Applied Psycholinguistics **38** (2017), 909–942
doi:10.1017/S0142716416000497

# Quantifying contextual effects in second language processing of phonolexically ambiguous and unambiguous words

ANNA CHRABASZCZ
*National Research University Higher School of Economics, Moscow*

KIRA GOR
*University of Maryland*

Received: November 3, 2015 Accepted for publication: November 5, 2016

ADDRESS FOR CORRESPONDENCE
Anna Chrabaszcz, National Research University Higher School of Economics, 21/4 Staraya Basmannaya, Moscow 105066, Russian Federation. E-mail: anna.lukyanchenko@gmail.com

ABSTRACT
Second language (L2) speakers often experience difficulty discriminating speech sounds of the non-native language, which can result in phonolexical ambiguity. We report two experiments that examine how L2 Russian speakers may utilize contextual constraints for phonolexical ambiguity resolution during speech comprehension. L2 ambiguous words constitute minimal pairs with palatalized and unpalatalized consonants in the Russian language, where the phonological feature of palatalization marks semantic, morphological, or syntactic distinctions between words. L2 performance is compared to that of a control group of Russian native speakers. The results demonstrate that L2 listeners rely on contextual information for meaning disambiguation during sentence comprehension, but that the relative reliance on different types of context is task specific.

Speech comprehension is a critical component of communication; however, it presents a challenge for second language (L2) learners (Goh, 2000; Rubin, 1994; Vandergrift, 2007). One source of difficulty involves routine encounters with ambiguous speech input due to L2 listeners' less accurate perception of certain nonnative phonological contrasts, especially those that are not represented in the native (L1) phonology (Best & Tyler, 2007; Flege, 1993; Kuhl, 1991). When phonological difficulties systematically and persistently affect lexical access and retrieval, this creates phonolexical ambiguity. As a result, similar sounding words can become lexically confusable and may start competing for lexical selection, which slows down word recognition or leads to the retrieval of an incorrect lexical item. One of the best known examples of phonolexical ambiguity is the difficulty that Japanese learners of English have in discriminating English words with the /r/–/l/ contrast, such as "rock" and "lock" (Cutler & Otake, 2004; Lively, Logan, & Pisoni, 1993; Ota, Hartsuiker, & Haywood, 2009). Some other notorious examples include phonolexical difficulties of Dutch learners of English discriminating English words with the /æ/-/ɛ/ contrast (Broersma & Cutler, 2008, 2011; Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012), Spanish–Catalan bilinguals discriminating Catalan words with the /e/–/ɛ/ contrast (Pallier, Colomé, & Sebastián-Gallés, 2001; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer, & Díaz, 2006), and English learners of Russian discriminating words that are contrasted by hard (unpalatalized) and soft (palatalized) consonants (Bondarko, 2005; Chrabaszcz & Gor, 2014; Lukyanchenko & Gor, 2011). Of note, phonolexical ambiguity is not only confined to difficulties in discriminating between phonological minimal pairs that involve particular L2 phonemic contrasts. Evidence from eye-tracking studies demonstrates that words that contain L2-difficult sounds but do not constitute phonological minimal pairs can also become temporarily ambiguous and create spurious lexical competition. For example, hearing the onset *pa-* in "panda" activates "pencil" for Dutch listeners (Weber & Cutler, 2004) and the onset *r-* in "rocket" activates "locker" for Japanese listeners (Cutler, Weber, & Otake, 2006). Phonolexical ambiguity can also occur in quite distinct words due to weak and fuzzy form-to-meaning mappings in the L2 (e.g., "quarrel"–"squirrel"; Cook & Gor, 2015; Cook, Pandža, Lancaster, & Gor, 2016).

In recent years, the topic of L2 phonolexical ambiguity has gained much interest in the second language acquisition (SLA) literature, with many researchers aiming to identify its source and establish its consequences for spoken word recognition. Some authors claim that phonolexical ambiguity originates from representational deficits in the encoding of the L2 phonemic contrasts and results in homophonous lexical representations (Ota et al., 2009; Pallier et al., 2001; Sebastián-Gallés et al., 2005). Others show that L2 listeners encode lexical distinctions between ambiguous words despite compromised phonetic processing, and that confusable words are not treated as homophones in the mental lexicon (e.g., Cutler et al., 2006; Darcy, Daidone, & Kojima, 2013; Hayes-Harb & Masuda, 2008; Weber & Cutler, 2004). These studies have used different methodologies and various populations of L2 speakers to examine how phonological difficulties affect spoken word recognition in L2; accordingly, most of them have focused on the word-level analysis. However, performing a lexical decision task on words presented in isolation may draw upon different types of information compared to the task of comprehending natural speech. The latter entails attending not only to the low-level information but also to the higher level relationships between words (lexical–semantic and structural), which can mediate the interpretation of the spoken input and trigger linguistic predictions that may disambiguate ambiguous language signal. For example, comprehension difficulty can be reduced and word recognition can be speeded up if words are embedded in predictable sentence contexts rather than presented in isolation (Grosjean, 1980; Kellas, Paul, Martin, & Simpson, 1991; Salasoo & Pisoni, 1985). Moreover, it has been shown that context effects can accumulate over the course of a sentence. Thus, sentence-final words tend to be recognized faster than words occurring at the beginning of a sentence (Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978). Similarly, the amplitude of the N400 component has also been shown to reduce progressively over the course of contextually constraining sentences (Kutas, Van Petten, & Besson, 1988; Van Petten & Kutas, 1991), indicating that contextual information builds over time, across the sentence, to facilitate lexical access.

The present study sets out to fill the gap in the literature on phonolexical ambiguity in L2 by examining what kinds of sentential contexts contribute to the processing of ambiguous words during oral sentence comprehension. To that end, we examine L2 listeners' performance in two sentence-level tasks: a lexical decision task (LDT) with contextual priming and a self-paced listening task (SPLT). In the LDT, ambiguous words are presented at the end of the sentence, where contextual effects may be the strongest. In the SPLT, the same words are embedded in the middle of the sentence, where context effects may be weaker. Using the same stimulus words with the same phonological contrast (the hard and soft consonant distinction in the Russian language) in the two tasks, we want to see whether the well-known perceptual difficulties that L1-English speakers of L2-Russian have with the hard/soft consonant distinction carry over to discriminating the words that are differentiated on the basis of this feature during sentence comprehension. The idea is that if two given words are phonolexically confusable, L2 listeners will experience difficulty discriminating between them. During sentence comprehension, listeners may benefit from having contextual constraints, and may be able to build linguistic expectations that will guide them toward the intended interpretation of the word. We can examine the strength of contextual constraints by replacing a phonolexically ambiguous but contextually congruent word with its contextually incongruent counterpart. If L2 listeners cannot resolve phonolexical ambiguity, they will resort to context for help and will therefore experience a contextual bias effect, which is operationalized as a reaction time (RT) difference between contextually congruent and incongruent words. If this difference approaches zero, one can assume that L2 listeners rely on sentential constraints to the extent that they treat contextually incongruent ambiguous words as congruent. In the present paper, we consider three types of sentential constraints (semantic, morphological, and syntactic) to quantify which of them are potentially more effective in constraining the interpretation of ambiguous words during oral sentence comprehension.

## CONTEXT EFFECTS IN AMBIGUITY RESOLUTION

When ambiguous words are presented in isolation or without preceding context, their recognition is dependent only on lexical frequency. When the same words occur in sentential context, their identification also depends on the strength of constraints imposed by prior text (Duffy, Morris, & Rayner, 1988; Kellas et al., 1991; Rayner, Pacht, & Duffy, 1994; Tabossi, 1988; Tabossi, Colombo, & Job, 1987). Recent research also provides evidence that people can engage in predictive processing and preactivate contextually appropriate linguistic units and features in advance based on contextual cues (DeLong, 2009; Kutas, DeLong, & Smith, 2011; Lau, 2009). Therefore, context may facilitate word recognition and may be used to speed up resolution of linguistic ambiguities. Although the precise timing of ambiguity resolution and the relative weight of context and frequency effects are still a matter of debate, much research on ambiguity suggests that there are differences in how semantic and morphosyntactic ambiguities are resolved in speech comprehension (e.g., Lee & Federmeier, 2009).

Research on semantic ambiguity resolution suggests two stages in the processing of ambiguous words: at the first stage, multiple meanings of ambiguous words are activated; at the second stage, contextually appropriate meanings are preserved while contextually inappropriate meanings are deactivated and suppressed (by around 300–500 ms; Duffy et al., 1988; Faust & Chiarello, 1998; Kintsch & Mross, 1985; Rayner et al., 1994; Swaab, Brown, & Hagoort, 2003; Van Petten, & Kutas, 1987). People are consciously aware of only the resulting meaning and not the initial activation of multiple meanings.

While the effect of semantic constraints builds incrementally over the sentence, the effect of syntactic constraints seems to be more localized and deterministic, driven by the local phrase structure and not the global syntactic context (Marslen-Wilson & Tyler, 1980; Tyler & Warren, 1987). According to one of the most influential accounts proposed for syntactic ambiguity resolution, the parser generates a phrase structure based on the syntactic information about the words (e.g., grammatical classes) that have already been processed (e.g., after determiner a noun should follow). On a serial view of sentence processing, the parser initially commits only to one interpretation of the ambiguous structure. At the next stage, it integrates the resulting phrase structure with semantic and discourse information. If the generated interpretation does not bear out, the parser reparses the sentence, which usually incurs additional processing costs (Frazier, 1979; Frazier & Rayner, 1982). Studies on a subset of syntactic ambiguities, ambiguities that involve a change of syntactic category (e.g., the word "park" referring to a noun meaning "a recreational area" and to a verb meaning "to maneuver a car into a space"), demonstrated that syntactic context can be quite selective (Folk & Morris, 2003). This means that in syntactic ambiguity resolution, syntactic context can be used early on to reorder the available meanings of an ambiguous word and to eliminate competition from syntactically inappropriate interpretations (e.g., a verb instead of a noun in "We went to the park"). However, according to some constraint-based models of sentence processing, both lexically specific biases and referential pragmatics are computed in parallel and interact immediately to determine initial sentence structure (MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton & Sedivy, 1995; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993). The interpretation that is usually chosen is the one that best fits the context or involves the fewest presuppositions.

The views regarding the nature of morphological ambiguity resolution parallel those in the sentence parsing literature and suggest that context can have an effect on the interpretation of morphologically ambiguous words. For example, a study by De Almeida and Libben (2005) examined the effect of context on the morphological choice between two derivationally ambiguous words of the type "unlockable" (meaning either "un-lockable," i.e., "not able to be locked," or "unlock-able," i.e., "able to be unlocked") and found that only one of the word's meanings gets activated depending on the sentence context.

Collectively, a large body of work suggests that context can play a significant role in the processing of ambiguous words; however, none of the individual studies has directly contrasted effects of different kinds of contextual constraints (semantic, syntactic, and morphological) on ambiguity resolution. Moreover, most of the documented evidence comes from research on native language acquisition and processing. Because L1 and L2 speakers are faced with similar tasks in language comprehension (e.g., a need to cope with ambiguous speech signal), it is tempting to assume that, in order to resolve different kinds of ambiguities, L1 and L2 speakers resort to similar processing strategies. However, empirical evidence is needed to test such hypotheses.

## L2 SENTENCE PROCESSING

In order to benefit from contextual constraints, one must have the necessary knowledge about the sentence structure and its constituents, and be able to integrate multiple cues in a rapid manner. While native speakers have robust linguistic representations and efficient processing strategies that allow them to combine low-level and higher order linguistic information on the fly, L2 speakers have been shown to have both representational and processing deficits.

Much of the evidence in favor of representational deficits comes from studies documenting L2 speakers' high error rate in the use of morphosyntactic features and their insensitivity to grammatical violations, especially when the L2 grammatical element is not instantiated in the speakers' L1 (Chrabaszcz & Jiang, 2014; Franceschina, 2001; Hawkins & Chan, 1997; Jiang, Hu, Chrabaszcz, & Ye, 2015; Jiang, Novokshanova, Masuda, & Wang, 2011). Longitudinal studies show that even L2 speakers with extended language exposure, substantial practice, and long residency in the target language country continue to experience difficulties in accurately using grammatical morphemes in spontaneous speech (e.g., Jia, 2003; Lardiere, 2007). According to the shallow-structure hypothesis (Clahsen & Felser, 2006a, 2006b), the representations that L2 learners compute during sentence processing contain less morphosyntactic detail and lack complex hierarchical structure, as opposed to L1 representations in children and adults.

According to a different view, L2 speakers may represent L2 knowledge accurately, but may not be able to retrieve it due to processing constraints (McDonald, 2006). For example, one view holds that L2 listeners are predominantly bottom-up processors; they focus too much attention on identifying sounds and words that they do not have enough time or working memory capacity for building higher level units of representation (Berne, 2004; Block, 1992; Service, Simola, Metsanheimo, & Maury, 2002). Other evidence suggests that L2 speakers are more affected than native speakers by limitations on cognitive resources (such as less automaticity, slower speed of processing) and are less efficient at information integration. For example, Kilborn (1992) demonstrates that resource limitation prevents L2 speakers from successfully integrating multiple cues (syntactic and semantic), focusing on only one. Such processing difficulties of L2 speakers find confirmation in the fact that very often they perform at ceiling in offline tasks, yet show much worse results in more cognitively demanding online tasks (Ellis, 2005; Hopp, 2010; López Prego & Gabriele, 2014).

Yet another line of research argues that L1 and L2 processing is essentially the same as far as lexical–semantic processing (e.g., Moreno, Rodríguez-Fornell, &

Laine, 2008; Mueller, 2005) and morphosyntactic processing (e.g., Dekydtspotter, Schwartz, & Sprouse, 2006) are concerned. Note that the generally observed slowness in L2 does not necessarily point to distinct processing mechanisms from those deployed by natives. According to Kaan (2014), potential differences in L2 sentence processing can be explained by the same factors that drive individual differences in native speakers (such as differences in frequency biases, task-associated effects, competing information).

For example, much evidence points to the fact that L2 speakers, akin to L1 speakers, are sensitive to semantic and morphosyntactic constraints. L2 speakers systematically produce an N400 component in response to semantically incongruent words in electrophysiological studies, both in reading (Ardal, Donald, Neuter, Muldrew, & Luce, 1990; Moreno & Kutas, 2005; Ojima, Nakata, & Kakigi, 2005; Proverbio, Cok, & Zani, 2002; Weber-Fox, Davis, & Cuadrado, 2003) and in listening (FitzPatrick & Indefrey, 2007; Hahne, 2001; Hahne & Friederici, 2001; Mueller, Hahne, Fujii, & Friederici, 2005). Morphosyntactic processing in L2 has been shown to be more strongly mediated by proficiency compared to lexical–semantic processing. In their comprehensive review of event-related brain potentials (ERP) studies on syntactic processing in L2, Steinhauer, White, and Drury (2008) argue that absence of certain electrophysiological components associated with morphosyntactic processing holds only for less proficient L2 learners. At low levels of proficiency, morphosyntactic violations are not yet recognized as such by L2 learners, and as a result, the anomalies are perceived as a lexical problem (Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006). With the beginning of grammaticalization and proceduralization of L2 knowledge, learners begin to identify the structural nature of the problem, and attempt to integrate morphosyntactic information with other sources of information available in the input. In case of integration difficulty, a P600 component is elicited. As proficiency increases, the P600 amplitude also increases and starts to resemble that of native speakers, whereas at nativelike levels of proficiency, L2 speakers display LAN-like components preceding the P600 component, very similar to native speakers. For example, in a study by Rossi, Gugler, Friederici, and Hahne (2006), high-proficiency L2 learners of German and L2 learners of Italian showed the same ERP components (both E[LAN]s and P600) as native speakers of those languages for all syntactic violations.

However, even when L2 speakers have the right kind of semantic and morphosyntactic knowledge, it is not obvious whether they will use this knowledge to resolve linguistic ambiguities. While ambiguity processing in the native language has been investigated extensively, the topic has remained largely untouched in the SLA literature. Several studies that have investigated context effects on ambiguity resolution in the L2 (Elston-Güttler & Friederici, 2005, 2007; Frenck-Mestre & Prince, 1997) observe that L1 and L2 processing of lexical ambiguities is fundamentally similar. Using RT measures in conjunction with ERPs, Elston-Güttler and Friederici (2005, 2007) compared L1 and L2 processing of homonyms whose two most frequent meanings belong to the same syntactic category (e.g., nouns, as in "sentence"), or different syntactic categories (e.g., a noun and a verb, as in "trip"). At the stimulus-onset asynchrony (SOA) of 200 ms, both RT and ERP data showed priming effects for both L1 and L2 participants, indicating multiple access to homonym meaning. At the 500 ms SOA, RTs revealed that contextually inappropriate meanings were no longer active for either group, although the ERP results showed that activation of inappropriate meanings had not completely decayed for L2 speakers. However, at the stage of late processing (800 ms SOA), both RT and ERP data revealed that only contextually appropriate meanings were still active in both L1 and L2 groups. Such results suggest that L2 speakers show nativelike multiple access at an early stage of ambiguity processing, but are a little slower than L1 speakers later in processing when sentence context information is used to disambiguate meanings. However, both groups can achieve disambiguation based on semantic context by 800 ms. If L2 speakers are given a little more time, they can use that time to their processing advantage. For example, in a study by Feroce, Aziz, Chun, and Kaan (2016), Spanish learners of English silently read highly and weakly semantically constraining sentences word-by-word, in which the presentation of the critical word was sometimes delayed by 300 ms. They showed no N400 effect in the no-delay condition, but in the delay condition, an N400 effect was elicited, which was larger in the highly constraining contexts than in the weakly constraining ones. Thus, the delayed presentation of a critical word gave L2 speakers enough time to have activated the linguistic input and have combined it in a timely manner with the context in order to predict the upcoming word. A study by Hu and Jiang (2011) also reports that L2 speakers can benefit from the preceding context to anticipate upcoming words. Native English speakers and L1-Chinese speakers of L2-English performed a cross-modal priming task, in which the target word could be contextually congruent, neutral, or incongruent ("The girl mailed the letter without a stamp/sticker/stone"). Both native and L2 speakers showed facilitation effects for congruent targets, although L2 speakers' response latencies to neutral and incongruent trials did not differ. L2 speakers have been shown to engage not only in semantic prediction but also in morphosyntactic prediction. Foucart, Martin, Moreno, and Costa (2014) manipulated the ending of highly constrained sentences so that the critical noun was either gender congruent or not, and measured anticipation effects on the article preceding the noun. An N400 effect was observed at the article and the noun, followed by an anterior positivity at the noun region, suggesting that L2 speakers are able to anticipate morphosyntactic information in a similar manner as monolinguals.

The collective, available data indicate that L2 speakers (provided they are highly proficient) show similar, albeit slower, sentence processing mechanisms compared to L1 speakers, and that both groups can achieve disambiguation based on sentential context.

## THE PRESENT STUDY

Although much evidence on L2 speakers' use of context in sentence processing has accumulated over the past couple of decades, studies comparing effects of different kinds of contexts within the same experimental setup are not many. For example, a study by Chrabaszcz and Gor (2014) examined whether L2 listeners' identification of phonologically ambiguous words was biased by the context. L2 English speakers listened to sentences with contextually congruent or incongruent phonologically ambiguous targets (e.g., “The bees/*peas are in the hive”) and identified which word they heard (“bees” or “peas”). Morphologically and syntactically constraining contexts appeared to be more effective than semantically constraining contexts in biasing L2 listeners’ choice between the two ambiguous words. However, because the interpretation of the data was based on inferences from the accuracy measures taken at the endpoint of processing (a choice between the two words had to be made after sentence presentation) rather than continuously, it is not possible to say whether context effects took place during listening or at the stage of word selection at the time of the response, and whether the results reveal real-time context bias or participants’ strategic performance and explicit lexical–semantic and morphosyntactic knowledge.

The present study extends existing evidence on L2 use of context for ambiguity resolution by assessing L2 performance in conditions that are more sensitive to temporal and dynamic aspects of speech processing. The main goal of the study is to examine how L2 speakers can recruit semantic, morphological, and syntactic information arising from contextual constraints for anticipation and disambiguation of phonolexically ambiguous words during speech comprehension.

Phonolexically ambiguous words in this study are distinguished on the basis of the phonological feature of consonantal palatalization in the Russian language. Russian presents a unique case where the opposition of hard (unpalatalized, [–soft]) and soft (palatalized, [+soft]) consonants permeates almost the entire consonantal system and can be used contrastively, for example, вес (“weight” /v$^j$es/)–весь (“the whole of” /v$^j$es$^j$/).[1] The distinction between hard and soft consonants in Russian poses an allophonic split problem for English-speaking learners of Russian (Bondarko, 2005), because, unlike in Russian, hard and soft consonants are not separate phonemes, but can occur as allophones of the same phoneme in English (such as light and dark pronunciation of /l/ in different phonetic environments). The degree of discriminability difficulty between hard and soft consonants for L2 Russian listeners depends on a number of factors, such as position in the word, articulatory–acoustic correlates for each particular consonant pair, the effect of phonetic assimilation, and vowel environment (Chrabaszcz & Gor, 2014; Kochetov, 2002; Lukyanchenko & Gor, 2011; Rice, 2015). With this in mind, the materials for the present study were created such that the discrimination of the consonants by L2 listeners was deemed difficult based on the results of the previous studies, but not impossible, even if listeners only attended to the phonetic–acoustic information. Thus, all phonological minimal pairs in the study are contrasted on either /l–l$^j$/ or /t–t$^j$/ distinction; we intentionally did not include the most difficult contrasts, like /f–f$^j$/ and /p–p$^j$/ (Kochetov, 2002; Lukyanchenko & Gor, 2011). We hypothesized that discriminability difficulty of L2 phonemes should result in phonolexical ambiguity for L2 listeners, which will make them more reliant on context for word meaning anticipation and disambiguation.

The feature of palatalization in the Russian phonology allows for another very important affordance. Unlike in many previous studies, it is possible to examine effects of different kinds of contexts (semantic, morphological, and syntactic) on word anticipation and disambiguation within the same study and with the same L2 phonological contrast. Target words can be created such that a contrastive feature can mark lexical–semantic, morphological, and syntactic distinctions between words. By replacing a contextually expected (congruent) word with an unexpected word (incongruent) within a pair, it is possible to quantify semantic, morphological, and syntactic context effects on spoken word recognition when word identification poses a phonolexical ambiguity problem (ambiguous condition) and when it does not (unambiguous condition). Context effects are defined as the processing costs incurred when trying to integrate an incongruent word with the preceding segment of the sentence semantically, morphologically, and syntactically: the longer it takes to process the unexpected word, the greater the context effect is. Therefore, the magnitude of the context effect can be operationalized as an RT difference between RTs to a contextually congruent word and RTs to a contextually incongruent word ($RT_{congruent} - RT_{incongruent}$), where negative values should indicate inhibition associated with the processing of the incongruency. In the situation of phonolexical ambiguity, L2 listeners may weigh contextual information more heavily than the low-level phonetic–phonological information. They can be mistakenly guided by the context to interpret ambiguous incongruent words as congruent ones more often, in which case a context bias will be observed. Under such circumstances, RT difference between congruent and incongruent targets should approximate zero (attenuated inhibition) or constitute a positive value (facilitation).

With regard to the strength of each of the three contextual constraints under consideration, we do not make firm predictions. Because most of the previous studies have focused on either a semantic or a morphosyntactic aspect of sentence processing, it is difficult to predict how different context effects will compare within the same experimental framework and the same participants. If L2 listeners are more sensitive to lexical–semantic rather than morphosyntactic information during sentence comprehension (as the prevalence of literature suggests), context effects in the semantic condition will be the greatest. If, however, they pay more attention to morphosyntactic cues, those should prevail in word disambiguation.

Finally, the use of two different tasks, LDT and SPLT, with the same stimulus words will allow us to examine the effect of full versus partial context on word anticipation and recognition. Generally, because context effect builds over time and over the course of the sentence, words that occur at the end of the sentence are more expected and are recognized faster compared to words that occur earlier in the sentence (Kutas et al., 1988; Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978; Van Petten & Kutas, 1991). In our LDT, target words occur sentence-finally, whereas in the SPLT, they are embedded in the middle of the sentence. By varying the position of the target word in the two tasks, we want to examine whether listeners can benefit from both partial and full contexts in the LDT and SPLT, respectively, or only from the full context in the LDT task. Crucially, the LDT also targets word recognition directly and measures response latencies associated with deciding whether a given word is an existing word or not. Response latencies in the SPLT are a more covert measure of sensitivity to context. At the same time, both tasks are implicit in that participants are never asked to judge word incongruency explicitly. Both tasks have been shown to be sensitive to lexical and morphosyntactic variables as well as suitable for examining the time course of integrative auditory comprehension.

Table 1. *Linguistic profile of second language participants*

| | Mean | *SD* |
|---|---|---|
| Age when started learning Russian | 17.7 | 2.8 |
| Length of living in Russia (years) | 2.7 | 2.3 |
| Formal instruction in Russian (years) | 6 | 1.6 |
| Self-rated pronunciation | 7.2 | 1.5 |
| Self-rated proficiency | | |
| Oral | 7 | 1.2 |
| Listening | 7.8 | 1.1 |
| Reading | 7.6 | 1.4 |
| Writing | 6.7 | 1.5 |
| Self-rated knowledge of grammar | 7.2 | 1.3 |
| Cloze test | 21.7 | 1.8 |

*Note:* Self-reported responses were given on a scale from 1 (*minimal proficiency*) to 10 (*maximal proficiency*).

## METHOD

### *Participants*

L1 group included 24 native speakers of Russian (20 females: mean age = 32, range = 23–58). L2 group included 34 American speakers of Russian as a second language (20 females: mean age = 29.5, range = 21–50). Their brief linguistic profile is shown in Table 1. In order to make sure that L2 participants can perform the tasks and are familiar with the lexical items in the target phonological minimal pairs, only L2 speakers with advanced high level of proficiency or higher were included in the study. They were screened into the study based on their performance on a 25-item Russian cloze test, their responses to the language background questionnaire, and their Interagency Language Roundtable (ILR) score, if they reported one. Out of those people who were included in the study, 9 participants had an ILR score of 2+ corresponding to advanced high level of proficiency, 15 participants had a score of 3, and 3 people had a score of 3+, both scores corresponding to superior proficiency levels. Seven people did not report an ILR score, but were included in the study based on their cloze test score (equal to or higher than 19, roughly corresponding to advanced high proficiency or higher). All participants were familiarized with the experimental protocol approved by the institutional review board, after which they signed an informed consent form. All participants were reimbursed for their participation.

### *Target feature and contrasts*

Both tasks reported below used the same phonological minimal pairs as stimulus words. In the phonolexically ambiguous condition, target words were distinguished on the basis of consonant hardness or softness in the word-final position.

The feature could mark different kinds of contrasts. In the semantic condition, it marked lexical–semantic contrasts, for example, мат ("checkmate" /mɑt/)–мать ("mother" /mɑtʲ/). In the morphological condition, two verbal forms were contrasted: a verb infinitive and a third-person singular form in the nonpast tense, for example, говорить ("to speak" /gɑvɑˈrʲitʲ/)–говорит ("speaks" /gɑvɑˈrʲit/). In the syntactic condition, the phonological distinction marked different parts of speech, for example, a verb and a noun, as in брат ("brother" /brɑt/)–брать ("to take" /brɑtʲ/). The target minimal pairs did not always share the same orthographic representation; however, they were phonologically the same except for the final consonant, for example, балет ("ballet" /bɑˈlʲɛt/)–болеть ("to be sick" /bɑˈlʲɛtʲ/). The unambiguous (control) condition included a similar set of items with the same kinds of distinctions except that the minimal pairs differed on the basis of phonological contrasts, which were common to both Russian and English and were not expected to pose any discrimination difficulty for L2 speakers of Russian. For example, in суд ("court" /sut/)–суп ("soup" /sup/), word-final sounds /t/ and /p/ are easily differentiated by the place of articulation in both Russian and English.

### *Analytic approach*

Linear mixed effects models (LMEMs) with restricted maximum likelihood estimation were used to analyze log-transformed RTs for correct responses only. Experimental conditions and the interactions of interest were entered in the models as fixed factors. For random effects, we included subjects and items as varying intercepts to account for subject-specific and item-specific idiosyncrasies. By-subject random slopes were included to account for by-subject adjustment to the experimental conditions of theoretical interest only if they significantly improved model fit and resulted in converging models (Baayen, Davidson, & Bates, 2008; Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013).[2] Models were fitted with different intercepts (levels of the fixed factors) in order to allow for a comprehensive examination of the data; however, the output tables with the associated estimated coefficients, standard errors, degrees of freedom, the $t$ statistic, and the $p$ values are provided for representative models only. Estimated coefficients generated by the models can be interpreted as a change in the dependent variable brought about by a change of an independent variable from one level to another. All data analyses were performed in R statistical software (R Core Team, 2015), using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) and the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015).

## TASK 1: LEXICAL DECISION TASK WITH CONTEXTUAL PRIMING

### *Design and materials*

The stimulus materials consisted of four 240-sentence 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 (semantic, syntactic, or morphological), target (nonce words, congruent, incongruent: confusable or unrelated), and condition (ambiguous, unambiguous, or filler; see Table 2 for the materials design and examples).

Table 2. *Materials design and the targets' properties in Task 1*

| | | | | | | No. of Phonemes | | Lemma Frequency | | Surface Frequency | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Condition | Context | Target | Word | Transcription | Translation | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Ambig | Semantic | Congr | мать | mɑtʲ | mother | 3.5 | 0.5 | 67 | 93 | 32 | 53 |
| | | Conf | мат | mɑt | checkmate | 3.5 | 0.5 | 67 | 93 | 32 | 53 |
| | | Unrel | газ | gɑs | gas | 3.5 | 0.5 | 86 | 47 | 26 | 13 |
| | | Nonce | маф | mɑf | NA | 3.5 | 0.5 | NA | NA | NA | NA |
| | Morphol | Congr | говорить | gavaˈrʲitʲ | to speak | 6.2 | 0.5 | 361 | 691 | 361 | 128 |
| | | Conf | говорит | gavaˈrʲit | speaks | 6.2 | 0.5 | 361 | 691 | 361 | 128 |
| | | Unrel | говорим | gavaˈrʲim | we speak | 6.2 | 0.5 | 361 | 715 | 4 | 7 |
| | | Nonce | говорик | gavaˈrʲik | NA | 6.2 | 0.5 | NA | NA | NA | NA |
| | Syntactic | Congr | брать | brɑtʲ | to take | 4.1 | 1.1 | 260 | 429 | 59 | 112 |
| | | Conf | брат | brɑt | brother | 4.1 | 1.1 | 260 | 429 | 59 | 112 |
| | | Unrel | вниз | vnʲis | downward | 4.1 | 1.1 | 74 | 69 | 74 | 69 |
| | | Nonce | брам | brɑm | NA | 4.1 | 1.1 | NA | NA | NA | NA |
| Unambig | Semantic | Congr | храм | hrɑm | temple | 3.6 | 0.5 | 88 | 63 | 25 | 20 |
| | | Conf | храп | hrɑp | a snore | 3.6 | 0.5 | 88 | 63 | 25 | 20 |
| | | Unrel | долг | dolg | debt | 3.6 | 0.5 | 68 | 28 | 22 | 10 |
| | | Nonce | храк | hrɑk | NA | 3.6 | 0.5 | NA | NA | NA | NA |
| | Morphol | Congr | любим | lʲubʲim | we love | 6.5 | 1.5 | 918 | 686 | 22 | 20 |
| | | Conf | любишь | lʲubʲiʃ | you love | 6.5 | 1.5 | 918 | 686 | 22 | 20 |
| | | Unrel | любит | lʲubʲit | loves | 6.5 | 1.5 | 918 | 710 | 21 | 20 |
| | | Nonce | любик | lʲubʲik | NA | 6.5 | 1.5 | NA | NA | NA | NA |
| | Syntactic | Congr | жир | ʒir | grease | 3.6 | 0.7 | 203 | 298 | 36 | 40 |
| | | Conf | жил | ʒil | lived | 3.6 | 0.7 | 203 | 298 | 36 | 40 |
| | | Unrel | зря | zrʲa | in vain | 3.6 | 0.7 | 61 | 38 | 61 | 38 |
| | | Nonce | жих | ʒih | NA | 3.6 | 0.7 | NA | NA | NA | NA |

*Note:* Ambig, ambiguous; Congr, congruent; Conf, confusable; Unrel, unrelated; Morphol, morphological; Unambig, unambiguous.

Context manipulation concerned the type of the sentential constraint afforded by the pretarget fragment of the sentence. In semantically constraining sentences, the pretarget context created a semantic expectation for a certain word (target) to appear sentence finally, for example, "Grandfather took an ancient book from the bookcase and blew off the dust," where the target word "dust" is semantically highly predictable. In syntactically constraining sentences, the preceding context biased listener's expectations in favor of a certain syntactic category, for example, a verb rather than a noun: "An excited soldier announced that the general just arrived," where a sentence-final verb in the past tense is expected although the semantic content of the target can potentially vary (e.g., "landed," "left"). In morphologically constraining sentences, an expectation for a certain morphological form was created. For example, in the sentence, "After the yesterday's quarrel, Roman doesn't want to talk," an infinitive form and not any other form of the verb is expected.[3] There were a total of 96 constraining sentences per presentation list, with 32 sentences in each context condition (semantic, syntactic, or morphological). At the stage of materials design, the sentences were normalized. Native speakers of Russian who did not participate in the main experiment ($n = 18$) read the sentences and filled in the last word in the sentence. An agreement score for each sentence was then computed, with 100% indicating a unanimous agreement among all participants on the word that should conclude the sentence. The sentences to be used in the experiment were then recalibrated such that sentence-final words in the semantic condition were semantically predictable ($M = 72.2\%$, $SD = 18.8$), 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 semantically predictable ($M = 24.3\%$, $SD = 12.7$).

In the experiment, target words always occurred in the sentence-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 word-final phoneme. The same words were used as congruent and confusable targets, but the condition in which they occurred was balanced across the lists. Note, however, that for L1 listeners, confusable targets did not present an actual phonolexical ambiguity problem; they were only confusable for L2 listeners. Unrelated targets were not phonologically related to the highly predictable, congruent targets. Nonce words differed from the congruent and confusable targets only in the word-final phoneme. 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 controlled for word length, lemma, and surface frequency as best as possible given the constraints imposed by the available matches in the language (see Table 2).

One more experimental manipulation concerned the degree of the targets' perceptual difficulty. In order to examine whether the phonological contrast between hard and soft consonants in Russian created a lexical problem, we needed to compare phonolexically ambiguous items with those that pose no discrimination problem to L2 listeners. The ambiguous condition included words that were phonological minimal pairs based on the hard/soft consonant distinction. The unambiguous condition also included phonological minimal pairs, but they differed on phonological contrasts that are common to both Russian and English, and were hypothesized to pose no perceptual difficulty. An additional, filler, condition was added in order to make the experimental manipulations less obvious to the participants. Fillers were either neutral carrier sentences, which created no expectation whatsoever about the following word or a nonce word (e.g., "Now you will hear the word . . . sister/sisteb"; $n = 96$), or sentences with varying syntactic structures in order to make the syntactic pattern of the experimental sentences less repetitive (e.g., "Ivan is late for work because he overslept"; $n = 48$). All items within and between the conditions were randomized.

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 at the end of the congruent and incongruent sentences such that they did not differ physically or acoustically across the conditions. In order to ensure that the participants were attending to the pretarget context, comprehension of the sentences was evaluated with occasional comprehension questions ($n = 24$) following congruent trials. Eight practice sentences were added for task familiarization purposes.

### *Procedure*

The experiment was delivered with the DMDX software (Forster & Forster, 2003). Participants were randomly assigned to one of the four presentation lists. Stimuli in each list were presented in a semirandom 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 (target) in the sentence was a real Russian word or not, indicating their response by pressing the Yes or No button on the keyboard. The target was separated from the rest of the sentence by a 500-ms interval in order to provide enough time for L2 speakers to process and integrate the preceding context (e.g., Feroce et al., 2016) and was marked by an appearance of a fixation cross on the screen. When the response was made, accuracy feedback and response latency were briefly displayed on the screen, after which a new sentence started playing automatically. On some trials, sentences were followed by Yes/No comprehension questions. The total duration of the experiment was about 30–40 min.

### *Results*

Participants' comprehension of the sentences was analyzed: L1 listeners made 3.4% of errors ($SE = 0.9\%$) and the L2 group made 8.2% of errors ($SE = 0.9\%$). With regard to the lexical decision performance, the L1 group's error rate amounted to 6.5% ($SE = 1.3\%$) while the L2 group's error constituted 8.4% ($SE = 0.7\%$).

Table 3. *Mean reaction times (ms) across all experimental conditions in Task 1*

| | | | Context | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Semantic | | Morphological | | Syntactic | |
| Language | Condition | Target | Mean | *SE* | Mean | *SE* | Mean | *SE* |
| L1 | Ambig | Congruent | 1166 | 63 | 1261 | 54 | 1224 | 62 |
| | | Confusable | 1398 | 80 | 1402 | 80 | 1626 | 87 |
| | | Unrelated | 1342 | 130 | 1384 | 71 | 1530 | 123 |
| | Unambig | Congruent | 1044 | 44 | 1398 | 86 | 1168 | 86 |
| | | Confusable | 1271 | 71 | 1465 | 80 | 1403 | 83 |
| | | Unrelated | 1217 | 50 | 1494 | 107 | 1420 | 92 |
| L2 | Ambig | Congruent | 1289 | 57 | 1241 | 43 | 1224 | 54 |
| | | Confusable | 1208 | 54 | 1305 | 40 | 1318 | 51 |
| | | Unrelated | 1332 | 58 | 1429 | 57 | 1431 | 49 |
| | Unambig | Congruent | 1201 | 50 | 1260 | 38 | 1233 | 55 |
| | | Confusable | 1417 | 73 | 1489 | 67 | 1571 | 69 |
| | | Unrelated | 1335 | 60 | 1471 | 46 | 1491 | 63 |

*Note:* L1, First language; Ambig, ambiguous; Unambig, unambiguous; L2, second language.

The errors were evenly distributed across subjects and items. Filler condition and nonce words were not included in the RT analysis. Erroneous responses were dropped from the RT analyses; RTs were log-transformed. For reference, mean RTs in milliseconds for all experimental conditions are provided in Table 3.

*Phonolexical ambiguity.* In order to examine how L2 listeners process phonolexically ambiguous words in sentential contexts relative to unambiguous words and relative to a control group of native speakers, we built an LMEM with three fixed factors and their factorial interactions (language group: L1 or L2; condition: ambiguous or unambiguous; target: congruent, confusable, or unrelated), and two random intercepts for subject and item as well as a random by-subject slope for condition and target. The outcome of the model is provided in Table 4.

The results point to an inhibition effect in the L1 group for both types of incongruent targets (confusable and unrelated) compared to congruent targets in the ambiguous condition (confusable: coefficient estimate = 0.18, $SE = 0.02$, $t = 8.68$, $p < .001$; unrelated: coefficient estimate = 0.11, $SE = 0.02$, $t = 5.45$, $p < .001$). That the interaction between ambiguity condition and target type was not significant suggests that L1 participants processed both types of incongruent targets more slowly than congruent ones in both ambiguous and unambiguous trials. The difference between the two incongruent targets in the L1 group was not significant in either of the conditions, suggesting that L1 listeners perceived both of them as equally bad continuations of the sentences.

L2 listeners were not statistically different from the native listeners on congruent targets in either the ambiguous or the unambiguous condition. Of note, we observed a significantly different relationship between congruent and confusable

Table 4. *Linear mixed effects model output for the analysis of phonolexical ambiguity in Task 1*

| Fixed Effects | Estimate | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|
| Intercept group: L1; condition: ambiguous; target: congruent | 7.05 | 0.04 | 64 | 163.3 | <.001 |
| L1 effects | | | | | |
| Condition unambiguous | −0.03 | 0.02 | 288 | −1.04 | .3 |
| Target confusable | 0.18 | 0.02 | 148 | 8.68 | <.001 |
| Target unrelated | 0.11 | 0.02 | 100 | 5.45 | <.001 |
| Condition (unambiguous) × Target (confusable) | −0.02 | 0.03 | 66 | −0.68 | .5 |
| Condition (unambiguous) × Target (unrelated) | 0.001 | 0.03 | 135 | 0.09 | .93 |
| L1 × L2 Effects | | | | | |
| Group L2 | 0.03 | 0.05 | 56 | 0.57 | .57 |
| Condition (unambiguous) × Group (L2) | 0.01 | 0.02 | 693 | 0.23 | .82 |
| Target (confusable) × Group (L2) | −0.16 | 0.03 | 143 | −5.9 | <.001 |
| Target (unrelated) × Group (L2) | 0.001 | 0.03 | 102 | 0.12 | .9 |
| Condition (unambiguous) × Target (confusable) × Group (L2) | 0.2 | 0.04 | 66 | 4.59 | <.001 |
| Condition (unambiguous) × Target (unrelated) × Group (L2) | 0.04 | 0.04 | 138 | 1.12 | .27 |
| **Random Effects** | **Variance** | ***SD*** | | | |
| Intercept items | 0.01 | 0.08 | | | |
| Intercept subjects | 0.04 | 0.19 | | | |
| Condition unambiguous | 0.0001 | 0.01 | | | |
| Target confusable | 0.0008 | 0.03 | | | |
| Target unrelated | 0.002 | 0.05 | | | |
| Condition (unambiguous) × Target (confusable) | 0.008 | 0.09 | | | |
| Condition (unambiguous) × Target | 0.0008 | 0.03 | | | |
| (unrelated) | 0.0008 | 0.03 | | | |
| Residual | 0.048 | 0.22 | | | |

*Note:* L1, First language; L2, second language.

targets in the ambiguous condition in the L2 group (coefficient estimate = –0.16, $SE$ = 0.03, $t$ = –5.9, $p$ < .001) and a significant interaction between group and ambiguity condition for the confusable targets (coefficient estimate = 0.2, $SE$ = 0.04, $t$ = 4.59, $p$ < .001). This suggests that L2 listeners treated confusable targets differently from L1 listeners only in the ambiguous condition. In order to examine this finding more comprehensively, we ran the same LMEM with intercepts for different levels of the fixed factors. The results demonstrated that L2 listeners spent significantly more time on processing unrelated targets compared to congruent targets in both ambiguous (coefficient estimate = 0.12, $SE$ = 0.02, $t$ = 6.53,

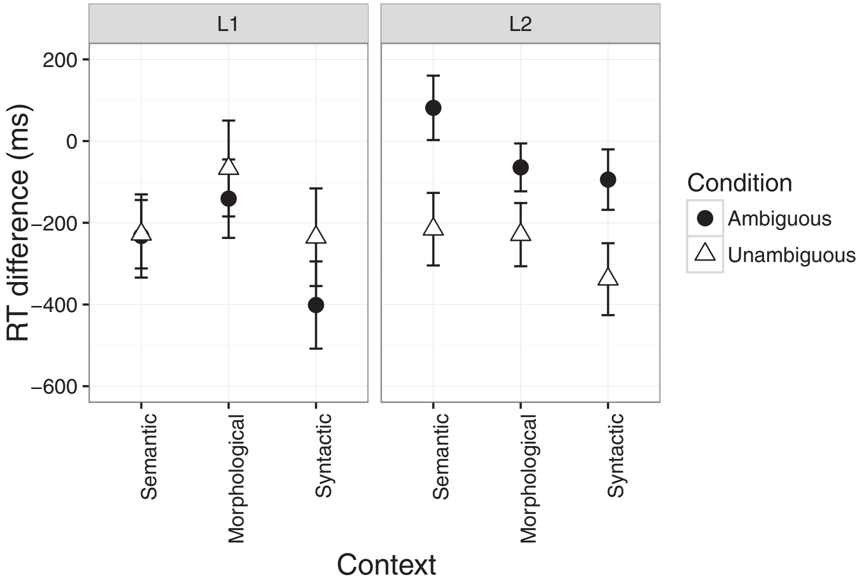


Figure 1. Contextual effects (ms) on word disambiguation in Task 1. Effects are calculated as $RT_{congruent} - RT_{confusable}$, where RT is the reaction time. Negative values indicate inhibition; positive values indicate facilitation.

$p < .001$) and unambiguous (coefficient estimate = 0.18, $SE = 0.02$, $t = 9.37$, $p < .001$) conditions, and that this trend was not statistically different from that observed in the L1 speaker group. However, unlike L1 listeners, L2 participants processed confusable targets longer compared to congruent targets only in the unambiguous condition (coefficient estimate = 0.2, $SE = 0.02$, $t = 9.6$, $p < .001$), but not in the ambiguous condition (coefficient estimate = 0.02, $SE = 0.02$, $t = 1.2$, $p = .2$). The difference between the two incongruent targets in the L2 group reached significance, with more inhibition observed for confusable targets relative to unrelated targets in the unambiguous condition (coefficient estimate = 0.04, $SE = 0.02$, $t = 2.1$, $p < .05$) and less inhibition for confusable targets relative to unrelated targets in the ambiguous condition (coefficient estimate = –0.1, $SE = 0.02$, $t = -4.7$, $p < .001$).

*Context effects.* To examine the strength of different types of contextual constraints on word disambiguation, we calculated RT differences between contextually congruent and incongruent confusable targets ($RT_{congruent} - RT_{confusable}$) for each context and ambiguity condition in both participant groups (see Figure 1 for graphical visualization of the effects across different experimental conditions). Incongruent unrelated targets were not included in this analysis because they constituted different lexical items unlike congruent and confusable targets. The latter were the same words used in different target conditions (e.g., мать "mother" used as a congruent target in one sentence and as a confusable target in a different sentence), which made it possible to compare these two types of targets directly. A negative RT difference indicates an inhibition effect; a positive RT difference indicates a facilitation effect.

It is already evident from Figure 1 that there is a difference in the strength of the contextual constraints in the ambiguous and unambiguous conditions in the L2 group, as we had predicted and discussed earlier. In the unambiguous condition, L2 listeners experience an inhibition effect when they encounter confusable targets. In contrast, in the ambiguous condition, L2 listeners produce a smaller inhibition effect (the RT difference between congruent and confusable targets is closer to zero in the morphological and syntactic conditions), or a facilitation effect (in the semantic condition). This suggests that although they do notice syntactic and morphological incongruences in the ambiguous condition, they do it less frequently compared to the unambiguous condition. Strikingly, in the semantic ambiguous condition, L2 listeners seem to treat incongruent confusable targets as legitimate continuations of experimental sentences suggesting a complete semantic bias effect.

To quantify the effect of the three types of contextual constraints on word identification statistically, we ran a series of LMEMs separately for each ambiguity condition and participant group. Each time, we fitted the models with different levels of the fixed factor (semantic, morphological, or syntactic) as intercepts in order to be able to compare the three contextual constraints with each other. The models included context type as a fixed factor and subjects as a random factor. The dependent variable was the magnitude of the context effect operationalized as the RT difference between congruent and confusable targets, averaged per subject and experimental conditions. The LMEM outcome is provided in Table 5.

In the L1 group, participants exhibited a significantly larger inhibition effect for syntactically constrained sentences compared to morphologically constrained sentences in both ambiguous (coefficient estimate = –260.32, $SE = 85.74$, $t = -3.04$, $p < .01$) and unambiguous (coefficient estimate = –194.77, $SE = 93.32$, $t = -2.087$, $p < .05$) conditions. The magnitude of context effect in the semantic condition did not differ significantly from that in the morphological or the syntactic conditions. In the L2 group, comparison of different types of context in the unambiguous condition did not yield significant results. Of note, in the ambiguous condition, L2 speakers experienced significantly more inhibition when they processed confusable targets in the morphological (coefficient estimate = –153.77, $SE =$ 51.57, $t = -2.98$, $p < .01$) and syntactic (coefficient estimate = 149.97, $SE = 85.74$, $t = 2.9$, $p < .01$) conditions compared to the semantic condition (where they showed facilitation). Morphological and syntactic conditions did not differ significantly from each other.

To summarize the results of Task 1, L2 speakers experienced less inhibition when they processed phonolexically ambiguous words occurring at the end of contextually incongruent sentences compared to (a) incongruent, but unrelated, words in the same sentences, (b) incongruent confusable targets in the control (unambiguous) condition, and (c) the group of L1 speakers processing the same words in the same sentences. Out of the three contextual constraints, the least inhibition (facilitation) for incongruent confusable targets was observed in the semantic condition compared to the morphological and syntactic conditions. These results suggest that L2 listeners experience a context bias effect when they encounter phonolexically ambiguous words with the hard/soft consonant distinction

Table 5. *Linear mixed effects model output for the analysis of contextual effects in Task 1*

| | L1 Ambiguous | | | | |
|---|---|---|---|---|---|
| Fixed Effects | Estimate | *SE* | *df* | *t* | *p* |
| Semantic–morphological | 91.32 | 85.74 | 45.94 | 1.06 | .29 |
| Morphological–syntactic | −260.32 | 85.74 | 45.94 | −3.04 | <.01 |
| Syntactic–semantic | 169 | 85.74 | 45.94 | 1.971 | .054 |
| Random Effects | Variance | *SD* | | | |
| Subjects | 244.4 | 15.63 | | | |
| Residual | 88212.8 | 297.01 | | | |
| | L1 Unambiguous | | | | |
| Fixed Effects | Estimate | *SE* | *df* | *t* | *p* |
| Semantic–morphological | 160.85 | 92.32 | 68.03 | 1.742 | .086 |
| Morphological–syntactic | −194.77 | 93.32 | 67.44 | −2.087 | <.05 |
| Syntactic–semantic | 33.92 | 93.32 | 68.03 | 0.363 | .71 |
| Random Effects | Variance | *SD* | | | |
| Subjects | 0 | 0 | | | |
| Residual | 102283 | 319.8 | | | |
| | L2 Ambiguous | | | | |
| Fixed Effects | Estimate | *SE* | *df* | *t* | *p* |
| Semantic–morphological | −153.77 | 51.57 | 65.95 | −2.982 | <.01 |
| Morphological–syntactic | 3.799 | 51.573 | 65.95 | 0.074 | .94 |
| Syntactic–semantic | 149.97 | 51.57 | 65.95 | 2.908 | <.01 |
| Random Effects | Variance | *SD* | | | |
| Subjects | 7769 | 88.14 | | | |
| Residual | 45215 | 212.64 | | | |
| | L2 Unambiguous | | | | |
| Fixed Effects | Estimate | *SE* | *df* | *t* | *p* |
| Semantic–morphological | −53.94 | 79.87 | 62.58 | −0.675 | .5 |
| Morphological–syntactic | −80.42 | 81.83 | 62.97 | −0.983 | .32 |
| Syntactic–semantic | 134.36 | 81.35 | 63.68 | 1.652 | .1 |
| Random Effects | Variance | *SD* | | | |
| Subjects | 21514 | 146.7 | | | |
| Residual | 106566 | 326.4 | | | |

*Note:* The first listed level of the fixed effect served as an intercept. For example, in the semantic–morphological comparison, the semantic condition served as an intercept. L1, First language; L2, second language.

and are mistakenly guided by the preceding context (especially in semantically constrained sentences) to treat contextually incongruent words as congruent.

## TASK 2: SELF-PACED LISTENING TASK

### *Design and materials*

The main stimulus manipulation involved the type of the target (congruent, confusable, or unrelated) and the type of the context condition (semantic, morphological, or syntactic). Manipulation of context condition was the same as in Task 1 with one exception. Instead of the sentence offset, the targets were embedded in the middle of the sentence, which was divided into eight regions. Regions could coincide with a word, or a phrase (sentence segment), and were 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 unrelated. Like in Task 1, confusable and congruent targets constituted a phonological minimal pair. They differed on the basis of consonantal hardness/softness in the word-final position. Each word in the minimal pair occurred as both a congruent and a confusable target, but never in the same presentation list. The experimental sentences with target manipulations constituted one-third ($n = 48$) of the total number of items ($n = 144$). The remaining items were filler sentences, which were added in order to make the experimental manipulation less obvious, and to balance the number of incongruent items per list such that 72 sentences were well formed and 72 sentences included semantic, syntactic, or morphological violations. Filler items were not included in the analysis.

Four presentation lists with pseudo-randomized 144 items each were created such that the same listener was never exposed to the same sentence or both words from the same minimal pair. In addition, comprehension Yes/No questions ($n = 72$) were presented visually on the computer screen to ensure that the listeners were attending to sentence meaning. After the button press, accuracy feedback briefly appeared on the screen. 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 & Weenink, 2010), and each segment was saved as a separate sound file. The sound files were strung together in such a way that the pretarget segments and the posttarget segments were acoustically identical across different target conditions, and the same targets did not differ physically between presentation lists.

### *Procedure*

The experiment was delivered with the DMDX software. Participants were assigned to one of the four presentation lists based on what list they completed in Task 1. The participants' task was to listen to the sentences presented through headphones and to answer comprehension questions as accurately as possible by pressing Yes or No buttons on the keyboard. 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 s if the button was not pressed. The total duration of the experiment was about 45–50 min.

### *Results*

First, accuracy of responses to comprehension questions was examined. L1 listeners made 3.4% errors ($SE = 0.7\%$) and the L2 group made 5.9% errors ($SE = 0.6\%$), indicating that both groups attended to sentence meaning. Second, erroneous responses were excluded (0.3% of data points) from the analysis of participants' listening latencies, which were computed as the time interval between the onset of each auditory segment until the button press. False alarms (RT $\leq$ 0) and timed-out responses (RT $\geq$ 4 s) were also removed, resulting in an additional 0.3% of data rejection. Figure 2 visually demonstrates how both groups of participants performed on the SPLT across all context conditions. Table 6 shows the output for the fixed effects structure in a series of LMEMs with log-transformed RTs as a dependent variable and target type as a fixed factor. The models also included random intercepts for subjects and items as well as a random slope for by-subject adjustment to the target. The analyses were run using the Nelder–Mead optimizer. The analyses were done for pretarget, target, and posttarget sentence positions, but because the effects did not turn out significant in the pre- or posttarget positions, the results of the analyses are only provided for the target position.

It can be seen from the figure and the table that L1 listeners slowed down in the target sentence position when they encountered incongruent targets (both confusable and unrelated) in all context conditions. The difference between listening latencies to congruent versus incongruent targets was significant in the semantic (confusable: coefficient estimate = 0.25, $SE = 0.09$, $t = 2.8$, $p < .01$; unrelated: coefficient estimate = 0.2, $SE = 0.08$, $t = 2.4$, $p < .05$) and the syntactic conditions (confusable: coefficient estimate = 0.14, $SE = 0.06$, $t = 2.3$, $p < .05$; unrelated: coefficient estimate = 0.14, $SE = 0.05$, $t = 2.7$, $p < .01$), but not in the morphological condition. In contrast, L2 listeners only slowed down in the semantic and syntactic conditions when they encountered unrelated targets (semantic: coefficient estimate = 0.13, $SE = 0.05$, $t = 2.72$, $p < .01$; syntactic: coefficient estimate = 0.12, $SE = 0.05$, $t = 2.56$, $p < .05$), but their listening latencies to confusable targets were not statistically different from those to congruent targets. Like native speakers, L2 speakers did not show a significant RT difference for processing incongruent targets in the morphological condition.

To summarize the results of Task 2, both L1 and L2 listeners spent more time on processing contextually incongruent unrelated words in the target position. However, only L1 listeners incurred processing costs (inhibition effect) for incongruent confusable targets, suggesting that L2 listeners treated incongruent phonolexically

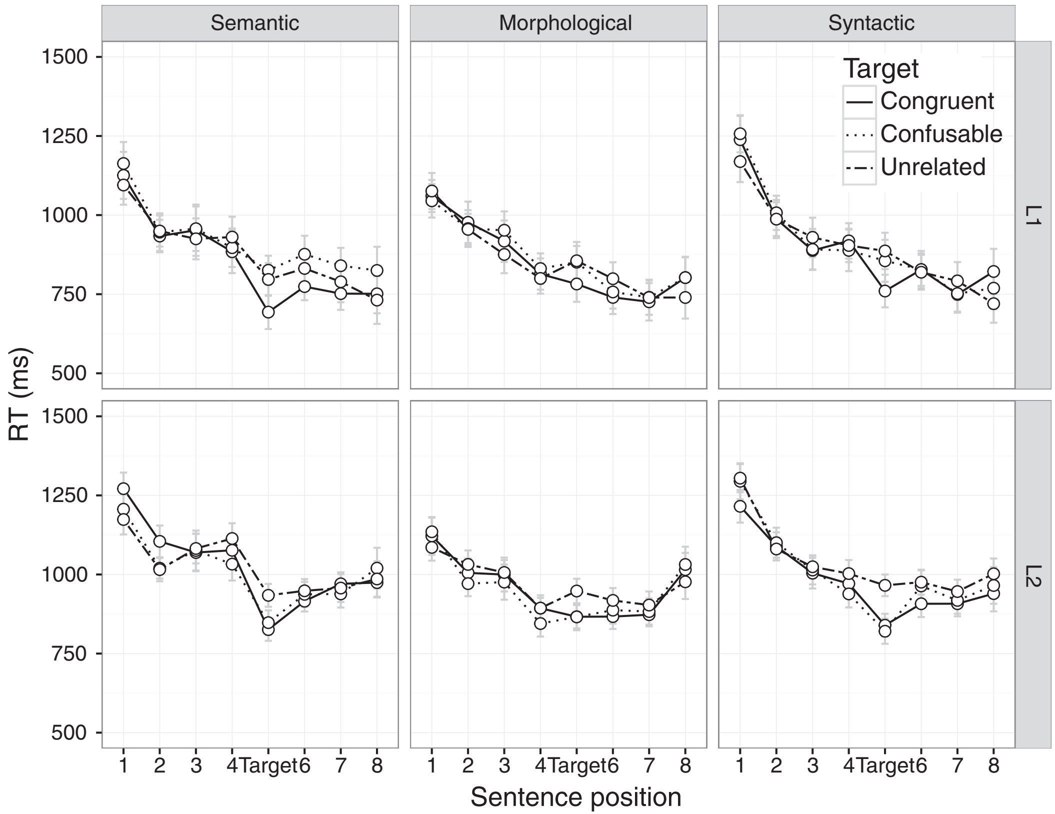


Figure 2. Mean listening latencies (ms) for each sentence position, computed across context conditions and participant groups.

Table 6. *Linear mixed effects model output (fixed effects) for the analysis of listening latencies in the target sentence position in Task 2*

| Context | Target | Estimate | *SE* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| First Language | | | | | | |
| Semantic | Intercept: Congruent | 6.35 | 0.13 | 24.7 | 50.68 | <.001 |
| | Confusable | 0.25 | 0.09 | 32.6 | 2.8 | <.01 |
| | Unrelated | 0.20 | 0.08 | 24 | 2.4 | <.05 |
| Morphological | Intercept: Congruent | 6.52 | 0.12 | 24.35 | 52.94 | <.001 |
| | Confusable | 0.10 | 0.07 | 68.9 | 1.57 | .12 |
| | Unrelated | 0.09 | 0.06 | 63.24 | 1.61 | .11 |
| Syntactic | Intercept: Congruent | 6.50 | 0.10 | 25.6 | 64.23 | <.001 |
| | Confusable | 0.14 | 0.06 | 198.2 | 2.3 | <.05 |
| | Unrelated | 0.14 | 0.05 | 315.2 | 2.7 | <.01 |
| Second Language | | | | | | |
| Semantic | Intercept: Congruent | 6.64 | 0.06 | 39.75 | 106.96 | <.001 |
| | Confusable | −0.04 | 0.06 | 37.47 | −0.67 | .51 |
| | Unrelated | 0.13 | 0.05 | 191.8 | 2.72 | <.01 |
| Morphological | Intercept: Congruent | 6.69 | 0.06 | 36.86 | 112.41 | <.001 |
| | Confusable | −0.03 | 0.05 | 74.78 | −0.66 | .51 |
| | Unrelated | 0.07 | 0.04 | 31.31 | 1.67 | .1 |
| Syntactic | Intercept: Congruent | 6.66 | 0.06 | 42.2 | 111.51 | <.001 |
| | Confusable | −0.07 | 0.06 | 115.1 | −1.26 | .21 |
| | Unrelated | 0.12 | 0.05 | 395.6 | 2.56 | <.05 |

ambiguous targets as congruent due to context bias. Out of the three context conditions, L1 listeners showed largest inhibition for syntactic and semantic violations, which points to a greater strength of semantic and syntactic sentential constraints on word anticipation and recognition. L2 listeners demonstrated a similar pattern only for unambiguous unrelated targets, but experienced a strong context bias in all three contextual conditions (semantic, morphological, and syntactic) when they encountered phonolexically ambiguous words.

## SUMMARY OF RESULTS FROM TASK 1 AND TASK 2

Both the LDT (Task 1) and the SPLT (Task 2) used the same target words, but in the LDT, they occurred at the end of the sentence, while in the SPLT, they were embedded in the middle of the sentence. By comparing the results of the two tasks, we wanted to examine the effect of contextual constraints on target word recognition in full (LDT) versus partial (SPLT) contexts. As can be seen from Table 7, L1 speakers demonstrate markedly larger context effects (operationalized as the magnitude of inhibition for incongruent targets) in the LDT compared to the SPLT for both types of incongruent targets (confusable and unrelated). The

Table 7. *Comparison of the magnitude of contextual effects (ms) for the ambiguous condition in Task 1 and Task 2 across different contexts and participant groups*

| | | LDT | | SPLT | |
|---|---|---|---|---|---|
| Context | Target | L1 | L2 | L1 | L2 |
| Semantic | Confusable | −232 | 81 | −132 | −22 |
| | Unrelated | −176 | −43 | −103 | −108 |
| Morphological | Confusable | −141 | −64 | −69 | 0.43 |
| | Unrelated | −123 | −188 | −73 | −81 |
| Syntactic | Confusable | −402 | −94 | −97 | 19 |
| | Unrelated | −306 | −207 | −127 | −126 |

*Note:* Effects are calculated as $RT_{congruent} - RT_{incongruent}$, where RT is the reaction time. Negative values indicate inhibition, and positive values indicate facilitation. LDT, Lexical decision task; SPLT, self-paced listening task; L1, first language; L2, second language.

observed pattern of results indicates that context effects tend to exert a stronger influence toward the end of the sentence (in LDT), although even partial context (in SPLT) is sufficient to trigger semantic, morphological, and syntactic expectations based on the available context. It should be noted that L2 participants also show larger context effects in the LDT compared to the SPLT for unrelated targets (except for the semantic condition), but they experience context bias for confusable targets, which is more pronounced in the SPLT (notice attenuated inhibition for the semantic condition and facilitation for morphological and syntactic conditions).

With regard to the strength of each type of context, we observed that L1 listeners show the largest inhibition when contextually incongruent words (both confusable and unrelated) occur in syntactically constrained sentences, a slightly lesser inhibition in semantically constrained sentences, and the least inhibition in morphologically constrained sentences: syntactic > semantic > morphological. The only exception is in the confusable target condition in the SPLT, where semantic context effects are larger than syntactic context effects: semantic > syntactic > morphological. The L2 group demonstrates a slightly different outcome for the strength of contextual constraints. In the LDT, L2 listeners experience the strongest inhibition in the syntactic condition, followed by the morphological condition, but an attenuated inhibition and even facilitation in the semantic condition: syntactic > morphological > semantic. The same trend is observed for both confusable and unrelated targets, but facilitation in the semantic condition for confusable targets suggests semantic context bias. In the SPLT, L2 speakers demonstrate a nativelike trend for context effects for unrelated targets: syntactic > semantic > morphological, but are more prone to context bias when they encounter confusable targets. The difference between the three contexts in the confusable target condition was not significant.

## DISCUSSION

We began this paper by introducing the term *phonolexical ambiguity* and argued that discriminability difficulty with certain phonemic contrasts in the L2 will have implications at the lexical level of processing.

When the phonological form is decoded, other properties of the word (e.g., morphological, syntactic, orthographic) are also accessed. In naturally occurring speech, words are embedded in sentences where they are combined with other words by means of complex semantic, syntactic, and morphosyntactic relationships. The results of the LDT and the SPLT show that both L1 and L2 listeners can use knowledge about such relationships and can anticipate the incoming input such that when their expectations are not met, a temporary breakdown in processing occurs. Moreover, comparison of the two experiments suggests that while listeners can benefit more from fully available contexts (in LDT), they can also use partially available contexts (in SPLT) to start disambiguating meaning. This is evidenced by greater processing costs (associated with recovery from a breakdown in contextual expectations) for both types of contextually incongruent targets in the L1 group and greater processing costs for incongruent unrelated targets in the L2 group. This means that listeners, including L2 participants, do not have to wait till the end of the sentence to integrate all the words within the sentence, but that they build their linguistic expectations incrementally as the sentence unfolds in time and use those to predict upcoming words. Thus, when L2 listeners have the necessary phonological representations in place, they can combine bottom-up information coming from low levels of linguistic analysis with contextual information coming from higher order processes in order to build expectations about the incoming signal similarly to L1 listeners.

However, incomplete or ambiguous phonological information can block lexical access, or make it more difficult. The results of our two experiments extend those of previous studies by establishing that phonolexical ambiguity can also carry over to comprehension at the sentence level. Specifically, by examining the distinctive feature of palatalization in the Russian language, we observe that L2 Russian listeners process contextually congruent and incongruent ambiguous words that differ minimally on this feature, such as мат (“stalemate” /mɑt/)–мать (“mother” /mɑt$^j$/), with similar processing costs. Thus, in the SPLT, L2 speakers of Russian only slow down significantly when they encounter unrelated (unambiguous) but not confusable (ambiguous) targets relative to contextually congruent targets, whereas native speakers of Russian slow down in the target sentence position when they encounter both types of incongruent targets (confusable and unrelated). Similarly, in the LDT, L1 listeners show a robust inhibition effect for both types of incongruent targets, while L2 listeners demonstrate significantly stronger inhibition for incongruent unrelated targets and only attenuated inhibition, and even facilitation, for incongruent confusable targets. These results suggest that when L2 listeners encounter phonolexically ambiguous words during oral sentence comprehension, they tend to rely on sentential contextual constraints for meaning disambiguation to the extent that they may treat contextually incongruent ambiguous targets as congruent (context bias). What this means outside of this particular experimental setting is that L2 listeners may be capable of compensating for low-resolution or fuzzy phonological representations by accessing and selecting the intended lexical item through its semantic, syntactic, and morphological characteristics, which can be derived through context.

The results of the present study indicate that there are both similarities and differences in how efficiently L1 and L2 listeners derive these kinds of information from context. Across both experiments, L1 participants experience the strongest context effect (the largest inhibition associated with target incongruency) in syntactically constrained sentences, a slightly weaker context effect in the semantic condition, and the weakest context effect in the morphological condition. The only exception is the confusable target condition in the SPLT, where L1 listeners show larger context effects in the semantic compared to the syntactic condition. In the SPLT, L2 participants demonstrate a nativelike pattern of context effects for unrelated targets: more inhibition in the syntactic condition, less inhibition in the semantic condition, and the least inhibition in the morphological condition (note that the degree of inhibition/facilitation for confusable targets did not differ significantly in semantically, syntactically, or morphologically constrained sentences, suggesting a similar context bias effect across all context conditions). A greater magnitude of syntactic and semantic context effects in the L1 group and in the L2 group (at least, for the unrelated targets in the SPLT) can be attributed to the fact that both syntactic and semantic violations necessarily including lexical violations (e.g., “brother–sister,” “brother–bring”), whereas morphological violations do not (e.g., “to bring–brings”). According to Levelt’s model of mental lexicon (Levelt, 1993), syntactic and semantic properties of words are closely connected in the lemma component, and are associated with word meaning. When listeners encounter semantic or syntactic violations in the sentence, they have to reanalyze the meaning component of the word, which, naturally, should increase processing times. Following the same logic, context effects are the smallest in the morphological condition because the phonological mismatch does not involve a lexical mismatch, as it does in the syntactic and the semantic conditions. When listeners encounter a morphologically incongruent word in the sentence, they have to recheck their morphological hypotheses while the meaning of the word remains unaffected. That is why it should be relatively easy to reevaluate 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 bring newspapers every day” as opposed to “They *brings newspapers every day”). In contrast, in case of semantic violations (e.g., “They bring *cities every day”) and syntactic violations (e.g., “They bring *put every day”), listeners need more time to recover from the comprehension breakdown.

Now, regarding L2 listeners’ performance in the LDT, they also demonstrate similar results to those observed in the L1 group in that they experience the strongest context effects in the syntactic condition; however, they differ from the native speakers in that they show weaker semantic context effects compared to the morphological condition. Thus, comparison of the L2 speakers’ results in the two tasks suggests that, when they have to disambiguate target words locally, with partially available sentential context (in the middle of the sentence in the SPLT), they are sensitive to sentential constraints in the following order: syntactic > semantic > morphological; when they process target words at the endpoint of sentence processing (in the LDT), they rely on syntactic and morphological constraints to a greater extent: syntactic > morphological > semantic.

Another interesting aspect of L2 participants' task-specific performance concerns the differences in the magnitude of inhibition that they display for ambiguous confusable targets in the two tasks. In the SPLT, L2 speakers show facilitation for the morphological and syntactic conditions and a very weak inhibition for the semantic condition (but because the differences in the magnitude of context bias for confusable targets in the SPLT were not significant among the three conditions, we will not focus on them). In contrast, in the LDT, L2 listeners reveal an opposite pattern: they show facilitation for incongruent confusable targets in the semantic condition, but demonstrate inhibition in the syntactic and morphological conditions. These results are intriguing and important in light of available literature on L2 discriminability of hard/soft consonants in Russian, but are difficult to reconcile. One explanation of an overall more pronounced context bias in the SPLT relates to the nature of the task. In the SPLT, participants are instructed to listen to the sentences segment-by-segment at their own comfortable pace; however, all listeners tend to press the button faster as they move through the sentence until the end of the sentence (Figure 2). Therefore, they may not have enough time in the middle of the sentence (target word position) to integrate contextual cues with low-level cues, which results in a strong context bias effect. In contrast, in the LDT, where target words are presented at the end of the sentences and are separated from the preceding sentence segment by 500 ms, L2 listeners may benefit from having more time and resources for integrating various cues and may pay more attention to the low-level phonological distinctions, although not as frequently or efficiently as they do in the unambiguous condition. To remind the reader, the perceptual distinction between hard and soft consonants is a difficult one for learners of Russian, but the degree of discriminability difficulty depends on a variety of factors, including the phoneme's position in the word, vowel environment, and the articulatory–acoustic correlates of individual consonant pairs. Although we did not specifically pretest our L2 participants on their perceptual discrimination of the target consonants for the present study, we purposefully selected only highly proficient L2 speakers of Russian and only those consonant pairs that have previously shown fair (and not homophonous) discriminability (Bondarko, 2005; Chrabaszcz & Gor, 2014; Kochetov, 2002; Lukyanchenko & Gor, 2011). The presence of some inhibition, albeit significantly reduced, for confusable targets in the morphological and syntactic conditions of the LDT suggests that L2 listeners attend to low-level phonological distinctions, but weigh contextual information higher when they process ambiguous words during sentence comprehension.

The difference in how L2 listeners treat ambiguous confusable targets in the syntactic and morphological conditions (inhibition) versus the semantic condition (facilitation) in the LDT requires additional consideration. On the one hand, this pattern of results may reflect differences in the encoding and the acquirability of the phonological feature, suggesting that the palatalization distinction in the Russian language might be more difficult to encode and to acquire (and thus more likely to be overridden by context) in arbitrary lexical items compared to when it signals morphological and syntactic distinctions. Remember, however, that the same trend (syntactic > morphological > semantic) was observed in the LDT for incongruent (unrelated) targets that do not pose an ambiguity problem. Taking this into account, weaker semantic effects in the L2 group in the LDT may not be rooted in the encoding problem of the palatalization feature per se, but may reflect the overall higher uncertainty and fuzziness at the lexical level. For example, Cook et al. (2016) argue that even words that do not involve phonologically difficult distinctions but that sound more or less similar, like "quarrel–squirrel," may be unfaithfully represented at the lexical level and can lead to the retrieval of incorrect semantic content (see "the fuzzy lexicon hypothesis" in Cook et al., 2016).

On the other hand, the observed pattern of results may indicate differences in L2 listeners' postlexical processing strategies in the LDT. Unlike SPLTs, where RTs reflect more implicit and automatic processing, LDTs register RTs that are associated with a more strategic and deliberate response when participants are required to make an explicit judgment about whether the target was a real word or not. Thus, a greater inhibition that L2 listeners display for incongruent confusable targets in the syntactic and morphological conditions compared to the semantic condition in the LDT may reflect their greater certainty about their knowledge of syntax and morphology compared to their higher uncertainty about lexical items. L2 listeners' greater reliance on syntactic and morphological constraints in the LDT in the present study is consistent with the results of another study by Chrabaszcz and Gor (2014) that used an explicit experimental paradigm. In their listening comprehension experiment, L2 speakers of Russian performed a forced choice task where they had to choose the word that occurred in the sentence out of the two given options, which could sometimes be ambiguous. The results also revealed that L2 learners rely on syntactic and morphological cues for identification of ambiguous words to a greater extent than on semantic cues.

In conclusion, the two experiments reported in this paper provide new data on L2 speakers' processing of phonolexical ambiguities and their use of contextual constraints for meaning disambiguation during oral sentence comprehension. This study also 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 SLA and classroom research. Our findings can potentially inform language teachers about difficult areas in L2 listening comprehension and about how L2 learners can compensate for such difficulties (e.g., which contextual cues they routinely employ or underuse). For example, teachers can encourage L2 learners to pay attention to contextual constraints by designing listening comprehension materials with high semantic and structural predictability and demonstrating to learners how they can compensate for their gaps in understanding by using contextual cues.

We acknowledge that although the present study provides some new insights into the problem of phonolexical ambiguity and context effects on meaning resolution in the L2, it is not without 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, and perceptual saliency). Second, the data were obtained from a sample of 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? Do mechanisms of context use change with changing proficiency? Examining performance of L2 speakers across different language proficiencies will provide a more complete picture and a better understanding of how bottom-up and top-down mechanisms develop and change.

## ACKNOWLEDGMENTS

The study was supported by the Graduate Dean's Dissertation Fellowship provided by the University of Maryland, College Park, and the Language Learning journal dissertation grant. The article was prepared within the framework of the Basic Research Program at the National Research University Higher School of Economics and supported within the framework of a subsidy by the Russian Academic Excellence Project 5-100. We are grateful to Professor Colin Phillips and Professor Nan Jiang for their thoughtful comments and suggestions. Any remaining errors are our own.

## NOTES

1. "ь" is a special letter in Russian that marks softness of word-final consonants in writing; "j" in the superscript of the IPA transcription represents consonantal softness.
2. In order to examine the effect of L2 proficiency, we carried out additional LMEMs with proficiency as a fixed factor for both lexical decision and self-paced listening tasks. However, L2 speakers with different proficiency levels did not demonstrate any significant differences in terms of ambiguity processing or context use; therefore, in the final analysis, they were combined into a single group of L2 speakers, which is contrasted with the control group of L1 speakers.
3. Although these types of expectations are not purely morphological in any given context, but morphosyntactic, here we refer to them as *morphological* since the target words in this condition are contrasted on the basis of the morphological form only, for example, говорить (/gɑvɑˈrʲitʲ/, speak$_{\text{INFINITIVE}}$)–говорит (/gɑvɑˈrʲit/, speak$_{\text{PRES/3rd/SING}}$).

## REFERENCES

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*, 328.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7 [Computer software]. Retrieved from http://CRAN.Rproject.org/package=lme4

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

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

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

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer software]. Version 4.5.16. Retrieved from 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.

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. *Quarterly Journal of Experimental Psychology, 64*, 74–95.

Chrabaszcz, A., & Gor, K. (2014). Context effects in the processing of phonolexical ambiguity in L2. *Language Learning, 64*, 415–455.

Chrabaszcz, A., & Jiang, N. (2014). The role of the native language in the use of the English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research, 30*, 351–379.

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.

Cook, S. V., & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *Mental Lexicon*, *10*, 247–270.

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology, 7*, 1345.

Cutler, A., & Otake, T. (2004). *Pseudo-homophony in non-native listening*. Paper presented to the 75th meeting of the Acoustical Society of America, New York.

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

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Mental Lexicon, 8*, 372–420.

De Almeida, R. G., & Libben, G. (2005). Changing morphological structures: The effect of sentence context on the interpretation of structurally ambiguous English trimorphemic words. *Language and Cognitive Processes, 20*, 373–394.

Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. A. (2006). The comparative fallacy in L2 processing research. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): The Banff Conference* (pp. 33–40). Somerville, MA: Cascadilla Proceedings Project.

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

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*, 680–689.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429–446.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.

Elston-Güttler, K., & Friederici, A. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language, 52*, 256–283.
Elston-Güttler, K., & Friederici, A. (2007). Ambiguous words in sentences: Brain indices for native and non-native disambiguation. *Neuroscience Letters, 414*, 85–89.
Faust, M., & Chiarello, C. (1998). Sentence context and lexical ambiguity resolution by the two hemispheres. *Neuropsychologia, 36*, 827–835.
Feroce, N., Aziz, P., Chun, E., & Kaan, E. (2016). *L2 learners need more time to predict*. Paper presented at the 29th Annual CUNY Conference on Human Sentence Processing, University of Florida, Gainesville, March 3–5.
FitzPatrick, I., & Indefrey, P. (2007). Effects of sentence context in L2 natural speech comprehension. *Nijmegen CNS, 1*, 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*, 87–99.
Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers, 35*, 116–124.
Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1461.
Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research, 17*, 213–247.
Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club.
Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.
Frenck-Mestre, C., & Prince, P. (1997). Second language autonomy. *Journal of Memory and Language, 37*, 481–501.
Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, 55–75.
Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*, 267–283.
Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*, 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.
Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research, 13*, 187–226.
Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research, 24*, 5–33.
Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*, 901–931.
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, PA: Benjamins.
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., Hu, G., Chrabaszcz, A., & Ye, L. (2015). The activation of grammaticalized meaning in L2 processing: Toward an explanation of the morphological congruency effect. *International Journal of Bilingualism*. Advance online publication.

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

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism, 4*, 257–282.

Kellas, G., Paul, S., Martin, M., & Simpson, J. (1991). Contextual feature activation and meaning access. In G. B. Simpson (Ed.), *Understanding word and sentence*. Amsterdam: Elsevier.

Kilborn, K. (1992). On-line integration of grammatical information in a second language. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 337–350). Amsterdam: Elsevier.

Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. *Journal of Memory and Language, 24*, 336–349.

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

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.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York: Oxford University Press.

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

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2015). ImerTest: Tests in linear mixed effects models. R package version 2.0-25 [Computer software]. Retrieved from http://CRAN.R-project.org/package=lmerTest

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: A case study*. Mahwah, NJ: Erlbaum.

Lau, E. F. (2009). *The predictive nature of language comprehension* (Unpublished doctoral dissertation, University of Maryland).

Lee, C. L., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language, 61*, 538–555.

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

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*, 1242–1255.

López Prego, B., & Gabriele, A. (2014). Examining the impact of task demands on morphological variability in native and non-native Spanish. *Linguistic Approaches to Bilingualism, 4*, 192–221.

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

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

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.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*, 381–401.

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.

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

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.

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

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.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition, 111*, 263–269.

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.

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

Rayner, K., Pacht, J. M., & Duffy, 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.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rice, H. (2015). *Perceptual acquisition of secondary palatalization in L2: Strengthening the bonds between identification and discrimination through multi-sequence category mapping* (Unpublished doctoral dissertation, Indiana University).

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*, 2030–2048.

Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal, 78*, 199–221.

Salasoo, A., & Pisoni, D. (1985). Interaction of knowledge sources in spoken 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*, 1277–1291.

Service, E., Simola, M., Metsanheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology, 14*, 383–408.

Spivey-Knowlton, M. J., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*, 227–267.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*, 276.

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

Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language, 86*, 326–343.

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

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

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

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*, 191–210.

Van Petten, C., & Kutas, M. (1987). Ambiguous words in context: An event-related potential analysis of the time course of meaning activation. *Journal of Memory and Language, 26*, 188–208.

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

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

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.