**Examining the Difference in Error Detection when Listening to Native and Non-Native Speakers**

Grace Sanders1 & Angela de Bruin1\*

1 Department of Psychology, University of York, York, United Kingdom.

\* Corresponding author:

Angela de Bruin

Department of Psychology

University of York

York YO10 5DD

UK

angela.debruin@york.ac.uk

**Abstract**

As communication with non-native speakers becomes increasingly common, it is important to understand how foreign-accented speech might influence language processing. Non-native speech can require the listener to process errors such as grammatical violations or unexpected word choices. The present study examines how listeners process different types of errors across native and non-native speakers. Using a self-paced listening task measuring reaction times to target words, 30 participants listened to sentences that contained either no error, a grammatical error (e.g.,” *Do you* ***wanting*** *anything?”*), or a contextual formal/informal word-choice error (e.g. *“Do you* ***require*** *anything?”* in an informal context). Participants responded more slowly while processing grammatical and word-choice errors compared to control sentences, especially when listening to non-native speech. This suggests that errors spoken by non-native speakers take longer to process, both in the case of grammatical errors as well as when contextually inappropriate words are used.

Keywords: language processing, error detection, non-native speech

**Introduction**

As travel has become more affordable and businesses have globalised, communication with non-native speakers has become more common. This has introduced us to a range of accents from all over the world (Barner-Rasmussen & Aarnio, 2011; Bloch, 1995). Still, learning a second language is not easy, and whilst all speakers may fall victim to committing linguistic errors from time to time, non-native speakers can struggle more with this (Corder, 1971; Schepens, van Hout, & Jaeger, 2020). Various components of language can differ between native and non-native speech, including accent (e.g., Abrahamsson & Hyltenstam, 2009), vocabulary or lexical diversity (e.g., Dewaele & Pavlenko, 2003), grammar (e.g., Clahsen & Felser, 2006), and discourse (e.g., Fuller, 2003). Despite deviations from native-speech patterns[[1]](#footnote-1) often being noticeable, native speakers only correct non-native speakers around 10% of the time (Chun, Day, Chenoweth, & Luppescu, 1982). Studies of error detection suggest that listeners process these linguistic errors, in particular grammatical errors, differently when they are produced by non-native speakers as compared to native speakers (e.g., Fairchild & Papafragou, 2018; Grey & van Hell, 2017; Hanulíková, Van Alphen, Van Goch, & Weber, 2012). However, less is known about how errors influence the actual speed of language processing. Furthermore, much of the previous research addressing the difference between native and non-native error detection focused on grammatical errors and/or salient semantic violations (e.g., Fairchild & Papafragou, 2018; Hanulíková et al., 2012). Non-native speakers, however, often also show more subtle differences from native speech, such as incorrect use of words that mismatch the context (Danesi, 2008; Romero-Rivas, Martin, & Costa, 2016). Because of this, the present study examined how non-native speech influences processing of both grammatical and word-choice errors.

**Behavioural effects of errors produced by native or non-native speakers**

Behavioural research assessing processing of errors made by a native or non-native speaker has mainly focused on offline measures asking listeners to evaluate or respond to sentences. For example, Gibson and colleagues (2017) asked listeners to interpret implausible sentences (e.g., “the mother gave the candle the daughter”) produced in a native or non-native accent. Listeners were asked a question after each sentence (e.g., “Did the daughter receive something?”) and were asked to answer according to what they thought the speaker intended to say. If participants answered “no” to this type of questions, their response was interpreted as listeners using the literal meaning. However, if participants answered “yes”, this was interpreted as listeners using a more plausible, non-literal interpretation. Listeners were more likely to follow a plausible (non-literal) interpretation for implausible sentences when the sentence was produced by a non-native speaker. Fairchild and Papafragou (2018) obtained similar findings when asking listeners to rate under-informative sentences (e.g., “Some people have noses with two nostrils”). While written sentences were used to remove any potential differences in processing difficulty as a consequence of listening to non-native speech, participants were told that the sentences came from either native or non-native speakers. They then had to judge how good (i.e., making perfect sense) or bad (i.e., making no sense) a sentence was. Reading about the writer’s background (native or non-native) influenced the ratings, with non-native under-informative sentences being rated more positively than native sentences. Fairchild and Papafragou interpret these findings by contrasting an intelligibility account (understanding non-native speech comes with additional processing demands) and an expectation-based account (arguing that listeners have different expectations for non-native versus native speech, with the former expected to be more variable and error prone). The expectation-based account furthermore predicts that listeners might rely more on top-down information (such as background knowledge and context) when processing non-native speech. Given that Fairchild and Papafragou used written sentences (which were identical for native and non-native speakers and thus did not differ in processing difficulty), they suggest the findings are best explained by an expectation-based account. Based on the description of the speaker, listeners might have used a more lenient approach when accepting sentences produced by a non-native speaker.

 These findings and interpretations are in line with the argument made by Lev-Ari (2015a,b), positing that when processing language produced by non-native speakers, listeners might alter their expectations. As a consequence, listeners might filter out some of the linguistic information and might end up with a less precise linguistic representation. To compensate, they might rely more on context than when listening to native speech. One example given is that when a non-native speaker refers to a “cake” instead of a “pie”, listeners might overlook this unexpected word choice and assume they are talking about the pie present in the room. This is compatible with the “good enough” approach listeners might use (e.g., Ferreira, Bailey, & Ferraro, 2002). Listeners might not always fully process the language input they are given. Instead they might only partially process the input, up to the level needed for the purpose (e.g., understanding what they speaker intends to say). This good enough approach might be used even more strongly with non-native speech, which might be processed less precisely.

Following this line of reasoning, listeners would be expected to show a smaller processing cost when hearing errors produced by a non-native speaker than when the same error is produced by a native speaker. In contrast, an intelligibility-based account would expect a larger processing cost as a consequence of increased processing demands associated with non-native speech. Behavioural research (e.g., Fairchild & Papafragou, 2018) supports the former explanation (differences in expectations and less precise processing of non-native speech leading to smaller error costs). Most of this research, however, has focused on offline tasks that ask participants to make judgements about sentences they hear or see (e.g., Fairchild & Papafragou, 2018), to respond to these sentences (e.g., Gibson et al., 2017), or to make judgements about similarities between utterances with and without errors (e.g., Pelzl, Lau, Jackson, Guo, & Gor, 2021, show that tone errors were less likely to be rejected when made by a non-native speaker of Mandarin). While this suggests that listeners might be more lenient in *response to* non-native language, these studies do not necessarily show us how native versus non-native error detection operates online (i.e., while processing speech). In other words: does processing of errors incur larger processing demands when the error is made by a non-native speaker (larger error cost) or, conversely, are listeners more lenient and “overlooking” errors made by a non-native speaker (smaller error cost)? Although we focus on behavioural processing of errors in our study, most research looking at online processing (as opposed to offline evaluations) has used electroencephalography (EEG). We will therefore briefly discuss this literature too.

**Grammatical Error Detection in ERP studies**

Focusing on grammatical errors first, Hanulíková et al. (2012) investigated grammatical error detection (gender disagreement between the definite determiner and noun, such as “de huis” (“the house”) instead of “het huis”) in native and non-native (Turkish) speakers of Dutch. Grammatical violations produced different ERPs (event-related potentials) in participants when listening to the native and non-native speakers. Grammatical violations elicited a larger P600 effect than control sentences in the native speech condition while no such difference between grammatical and control sentences was observed in the non-native speech condition. The authors interpret these findings in line with the P600 reflecting repair of grammatical errors. In line with the behavioural interpretation discussed above, listeners might modify their expectations when listening to non-native speech and might take a more tolerant approach overlooking grammatical errors. As a result, they may not attempt to repair grammatical errors in non-native speech. In a comparison between common and less common grammatical errors, Caffarra & Martin (2019) also showed that the P600 was absent when listening to common grammatical errors (gender violations such as “la color” instead of “el color”) in non-native speech, suggesting that listeners might not attempt to repair errors in non-native speech when they are expected. However, uncommon grammatical errors (number violations such as “los color”) did elicit a P600 in non-native speech too, suggesting listeners do attempt to make repairs when errors are not predictable.

Grey & van Hell (2017) also examined sentences containing a grammatical error such as “*Thomas was planning to attend the meeting but****he/\*she*** *missed the bus to school”,* reflecting an error in English subject pronouns that is frequently found in Chinese-English accented speech (used in their study). Similar to Hanulíková et al., grammatical errors elicited a larger ERP effect compared to control sentences in the native-speech condition, while this was not the case in the non-native speech condition. The study furthermore examined the role of familiarity with the accent. Contrary to the studies described above, which suggest that listeners are less likely to repair errors when they are familiar with the accent and/or type of errors, listeners familiar with the accent showed an ERP effect in non-native speech too while listeners *not* familiar with the accent did not. This could suggest that when listeners are not familiar with the accent, they might not have sufficient cognitive resources available to reanalyse the sentence.

EEG studies thus suggest that grammatical error processing differs between native and non-native speech, with listeners showing no difference between errors and control sentences in non-native speech (potentially depending on familiarity with accent and errors). These findings could be and have been interpreted in line with the “good enough” approach suggesting that listeners are more likely to overlook errors and rely on top-down knowledge when processing non-native speech. However, multiple interpretations are possible, including the interpretation that listeners unfamiliar with an accent might require too many resources to understand the speech and might therefore have insufficient resources left to repair or reanalyse any errors (Grey & Van Hell, 2017). Furthermore, from these ERP studies it is less clear how differences between native and non-native speech might influence behavioural outcomes. While comprehension accuracy was high in the studies discussed above, we do not know how accent might influence speed when processing errors, an important aspect of communication. Does accent influence how quickly listeners can process sentences with grammatical errors? The current study will therefore investigate how grammatical error processing, across native and non-native speakers, influences measures of behavioural processing.

**Semantic error detection**

While grammatical errors studied in the research described above often resemble the type of errors non-native speakers might make, this is often not the case for materials used when studying semantic or pragmatic violations. The behavioural studies described at the beginning often present participants with implausible or under-informative sentences. Similarly, ERP studies assessing semantic errors typically use words that are impossible in the context and that violate world knowledge (e.g., *“Kaitlyn travelled across the ocean in a****plane/\*cactus****to attend the conference”;* Grey & van Hell, 2017). ERP studies have shown mixed findings when assessing these types of semantic violations. Some studies show ERP effects in both native and non-native speech (e.g., Hanulíková et al., 2012; Grey & Van Hell, 2017), although the exact pattern of distribution might still differ between native and non-native speech (e.g., Hanulíková et al., 2012). Others have suggested that listeners experience greater difficulty processing and re-analysing semantic violations in non-native speech (Romero-Rivas et al., 2015) or integrating unrelated words (Romero-Rivas, Martin, & Costa (2016).

The type of semantic violations studied, however, are very unlikely to occur in either native or non-native speech. The semantic errors used in studies are very salient. For example, in the sentence “*Kaitlyn travelled across the ocean in a* ***plane/\*cactus***”, changing the semantic target ‘*plane*’ to ‘*cactus*’ is immediately more noticeable than incorrect gender identifiers used in grammatical error conditions (e.g., Grey & van Hell, 2017). Similarly, neither native nor non-native speakers are likely to make statements about a nose having two nostrils (Fairchild & Papafragou, 2018). Because of this, Hanulíková et al. (2012) argued that this type of semantic errors results in similar processing across accents because it is very rare to hear these kinds of linguistic errors, even from non-native speakers, whilst grammatical violations are fairly common in non-native speech (Van Hout, 2003). This critique shows the importance of investigating whether semantic error processing remains the same across accent type even with less apparent violations of world knowledge.

 In recent research, Gosselin, Martin, Navarra-Barindelli, & Caffarra (2021) created semantic errors that non-native speakers make in daily life as a consequence of mispronunciations. In the study, native Spanish listeners heard both native and non-native (Chinese-accented) speakers read correct sentences and incorrect sentences with typical and atypical phoneme substitutions (for example, substitution of /r/ to /l/, which is often made by Chinese speakers, or /n/ to /s/, which is less common). Sentences with phonemic substitutions led to semantic anomalies but, in the case of typical substitutions, represented errors non-native speakers could make in natural speech. Whilst EEG data showed that there was no difference between well-formed and semantically incorrect sentences in an early time window (300-600ms), the later time window (600-900ms) showed a larger negativity for incorrect versus correct sentences in non-native speech than in speech with a native accent. This was found when listening to both common and uncommon semantic errors. This could suggest that whilst semantic errors may initially seem ‘overlooked’ (cf. Hanulíková et al., 2012), error detection may be present in later processing, potentially because listeners have a harder time integrating semantic errors when listening to non-native speakers.

**Contextual Error Detection**

While Gosselin and colleagues (2021) created semantic violations that can occur in daily life, the semantic violations might still have been stronger than what would be expected in natural speech, where mispronounced phonemes do not always lead to strong anomalies. Rather than focusing on semantic anomalies we therefore studied a different type of semantic inconsistencies in our study, namely contextually inappropriate language use. Acquiring a new language does not just requiring learning new vocabulary and grammar rules but also learning how to use this language in different contexts. Non-native speech has been found to show violations of pragmatics or discourse. For example, Fuller (2003) investigated native and non-native speech in more formal contexts (interviews) and informal contexts (conversations). Native speakers used their discourse markers (such as “well” or “like”) in different ways depending on the type of interaction and familiarity with the interlocutor. Non-native speakers, despite having a very high proficiency, did not adapt their use of discourse markers to the interaction they were in to the same extent as the native speakers. This suggests that non-native speech might show less adaptation to the context it is used in and might be less sensitive to register differences (Gilquin & Paquot, 2008). Non-native speakers might struggle to acquire and correctly use the more colloquial language that is often used in informal contexts (e.g, Frumuselu, De Maeyer, Donche, & Plana, 2015), especially when they learn their second language in a classroom setting. At the same time, non-native speakers also struggle with formal contexts and often use less “polite” language when in contexts requiring more formal use (e.g., overly colloquial use when emailing a teacher, Biesenbach-Lucas, 2007; Economidou-Kogetsidis, 2011, or in academic settings, Gilquin & Paquot, 2008).

Earlier research has utilised contextual errors to examine what effect these violations have on linguistic processing in native speakers. Van Berkum (2009) used contextual errors to examine how ‘referential pragmatism’ (the process of building semantic meaning from sentences) can influence linguistic processing. Across multiple studies, participants showed to be sensitive to violations when sentences are not expected given the speaker or context *(e.g. “Every evening I drink some wine before I go to sleep”* spoken by a child*).* Thisshows that listeners are not just sensitive to implausible words that violate world knowledge but also to contextually inappropriate word choices when building semantic meaning from sentence content (e.g., Van Berkum, Brown, Hagoort, & Zwitserlood, 2003; Van Berkum, Hagoort, & Brown, 1999; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). However, this was only investigated within native speakers, leaving open the question of how contextual violations can influence processing of non-native speech.The current study therefore examined how, in addition to grammatical errors, listeners process contextual “word-choice” errors (i.e., overly formal words in an informal setting or overly informal words in a formal setting) in native and non-native speech.

**The Current Study**

The literature thus suggests that listeners might show different patterns in response to errors in native speech than in non-native speech, suggesting that listeners might be more tolerant when listening to non-native speech and follow a “good enough” approach in which they do not repair errors (Lev-Ari, 2015a,b). Studies assessing semantic or pragmatic violations, however, show more mixed results (e.g., Hanulíková et al., 2012; Romero-Rivas et al., 2015) but also often use errors that do not typically occur in either native or non-native speech. Hence, the current study’s aims are twofold. First, we aimed to examine behavioural outcomes when listeners hear native and non-native speakers commit linguistic errors. Most previous work has focused on ERP effects or on offline behavioural processing (e.g., ratings or response to sentences), but little is known about how errors might influence the speed of language processing in different types of speakers. Second, we aimed to examine whether error processing differs between grammatical and contextual errors, when those contextual errors represent word-choice errors non-native speakers make in daily life as opposed to salient violations of world knowledge.

To answer these questions, native English speakers listened to a set of sentences (spoken by a native speaker or a non-native speaker of English) which were divided between two contexts (either a ‘friend’ or a ‘teacher’ speaking to them) and three error variants (no-error, grammatical error, or a contextual word-choice error). The task measured reaction times (RTs) while listening to the sentences in a self-paced listening task. This acted as a measure of behavioural processing; the more processing needed, the longer we expected the RTs to be. Previous research has shown that these self-paced listening tasks are sensitive to effects detected in other language paradigms (e.g., Waters & Caplan, 2004).

Considering previous research (e.g., Fairchild & Papafragou, 2018; Gibson et al., 2017; Grey & van Hell, 2017; Hanulíková et al., 2012), we expect that processing will differ when listening to native and non-native speakers commit errors, signifying that error processing varies across native and non-native speech. If listeners are indeed more tolerant of errors in non-native speech and follow a “good enough” approach more strongly for non-native than native speech (Lev-Ari, 2015a,b), we expect the RT error cost to be smaller when listening to non-native speech as compared to native speech. The comparison between grammatical and contextual word-choice errors will allow us to examine whether processing differences between native and non-native speech depend on the type of errors made.

**Methods**

The study was pre-registered on the Open Science Framework (<https://osf.io/tyu94>). The data are available on <https://osf.io/u367c/>. The study received ethical approval from the ethics committee in the Department of Psychology at the University of York (number 20216).

**Participants**

According to power calculations based on medium-sized effects using *G\*Power*, 28 participants were needed in order to satisfy 80% statistical power, thus the sample size aim was 30 (Faul, Erdfelder, Lang, & Buchner, 2007). The final sample included 30 participants. An additional four participants were recruited but three were omitted due to not meeting eligibility requirements (indicating they were born outside of the UK and had an ‘American’ accent, while the native speaker in our study had a British accent), whilst another participant was removed due to showing abnormally fast responses (2 standard deviations, SD, below the group mean). It should be noted that our pre-registered exclusion criteria focused on low accuracy on the comprehension questions and not fast RTs in the listening task as a reason for exclusion. Given, however, that these RTs were faster than the duration of the audio they were asked to listen to, it was unlikely that this participant paid sufficient attention to the sentences. Participants were recruited through word-of-mouth and social media platforms using a link to the online experiment. Eligibility criteria required participants to be 18 years or older, monolingual native English speakers, and currently residing in the UK. Gender and age of participants was not recorded. There was no incentive for participation.

**Design**

This study included two within-subject variables: type of speaker (either native or non-native) and type of error (no-error, grammatical error, or a contextual formal/informal word-choice error, which will be referred to as a word-choice error). The dependent variable was RTs in response to the target cluster, which was measured in milliseconds. RT was calculated as the time participants needed (relative to onset of the audio) to indicate (by pressing the spacebar) that they were ready to move onto the next word-cluster.

**Materials/Stimuli**

Stimuli consisted of 20 sentence triplets, half of which were assigned to a ‘friend’ context, meaning they were common things a friend would say, and half to a ‘teacher’ context, meaning they were common things a teacher would say. A teacher and friend context were used to create a formal and informal environment respectively. Participants were told about the context they were in, with associated images provided to enhance context salience. For the friend context, the image supplied was of two male cartoon boys standing with their arms around each other, and for the teacher context, a similar style image was provided of an elderly male with a notepad and a pencil (See Figure 1 and Appendix C).

Each sentence triplet was created to have three versions of the target-word cluster: a no-error, a grammatical error, or a formal/informal contextual word-choice error on a specific target word (see Appendix A and Table 1 for an example). Grammatical error conditions contained a range or errors (see Appendix A for all sentences). This included, for example, incorrect use of verbs (e.g., “do you wanting” or “did you gone”), plural instead of singular nouns (e.g., “an emails”), and incorrect use of nouns or adjectives (e.g., using “danger” instead of “dangerous”).

The word-choice errors were designed so that the target-cluster contained either a formal word, which would be unusual if heard from a friend, or an informal word, that would be unusual if heard from a teacher (Krashen, 1976). In neither the grammatical errors nor the word-choice errors did we specifically consider the type of errors Spanish native speakers might make. In both conditions, errors reflected grammatical and choice mistakes that could be made by English non-native speakers more generally. Considering that we did not know how familiar participants would be with Spanish speakers specifically, we aimed to include errors that would be salient regardless of familiarity with the specific accent. Furthermore, considering that our word-choice error condition included a range of word classes (including nouns, adjectives, and verbs), we did not pick one type of grammatical errors (e.g., number agreement) but used a range of errors.

**<**Insert Table 1 around here>

Apart from the target word/words, sentences were identical. This ensured that any effects of error were not due to differences in the sentence context used. Sentences did differ between the friend and teacher context (to create sentences that are likely to be produced by either a teacher or friend), but a comparison between these two contexts was not part of this study and the analysis collapsed across these two parts. Nevertheless, we ensured sentence length was similar for the two contexts.

 Target words in the “word-choice error” condition differed from target words in the no-error and grammatical error condition. Due to the nature of these words, there was a difference in frequency between the word-choice and control/grammatical target words. Formal or informal words (e.g., “perspire” instead of “sweat” or “lads” instead of “males”) are unavoidably less frequent (*M* log frequency = 0.95, *SD* = 0.62; Davis et al., 2008) than their more common control equivalents (*M* log frequency = 2.00 *SD* = 0.70). Additionally, the control words were slightly shorter (*M* syllables = 1.45; *SD* = 0.60; *M* phonemes = 4.25, *SD* = 1.45) than the word-choice condition (*M* syllables = 2.05; *SD* =0.85; *M* phonemes = 5.47, *SD* = 1.71). As an additional control condition, we therefore also presented participants with these target words in the more appropriate context (i.e., the more formal words such as “perspire” in the teacher context and the more informal words such as “lads” in the friend context).

Sentences were recorded in English by a native speaker (who had what could be described as a British ‘received pronunciation’ accent) and a non-native speaker (with a Chilean Spanish accent). Both speakers were male and around the same age. Audio from the speakers was processed using ‘*Praat*’ audio-editing software (Boersma & Weenink, 2021). Sentences were divided into clusters of words that were deemed to ‘naturally flow’ on from one another (see Appendix A). The word-cluster containing the target word was referred to as the ‘target-cluster’; each target-cluster ended with the target word and target-clusters were never at the end of sentences, to accurately measure RT of participants when indicating they were ready to move on to the next part of the sentence.

For the experiment, non-target clusters were taken from the same reading of a sentence (usually the no-error condition), the target cluster from each condition was inserted, and volume was standardised to 70db. Each of the clusters were cut to ensure they had around 50ms of silence between the start and end of the clip. Finally, the non-native speaker clusters were increased in speed by 20%, to match speaking rate between the native and non-native audio.

As an attention check and to ensure sentences were comprehendible in both the native and non-native speaker condition, we also generated a set of forced-choice comprehension questions asking about the content of the sentences. Each sentence had its own comprehension question (see Appendix B). These were distributed across participants so that each participant saw a question on 12.5% of trials, distributed across speaker and error condition.

**Procedure**

The study was run using *Gorilla.sc* (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020).

Participants accessed the experiment through their own computer device and we ensured that the study could only be completed on a desktop or laptop and not on a phone or tablet.

Participants first read and completed their information and consent forms, confirmed they were 18 years or older, and confirmed they met the necessary eligibility criteria (monolingual native speaker of English, where they were born, and what accent they had). Next, participants completed a sound check, prompting participants to move to a quiet area where they would not be distracted. This sound check allowed participants to adjust their volume if necessary and made sure that sound clips could be played automatically on the participant’s device.

The programme then randomly allocated which half of participants heard the friend sentences first and which half heard the teacher sentences first. The order of trials in the experiment was also pseudo-randomised so that no more than three consecutive trials had the same speaker or error condition. Instructions stated that participants would hear a range of sentences spoken to them and they should imagine the speaker as their friend or teacher, dependent upon which context they were completing. Each sentence was split into clusters of words and they had to indicate, by pressing the spacebar, when they were ready to move onto the next cluster (see Figure 1). Additionally, after each trial, they saw a ‘Start Next Sentence’ screen or a forced-choice comprehension question, which they had to answer through keys on their keyboard (see Figure 1).

<insert Figure 1 around here>

Participants first heard a practise sentence (produced by the researcher) with a comprehension question. Once participants either worked through the friend or teacher sentences, instructions were shown again (this time with the alternative context) and the practise question was repeated (again, with the alternative icon), familiarising them to the change of context, before completing the alternative block. The experiment included 120 experimental trials, 60 per context (friend/teacher). This included 40 no-error, 40 grammatical, and 40 word-choice trials (20 per speaker).

Additionally, to control for differences in frequency between the word-choice targets and control/grammatical targets, participants also saw “word-choice” controls in their corresponding context. This meant that in the ‘friend’ context block, participants also heard the informal word-choice sentences and in the ‘teacher’ block participants heard the formal word-choice sentences (see Table 1). This was done to investigate if processing of the inappropriate word-choices (formal in friend context and informal in teacher context) were indeed impacted due to the use of contextually incorrect words or confounded by the frequency of these words, as uncommon words are harder to process than common ones (‘The Word Frequency Effect’; Brysbaert et al., 2011).

After completing the self-paced listening task, participants completed post-study questions which asked where they believed both of the speakers to be from, what the native languages of the speakers were, and if they believed themselves to be familiar with the non-native accent heard. Participants then read through a short debrief.

**Data Analysis**

Participants’ data were taken from *Gorilla.sc* and responses to questionnaires were examined to ensure participants met the study’s inclusion requirements. Next, comprehension questions (used as an attention check) were assessed to make sure participants reached the 70% accuracy threshold for inclusion. The main analysis focused on RTs from the self-paced listening task. The RT data file was run through a script on *RStudio* to remove non-target word-clusters from the data, as only target-clusters were considered in analysis, as well as individual RTs 2.5SD above or below the participant’s mean per speaker and error condition (*trimr,* Grange, 2015).

Means were computed per condition, which was then inserted into *SPSS* (Version 26.0) to perform the main analysis comparing main effects of ‘speaker’ (native and non-native) and ‘error condition’ (no-error, grammatical error, word-choice error), as well as the interaction between these variables, on RT.

It should be noted here that whilst a 2x3 Repeated Measures ANOVA was used, two conditions failed Shapiro-Wilk normality checks: non-native control targets (*p* = .045) and non-native word-choice targets (*p* = .037). However, after inspection of Q-Q plots, the ANOVA was still considered to be viable; the other conditions furthermore met the normality checks. In addition, we log-transformed RTs to improve normality of distribution. ANOVAs using these log-transformed RTs showed the same outcomes as the results reported below with untransformed RTs.

After the main analysis (and pairwise comparisons between error conditions) was run, a post-hoc analysis was conducted to examine specifics of the interaction found in the main 2x3 ANOVA and to compare the two error types. For each participant we computed their ‘error cost’ for grammar errors as the RT difference between the grammar and no-error condition, and for word-choice errors as the RT difference between word-choice and no-error conditions. This was done for the native and non-native conditions separately. These error costs were then analysed using a 2x2 Repeated Measures ANOVA including the factors error type (grammatical or word choice) and speaker (native and non-native). A paired t-test was also run between the RTs from the native and non-native no-error conditions to assess whether there were any processing differences between native and non-native speakers in sentences without errors.

Finally, we examined whether any effects related to word-choice errors were related to contextually inappropriate use or rather to differences in frequency between word-choice targets and no-error/grammatical targets. We therefore ran another Repeated Measures ANOVA including Speaker and Word-Choice Condition (Word-Choice Error and Word-Choice Control). In this analysis, we examined responses to the same target words, presented either in the wrong context (word-choice error) or in the appropriate context (word-choice control).

 In addition to the ANOVAs we also conducted Bayesian ANOVAs (not pre-registered) to formulate how much evidence there was for or against the crucial interactions between error and speaker (the effect of interest in this study). We therefore compared a model with the main effects + the interaction of interest to a model with the main effects but no interaction (using JASP version 13.1, with 1 million iterations). We will report these results in the form “BF10”, showing the evidence for the alternative hypothesis (significant interaction) over the null hypothesis (no interaction). Values below 1 indicate evidence against an interaction; values above 1 indicate evidence for an interaction.

**Results**

**Sentence Comprehension and Post-Study Questions**

In the comprehension questions used to measure attention levels, all participants superseded the required 70% accuracy levels, showing that participants paid sufficient attention to the content of the sentences (*Maccuracy* = 96%, *SD* = 4.7, Range = 85% - 100%)

When answering the post-study questions, none of the 30 participants correctly guessed the accent or place of birth of the non-native speaker. Despite this, 7 participants still indicated that they believed themselves to be ‘familiar’ with the accent (another 13 indicated they were not familiar, and the remaining 10 chose to say they were ‘unsure’). It is worth noting that 7 participants guessed that the non-native speaker had a ‘Mediterranean’ accent, which, whilst perceptually close to a Chilean Spanish accent, is not strictly correct. However, there was no relationship between these close guesses, and confidence in ‘familiarity’, suggesting that even the participants who guessed a Mediterranean accent were not certain of that guess.

**Main analysis**

A 2x3 Repeated Measures ANOVA was carried out on participants' mean RTs, comparing the main effects of ‘speaker’ and error condition’, as well as the interaction effect between these two variables, on RT (see Figure 2 and Table 2).

<Insert Figure 2 around here>

<Insert Table 2 around here>

The ANOVA revealed that the type of speaker had a significant effect on RT (*F*(1,29) = 8.73, *p =* .006, ηp2= .231). RTs when listening to the non-native speaker (*M* = 1152.70, *SD* = 201.38) were significantly longer than when listening to the native speaker (*M* = 1121.45, *SD* = 190.10).

The main effect of error condition on RT was also significant (*F*(2,58) = 47.17, *p <* .001, ηp2  = .619). To further investigate this, pairwise comparisons (Bonferroni correction) were run. RTs associated with the no-error condition (*M* = 1071.83, *SD* = 202.22) were significantly shorter than both the grammar (*p <* .001) and word-choice errors (*p <* .001). However, there was no difference in RT between grammar (*M* = 1169.74, *SD* = 196.52) and word-choice errors (*M* = 1169.02, *SD* = 192.16) themselves (*p* > 0.999). This outlines that processing was significantly slower for both the grammar and word-choice error conditions than the no-error condition.

There was a significant interaction between speaker and error condition (*F*(2,58) = 12.71, *p <* .001, ηp2= .305). Bayesian analyses strongly supported this interaction (BF10 =130.3, error = 0.5%), suggesting the data were 130 times more likely to be observed in a model with an interaction between speaker and error than in a model without the interaction. As can be seen in Figure 2, responses to the native and non-native speaker were comparable in the no-error condition (Native speaker: *M* = 1083.23, *SD* = 203.43; Non-native speaker: *M* = 1060.11, *SD* = 209.02). This was confirmed in a paired t-test, revealing that RTs were not significantly different from one another, (*t*(29)= 1.68, *p* = .104) when processing non-native and native speech when sentences were well-formed. However, RTs when listening to the native speaker seemed to be shorter than when listening to the non-native speaker in both the grammatical (Native speaker: *M* = 1143.57, SD = 191.58; Non-native speaker: *M* = 1196.37, *SD* = 209.85) and in the word-choice error conditions (Native speaker: *M* = 1137.01, *SD* = 189.42; Non-native speaker: *M* = 1200.02, *SD* = 205.45). Therefore, this interaction between speaker and error condition suggests that error processing requires more time than processing of no-error sentences, but especially when these errors are produced by non-native speakers. To further examine this, and any potential differences between grammatical and word-choice errors, a Post-Hoc ANOVA was carried out on the error costs.

**Cost Analysis**

After computing error costs relative to no-error sentences for each participant and speaker for the grammatical error condition and word-choice condition, a 2x2 Repeated Measures ANOVA on RT cost was carried out.

There was a significant main effect of speaker on RT (*F*(1,29) = 32.79, *p <* .001, ηp2 = .531). This confirmed that the RT cost associated with error processing was significantly larger when listening to the non-native speaker (*M* = 138.09, *SD* = 85.80) than the native (*M* = 57.05, *SD* = 76.78) speaker; listeners needed more time when hearing errors made by the non-native over the native speaker.

However, the ANOVA also revealed that there was no significant main effect of type of error (*F*(1,29) = 0.20, *p* = .890, ηp2 = .001), meaning that the RT cost was similar for grammatical (*M* = 98.30, *SD* = 86.45) and word-choice conditions (*M* = 96.84, *SD* = 95.50). Furthermore, the interaction between the variables of speaker and error type was not significant (*F*(1,29) = 0.24, *p* = .626, ηp2 = .008), meaning that the effect of speaker on error processing was similar for both types of errors. The difference between native and non-native speakers was similar for the grammatical (Native: *M* = 60.33, *SD* = 70.78; Non-native *M* = 136.26, *SD* = 84.92) and for the word-choice condition (Native: *M* = 53.77, *SD* = 83.44; Non-native: *M* = 139.91, *SD* = 88.08, see Figure 2). This was confirmed by the Bayesian analysis, which showed evidence against the alternative hypothesis including an interaction between speaker and error type (BF10 *=* 0.28, error = 1.3%).

**Word-choice Control Analysis**

Finally, another 2x2 Repeated Measures ANOVA was run to check the validity of the word-choice error condition by comparing the word-choice error and word-choice control conditions (i.e., two conditions using the same target words but in either correct or in “incorrect” contexts). As the main ANOVA shows that processing word-choice errors was significantly slower than well-formed speech, this test was run to ensure contextual error detection was not confounded by frequency differences between stimuli used in the word-choice condition and in the no-error/grammatical error conditions (Brysbaert et al., 2011).

<Insert Figure 3 around here>

Once again, a significant main effect of the speaker was found (*F*(1,29) = 9.63, *p =* .004, ηp2 = .249), meaning that RTs were significantly longer when listening to the non-native speaker (*M* = 1176.48, *SD* = 197.37) over the native speaker (*M* = 1140.59, *SD* = 192.97). The main effect of word-choice condition was also significant (*F*(1,29) = 4.59, *p =* .041, ηp2  = .137), with word-choice error RTs (*M* = 1169.02, *SD* = 192.16) being significantly longer than word-choice control RTs (*M* = 1148.07, *SD* = 196.53). This implies that the longer RTs associated with the word-choice error condition in the main and cost-analysis AVOVAs are in fact related to the use of words in an incorrect context, rather than the frequency of words themselves.

The interaction between these two variables (speaker and word-choice errors versus word-choice controls) was not significant (*F*(1,29) = 3.63, *p =* .067, ηp2  = .111). The Bayesian analyses showed some, but weak, evidence in favour of an interaction between speaker and condition (BF10 = 2.31, error = 0.75%). As can be seen from Figure 3, numerically the effect of speaker was driven by the word-choice error condition, as the non-native trials have longer RTs (*M* = 1200.02, *SD* = 205.45) than native trials (*M* = 1137.01, *SD* = 189.42). However, this magnitude of difference is not seen between non-native (*M* = 1151.90, *SD* = 202.51) and native (*M* = 1143.86, *SD* = 206.32) word-choice controls. This suggests that speaker effects on word-choice error conditions are related to incorrect contextual use, instead of frequency effects; but this should be interpreted cautiously, as the interaction does not reach significance in the frequentist analysis and is only supported weakly by the Bayesian analysis.

**Discussion**

This experiment investigated behavioural processing of linguistic errors across native and non-native accents. The study aimed to examine whether error processing varies when listening to native and non-native speakers, as well as to compare processing across error type, utilising frequently studied (grammatical) and novel (contextual) errors. The results showed that RTs when listening to the non-native speaker were longer than when listening to the native speaker. This was due to the differences in RT when listening to the speakers commit grammatical/word-choice errors, while no speaker difference was found during no-error speech. As RT was used as a proxy for linguistic processing, this suggests that more processing is needed when listening to a non-native speaker commit linguistic errors, over a native speaker. Moreover, there was no difference in RT across grammatical or contextual errors, suggesting that error processing was more demanding when listening to non-native speakers during both grammatical and word-choice errors.

**Error Processing Across Accents**

The overall finding that errors are processed differently when listening to native or non-native speech is in line with many previous studies, including those showing differences in how people respond to implausible sentences with errors (e.g., Gibson et al., 2017), how people rate under-informative sentences (e.g, Fairchild & Papafragou, 2018), as well as EEG studies showing different ERP effects in response to errors made in native versus non-native speech (e.g., Grey & van Hell, 2017, Hanulíková et al., 2012).

Listeners might follow a “good enough approach” (e.g., Ferreira et al., 2002), in which linguistic input might not be processed fully. It has been proposed that this might apply even more strongly when listening to non-native speech, where listeners might expect errors and might therefore process the linguistic input less precisely or fully. They might be more tolerant to errors and instead rely on context to facilitate their understanding (Lev-Ari, 2015a,b). Indeed, several ERP results suggest that errors committed by non-native speakers may be ‘overlooked’ in processing (Hanulíková et al., 2012), perhaps especially so when errors are expected (Caffarra & Martin, 2019).

However, the current study shows that responses are slower when listening to errors spoken in a non-native accent than when listening to these errors in native speech. This suggests that listeners are not more tolerant but rather that these errors produced by non-native speakers are even more difficult to process. This suggests that more cognitive recruitment is needed (e.g., Gagné, Besser, & Lemke, 2017). Listening to non-native speech might pose additional demands (e.g., processing slight mispronunciations, different intonation patterns) and this has been found to make speech comprehension more difficult (e.g., Braun et al., 2011, Floccia et al., 2009; Van Engen & Peelle, 2014). In line with the intelligibility account proposed by Fairchild and Papafragou (2018), as a consequence, listening to non-native speech might reduce cognitive resources available for error processing, resulting in longer response times. While there was no difference between listening to native and non-native control sentences, listeners might apply additional resources to quickly process non-native speech (without errors), which reduces the resources needed to quickly manage errors when they are encountered. This interpretation would be in line with dual-tasking effects that have shown, for example, that listeners show larger costs on a secondary task when speech is presented in (more) noise (cf. Gagné et al., 2017). Similar to speech in noise, non-native speech could require additional cognitive resources that make the “secondary task” (processing errors) more difficult.

It could be precisely because of the absence of early repair of errors in non-native speech (e.g., Caffarra & Martin, 2019, Hanulíková et al., 2012) that later stages of processing are more demanding (e.g., Gosselin et al., 2021, Romero-Rivas et al., 2015). When looking at behavioural processing (as done here), this can lead to longer processing times to ensure comprehension is still achieved. It is possible that this cost of processing errors made by non-native speakers could be especially effortful when the listener is not familiar with the accent, as was the case in this study, or when grammatical errors are not entirely predictable (e.g., when a range of errors is used, as in the current study). This is in line with Grey and Van Hell’s (2017) suggestion that listeners unfamiliar with the non-native accent might not have sufficient cognitive resources available to quickly re-analyse the sentence. When processing a familiar accent, the processing demands might be reduced. Similarly, these demands might be smaller when using written instead of spoken language (e.g., Fairchild & Papafragou, 2018), when stimuli are related and provide context (e.g., Lev-Ari, 2015a), or when listeners are given more time to respond (e.g., in studies using offline ratings, e.g., Gibson et al., 2017). A “good enough” approach might be more successfully applied by listeners in those types of situations. In contrast, when overall demands are higher (i.e., when having to respond quickly to an unfamiliar speaker without any other contextual factors facilitating understanding, as was the case in this study), this approach might be less successful and non-native speech might increase processing costs associated with errors.

**Processing of Different Error Types**

Reaction times in the current study did not vary as a result of error type (grammatical or word choice). This suggests that contextual word-choice errors are processed in a similar fashion to grammatical errors. Previous studies assessing “word-choice” errors have focused on semantic violations that are incompatible with our world knowledge (e.g., sentences referring to flying on a cactus). These implausible sentences might be evaluated differently in native and non-native speech (e.g., Gibson et al., 2017). However, ERP studies focusing on online processing provide more mixed results. Some conclude that semantic violations, contrary to grammatical errors, might elicit similar effects in native and non-native speech (e.g., Hanulíková et al., 2012) while others have shown different patterns for native and non-native speech (e.g., Romero-Rivas et al., 2015). Crucially, however, these errors might be so salient, and so uncommon in daily-life speech, that they stand out and need to be re-analysed in both native and non-native speech. Our study suggests that listeners are not just sensitive to these salient violations of world knowledge, but also to more subtle word choices that mismatch the context, in particular when these violations are presented in a non-native accent. These are errors that occur in non-native speech but that typically receive less attention in classroom settings (especially when compared to e.g., grammatical errors). Our study, however, shows that these errors can influence processing speed as much as grammatical errors.

 Our word-choice errors, however, differed from the no-error controls and grammatical errors in frequency (and to a lesser extent in length). Comparisons with “word-choice” control conditions in which the same “word-choice” target words were presented in their corresponding context (e.g., formal words in a formal context) suggested that effects of word-choice errors were indeed due to violation of context and not frequency of usage. Nevertheless, considering that this interaction did not reach significance and only showed weak support from Bayesian analyses, future research is needed to examine the role of context violation as opposed to frequency effects.

**Conclusion**

In conclusion, this study suggests linguistic error processing varies as a function of accent, both when listeners encounter unexpected word choices as well as when they hear grammatical errors. Our results suggest that more processing time is needed in order to handle linguistic errors when spoken in a non-native accent. Additionally, it may be that effects are comparable across different error categories, or at least the two examples (grammatical and contextual word-choice violations) used here. When listening to non-native speech, listeners might recruit additional resources and might consequently not have sufficient resources left to quickly process and repair errors. This has consequences for our everyday conversations and our perceptions of non-native speech. For example, non-native speech can be judged as being less credible, and this could be a consequence of increased processing difficulty (e.g, Lev-Ari & Keysar, 2010). Slower processing when encountering non-native errors might thus not only influence the ease with which but also *how* we engage with millions of non-native speakers.

**References**

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language:

Listener perception versus linguistic scrutiny. *Language Learning*, *59*(2), 249-306.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research methods, 52*(1), 388-407.

Barner-Rasmussen, W., & Aarnio, C. (2011). Shifting the faultlines of language: A quantitative functional-level exploration of language use in MNC subsidiaries. *Journal of World Business, 46*(3), 288-295.

Biesenbach-Lucas, S. (2007). Students writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning & Technology*, *11*(2), 59-81.

Bloch, B. (1995). Career enhancement through foreign language skills. *International Journal*

 *of Career Management*. *7*(6), 15-26.

Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer

 program]. Version 6.1.38, retrieved 25 January 2021 from <http://www.praat.org/>

Braun, B., Dainora, A., & Ernestus, M. (2011). An unfamiliar intonation contour slows down online

speech comprehension. *Language and Cognitive Processes*, *26*(3), 350-375.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*. *58*(5), 412-424.

Caffarra, S., & Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error

 typicality in foreign accented speech perception. *Cortex, 116*, 308-320.

Chun, A. E., Day, R. R., Chenoweth, N. A., & Luppescu, S. (1982). Errors, interaction, and correction: A study of native‐normative conversations. *Tesol Quarterly, 16*(4), 537-547.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing?. *Trends in Cognitive Sciences*, *10*(12), 564-570.

Corder, S. P. (1971). Idiosyncratic dialects and error analysis. *International Review of Applied Linguistics in Language and Teaching*, *9*(2) 147-160.

Danesi, M. (2008). Conceptual errors in second-language learning. *Applications of Cognitive Linguistics, 9*, 231.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65-70.

Dewaele, J. M., & Pavlenko, A. (2003). *Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects* (pp. 120-141). Multilingual Matters.

Economidou-Kogetsidis, M. (2011). “Please answer me as soon as possible”: Pragmatic failure in non-native speakers’e-mail requests to faculty. *Journal of Pragmatics*, *43*(13), 3193-3215.

Fairchild, S., & Papafragou, A. (2018). Sins of omission are more likely to be forgiven in non-native speakers. *Cognition*, *181*, 80-92.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\* Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11-15.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt?. *Journal of Psycholinguistic Research*, *38*(4), 379-412.

Frumuselu, A. D., De Maeyer, S., Donche, V., & Plana, M. D. M. G. C. (2015). Television series inside the EFL classroom: Bridging the gap between teaching and learning informal language through subtitles. *Linguistics and Education*, *32*, 107-117.

Fuller, J. M. (2003). Discourse marker use across speech contexts: A comparison of native and non-native speaker performance. *Multilingua, 22*, 185-208.

 Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using

 a dual-task paradigm: A review. *Trends in hearing, 21*, 1-25.

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017).

Don’t underestimate the benefits of being misunderstood. *Psychological Science*, *28*(6), 703-712.

 Gilquin, G., & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English*

*Text Construction*, *1*(1), 41-61.

Gosselin, L., Martin, C. D., Navarra-Barindelli, E., & Caffarra, S. (2021). The presence of a foreign

 accent introduces lexical integration difficulties during late semantic processing. *Language,*

 *Cognition and Neuroscience*, 1-21.

Grange, J.A. (2015). trimr: An implementation of common response time trimming methods.

 R package version 1.0.1. <https://cran.r-project.org/web/packages/trimr/index.html>

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics, 42*, 93-108.

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person’s mistake is another’s standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*(4), 878-887.

Krashen, S. D. (1976). Formal and informal linguistic environments in language acquisition and language learning. *Tesol Quarterly*, *10*(2), 157-168.

Lev-Ari, S. (2015a). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Frontiers in Psychology, 5*, 1546.

Lev-Ari, S. (2015b). Adjusting the manner of language processing to the social context: Attention allocation during interactions with non-native speakers. *Attention and vision in language processing*, 185-195.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093-1096.

Pelzl, E., Lau, E. F., Jackson, S. R., Guo, T., & Gor, K. (2021). Behavioral and neural responses to tone errors in foreign-accented Mandarin. *Language Learning*, *71*(2), 414-452.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-

accented speech. *Frontiers in Human Neuroscience*, *9*, 167.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates

 linguistic anticipatory processes. *Neuropsychologia, 85*, 245-255.

Schepens, J., van Hout, R., & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition, 194*, 1-14.

Van Berkum, J. J. (2009). The neuropragmatics of ’simple’ utterance comprehension: An ERP review. In U. Sauerland, & K.Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (276-316): Basingstoke: Palgrave Macmillan.

Van Berkum, J. J., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event‐related brain potentials reflect discourse‐referential ambiguity in spoken language comprehension. *Psychophysiology, 40*(2), 235-248.

 Van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences

 and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657-

 671.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580-591.

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human*

*Neuroscience*, *8*, 577.

Van Hout, R. (2003). *The lexicon-syntax interface in second language I* (Vol. 30).

 Amsterdam, Netherlands: John Benjamins Publishing.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology, 57*(1), 129-163.

**Appendices**

**Appendix A**

Examples of the sentences used in the present study. Table A1 shows the sentences used in the friend context, and Table A2 shows the teacher context. The subheadings show which sentence corresponds to each condition. The word(s) in bold indicate the target. The forward slashes in the first column of ‘No Error’ sentences indicate where the audio was cut for each of the clusters. These parameters correspond to each of the other conditions (grammatical and word-choice).

<Insert Table A1 about here>

<Insert Table A2 about here>

**Appendix B**

An example of the comprehension questions used in the study. Table B1 shows the questions used in the friend context, and Table B2 shows the teacher context. In each case the correct answer to the question is answer A.

<Insert Table B1 about here>

<Insert Table B2 about here>

**Appendix C**

Example images used in the experiment to increase the salience of context. Figure C1 shows the image used for the friend context, whilst Figure C2 shows the image used for the teacher context.

<Insert Figure C1 about here>

<Insert Figure C2 about here>

1. We will refer to violations of patterns expected in native speech as “errors” but this is not intended as a judgement of non-native speech being inferior to native speech. [↑](#footnote-ref-1)