoodsand the word–form representations of 14-month-olds. *Psychological Science, 13,* 480–484.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-659.

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., Arévalo, A., Vargha, A., & Bates, E. (2005). Times action and object naming. *Cortex, 41*, 7-25.

Tabossi, P. (1988).Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory & Language, 27*, 324-340.

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language, 32,* 359–72.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological research, 49*, 161-167.

Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics, 16*(4), 379–405.

Tanenhaus, M. K., & Donnenwerth-Nolan, S. (1984). Syntactic context and lexical access. *Quarterly Journal of Experimental Psychology, 36A,* 649-661.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. A. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal Behavior, 18,* 427-440.

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In A. K. Biller, E. Y. Chung, and A. E. Kimball (Eds.), *Proceedings of the 36th Boston University Conference on Language Development* (pp. 594-606). Somerville, MA: Cascadilla.

Tanner, D., Osterhout, L., & Herschensohn, J. (2009). Snapshots of grammaticalization: Differential electrophysiological responses to grammatical anomalies with increasing L2 exposure. In J. Chandlee, M. Franchini, S. Lord, and G.-M. Rheiner (Eds.), *Proceedings of the 33rd Boston University Conference on Language Development* (pp. 528-539). Somerville, MA: Cascadilla.

Tockowitz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event related potential investigation. *Studies in Second Language Acquisition, 2,* 173–204.

Trofimovich, P. (2008). What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research, 37*(5), 309-329.

Trubetzkoy, N. S. (1969). *Principles of phonology* (C.A. Baltaxe, translated). Berkley: University of California Press, Berkeley, CA. (Original work published in 1939).

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 528–553

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception and Psychophysics, 36,* 417–427.

Tyler, L. K. (1989). The role of lexical representations in language comprehension. In W. D. Marslen-Wilson (Ed.*), Lexical representation and process* (pp. 439-463), Cambridge, MA: MIT Press.

Tyler, L. K. & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics, 34,* 409-420.

Tyler, L. K. & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics, 38*, 217-222.

Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language, 26*, 638–657.

van Alphen, P., & McQueen, J.M. (2006). The effect of Voice Onset Time differences on lexical access in Dutch*. Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 178-196.

van Berkum, J. J. A., Zwitserlood, P., Brown, C. M., & Hagoort, P. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701–18.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13,* 967–85.

van Herten, M., Kolk, H. H. J., Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research,* 22, 241–255.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19*, 95–112.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 394–417.

Van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *Journal of the Acoustical Society of America, 111*, 1906–1916.

Vogely, A. J. (1995). Perceived strategy use during performance on three authentic listening comprehension tasks. *Modern Language Journal, 79*(1), 41–56.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 98–108.

Warren, P., & Marslen-Wilson, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics, 41*, 262–75.

Warren, P., & Marslen-Wilson, W. D. (1988). Cues to lexical choice: discriminating place and voice. *Perception and Psychophysics, 43*, 21–30.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256.

Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language, 85*, 231-244.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*, 1–25.

Werker, J. (1995). Exploring developmental changes in cross-language speech perception. In L. Gleitman & M. Liberman (Eds.), *Invitation to Cognitive Science, Volume 1: Language* (pp 87-106). Cambridge, MA: MIT Press.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49-63.

Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., & Münte, T. (1997). Brain potentials indicate differences between regular and irregular German plurals. *NeuroReport, 8*, 957–962.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics, 35,* 49–64.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics, 50*, 351–60.

Wolff, D. (1987). Some assumptions about second language text comprehension*. Studies in Second Language Acquisition, 9,* 307-326.

Yeung, N., Botvinick, M., & Cohen, J. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931–959.

Yokoyama, S., Miyamoto, T., Riera, J., Kim, J., Akitsuki, Y., Iwata, K., Yoshimoto, K., Horie, K., Sato, S., & Kawashima, R. (2006). Cortical mechanisms involved in the processing of verbs: An fMRI study. *Journal of Cognitive Neuroscience, 18*(8), 1304–1313.

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language, 39*, 14-32.

Zwitserlood, P. (1989). The locus of the effects of the sentential-semantic context in spoken word processing. *Cognition, 32*, 25-64.

Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition, 9,* 307-326.

Yeung, N., Botvinick, M., & Cohen, J. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931–959.

Yokoyama, S., Miyamoto, T., Riera, J., Kim, J., Akitsuki, Y., Iwata, K., Yoshimoto, K., Horie, K., Sato, S., & Kawashima, R. (2006). Cortical mechanisms involved in the processing of verbs: An fMRI study. *Journal of Cognitive Neuroscience, 18*(8), 1304–1313.

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language, 39*, 14-32.

Zwitserlood, P. (1989). The locus of the effects of the sentential-semantic context in spoken word processing. *Cognition, 32*, 25-64.