RESEARCH ARTICLE

Co-text, context, and listening proficiency as crucial variables in intelligibility among nonnative users of English

Veronika Thir 🆔

University of Passau, English Language and Culture, Passau, Germany
Email: Veronika.Thir@uni-passau.de

Abstract

Research on intelligibility in international encounters has long focused on issues of pronunciation to the detriment of factors such as linguistic co-text and extralinguistic context, which are comparatively well-studied variables in intelligibility research concerning L1 listeners. This paper seeks to expand the scope of international intelligibility research in this respect by reporting on a large-scale study involving 423 nonnative listeners at different proficiency levels, who transcribed words spoken with another nonnative accent under four conditions that varied in the availability of syntactic, semantic, and schematic cues. The results suggest that co-text and context as well as listening proficiency are crucial variables that ought to receive greater attention in research on international intelligibility. The pedagogical implications of these findings are addressed as well.

Introduction

Currently, most nonnative users of English primarily use the language as a means of international communication. As the vast majority speaks English with some degree of nonnative accent, research on international intelligibility has flourished during the past 2 decades. Pedagogical suggestions from this field have mostly centered on aspects of sound production and perception—that is, on equipping learners with a core of pronunciation features necessary for successful international communication and improving their receptive and productive phonological accommodation skills (e.g., Deterding, 2013; Jenkins, 2000, 2002; Walker, 2010)—for example, through increased exposure to nonnative accents. These suggestions arose from a considerable body of research that until now has strongly focused on the role of particular pronunciation features for mutual understanding in international contexts (e.g., Deterding, 2013; Gardiner, 2019; Jenkins, 2000, 2002; Kang, Thomson, et al., 2020). That is, the issue of international intelligibility has been approached primarily from a bottom-up perspective on the listening process, the focus of attention being the quality of the acoustic signal. What has received comparatively little attention in research on international

intelligibility so far is the role of top-down influences in the form of linguistic co-text and extralinguistic context. *Co-text* refers to the "verbal environment" of an utterance—that is, "the accompanying text" (Halliday, 1999, p. 3)—and relates to the linguistic relations encoded in a particular language system (Widdowson, 2011)—for example, in the form of collocations and lexical or morphosyntactic relations. *Context* refers to "the extralinguistic circumstances in which language is produced" (Widdowson, 2011, p. 221) and, besides the immediate situational context, also involves the sociocultural context that language users bring with them (Widdowson, 2011). The importance of co-textual and contextual information for linguistic comprehension is generally acknowledged by researchers: It is well established that both L1 and L2 listening are cognitively interactive processes in which listeners combine information from the acoustic signal with information from linguistic co-text and extralinguistic context (e.g., Field 2004, 2008; Goh & Vandergrift, 2021; see also Kennedy, 2021). This has repeatedly been demonstrated regarding intelligibility to L1 listeners in publications on speech science, psycholinguistics, or speech pathology (e.g., Bent et al., 2019; Garcia & Cannito, 1996; Kamide et al., 2003) and regarding L2 listening comprehension more generally (e.g., Macaro et al., 2005). However, research on co-textual and contextual effects has been comparatively scarce concerning intelligibility to L2 listeners, notably L2 listeners in lingua franca (LF) listening situations—that is, when listening to another nonnative accent. Yet, it is precisely in such contexts that an increased need to rely on co-textual and contextual information to compensate for unexpected pronunciation patterns can be expected. The present paper seeks to fill this research gap by expanding the scope of intelligibility research on co-textual and contextual factors to English as LF (ELF) listeners. Additionally, it adopts a differentiated perspective regarding nonnative ELF listeners' ability to rely on co-text and context, acknowledging that the latter might be affected by listeners' proficiency level. Based on a large-scale experimental study, it shows that nonnative listeners at different levels profit from various types of co-textual and contextual cues when recognizing words spoken with another nonnative accent. Its findings highlight the need for a greater recognition of co(n)textual factors and listening proficiency as relevant variables in research on international intelligibility, both of which have so far received relatively little attention in the field (see also Kang, Moran, et al., 2020).

Intelligibility as spoken word recognition: A cognitively interactive process

Intelligibility is a complex concept that has been defined in various ways (see also Munro & Derwing, 2015). The present paper is primarily interested in the level of spoken word recognition (SWR, cf. Smith's *intelligibility* 1992)—that is, the process of matching units in the stream of speech to a particular item stored in the mental lexicon (e.g., Magnuson, 2017). Crucially, SWR does not merely involve information from the acoustic level, which would then be processed as words, sentences, and so on (*bottom-up* processing, see Field, 2004) but is essentially interactive (e.g., Mirman, 2017) because listeners combine bottom-up information with *top-down* information—that is, higher level linguistic and extralinguistic information (or, in other words, co-textual and contextual information), which then affects the processing of smaller units such as speech sounds (see Field, 2004).

The use of top-down information can take various shapes in SWR. Studies have shown that listeners draw on lexical knowledge—for example word frequency knowledge or knowledge of the phonological "neighborhood" of a word (Bradlow & Pisoni,

1999; Luce & Pisoni, 1998)—and on aspects of the visual context, such as gestures (Garcia & Cannito, 1996) or referent availability (Tanenhaus et al., 1995), to identify spoken words. Additionally, numerous studies have provided evidence for the use of sentence co-text in SWR, with syntactic, semantic, and collocational constraints aiding listeners in recognizing words (Baese-Berk et al., 2021; Behrman & Akhund, 2013; Bent et al., 2019; Hilpert, 2008; Kennedy & Trofimovich, 2008). Sentence co-text has also been found to facilitate SWR by triggering certain schematic expectations that make listeners anticipate the intended word (Kamide et al., 2003). Finally, Garcia and Cannito (1996) and Hustad and Beukelman (2001) used verbal cues (e.g., "relocating to a new city") to prime listeners schematically for certain situations or scripts, which contributed positively to SWR.

Spoken word recognition in nonnative listeners

The research discussed above was concerned with co(n)textual effects on SWR in L1 listeners, often under adverse listening conditions such as when listening to dysarthric speech (e.g., Garcia & Cannito, 1996) or L2-accented speech (Baese-Berk et al., 2021; Behrman & Akhund, 2013; Bent et al. 2019; Kennedy & Trofimovich, 2008). Such research concerning L2 listeners is much harder to find, although other factors associated with intelligibility to nonnative listeners are well studied, such as the effect of various talker–listener pairings. Thus, various studies examined whether L2 listeners experience a benefit when listening to nonnative rather than native speech, either in their own accent or a different L2 accent (Bent & Bradlow, 2003), but evidence regarding this phenomenon is mixed (cf. Stibbard & Lee, 2006), and research suggests that its manifestation depends (at least partly) on listening proficiency and the acoustic phonetic similarity of the talker's and listener's accents (e.g., Pinet et al., 2011; Stringer & Iverson, 2019).

Research on co(n)textual effects on intelligibility to L2 listeners, however, is comparatively scarce, although of particular interest as results for L1 listeners in this respect may not apply to (all types of) L2 listeners. This is because nonnative listeners may be less apt at combining bottom-up and top-down information, the interactive processing of the speech signal here being less "automatic" because their linguistic knowledge is typically narrower (Goh & Vandergrift, 2021, p. 20) and not as deeply entrenched (Mauranen, 2018). It has thus been suggested that L2 listeners may be more dependent on acoustic information than native listeners (e.g., Cutler, 2012; Jenkins, 2000). Crucially, this is not to say that nonnative listeners solely rely on bottom-up information in SWR. Bradlow and Pisoni (1999) and Yoneyama and Munson (2010) found evidence that nonnative listeners draw on top-down lexical information (word frequency knowledge and knowledge of the phonological neighborhood of a word) when identifying spoken words. There is also evidence that nonnative listeners profit from semantic and schematic relations in the sentential co-text of a word in SWR (Lagrou et al., 2013; Mack, 1992; see also Bradlow and Alexander, 2007, for a "clear" speaking style but not a "plain" speaking style).

Thus, nonnative listeners also seem to rely on top-down information in SWR, though not necessarily in the same way or to the same extent as native listeners, which may sometimes indeed make them more dependent on acoustic information. Still, top-down influences may play a considerable role in L2 listening, especially when they are extralinguistic. Interestingly, Wolff (1987) argues that nonnative listeners may rely more heavily on *top-down* processing than native listeners due to problems with

bottom-up processing owing to a lack of L2 knowledge (see also Field, 2004). He assumes that once a certain threshold has been attained, nonnative listeners will resort to bottom-up processing more intensively (1987, p. 313), implying that top-down processing is particularly pronounced in lower proficiency listeners. Evidence for this is provided by Koster (1987), who found that lexical cues improved SWR the most for the L2 listeners with the lowest proficiency in her study, compared with high-proficiency L2 listeners and L1 listeners. Similarly, Field (2004) found evidence for the dominance of top-down knowledge over bottom-up information in lower intermediate listeners' SWR, albeit only when the available sentence co-text was "highly constraining" (2004, p. 373). These findings stand in stark contrast to the established assumption of "bottom-up dependency" in low-proficiency L2 listeners (Field, 2004, p. 364), with their (necessarily greater) attention to the linguistic code detracting them from exploiting co-text and context effectively (Field, 2008, p. 132), for which there is also some empirical support (e.g., Gu et al., 2005).

Given these somewhat inconclusive findings regarding top-down processing in low-proficiency L2 listeners, it has been suggested that their reliance on top-down information simply has a different function than in accomplished listeners—namely, a compensatory function—to supplement incomplete understanding (Field, 2008). As discussed above, native listeners rely equally on co-text and context when confronted with difficult input, so language users seem to naturally adopt this approach when encountering an incomplete or ambiguous acoustic signal. Crucially, this becomes an essential skill for both native and nonnative listeners in international contexts, where they typically encounter numerous unexpected pronunciation patterns.

International intelligibility and the co-text/context problem

From the above discussion, one might expect pronunciation to play a subordinate role for intelligibility in international communication, as nonnative listeners, particularly after having attained a certain proficiency level, should be able to rely on cues from the linguistic co-text and the extralinguistic context in SWR of another nonnative accent. In contrast, Jenkins (2000, 2002) observed that in international encounters, nonnative listeners tended to have difficulty drawing on such cues when trying to understand their nonnative interlocutor, with pronunciation consequently being the number one cause of miscommunication in her data. Interestingly, the listeners she observed were by no means low-proficiency listeners but at B2–C1 level according to the *Common European Framework of Reference* (CEFR; see Jenkins, 2002). Jenkins' findings, which were derived from interactive and multimodal speaking tasks, led her to conclude that nonnative listeners "below the level of bilingual proficiency appear unable to process contextual cues to compensate for their interlocutors' pronunciation errors" (2002, p. 87) and thus tend to overrely on the acoustic input. Interestingly, this stands in stark contrast to more general accounts of ELF communication that stress its situatedness in a particular context of communication and international users' ensuing attention to aspects beyond the linguistic code (e.g., Seidlhofer, 2011). Nevertheless, few studies have hitherto investigated the role of co-textual and contextual information for international intelligibility. Osimk (2009) found that nonnative ELF listeners profit from sentential co-text in SWR, and Thir (2021) showed how nonnative ELF listeners at a proficiency level similar to those in Jenkins (2000, 2002) draw on various types of linguistic and extralinguistic cues in SWR during interactive and multimodal speaking

tasks. Interestingly, in contrast to Jenkins' participants, who tended to "adjust[] the context and/or co-text to bring them into line with the acoustic information rather than vice versa" (2002, p. 90), one listener in Thir (2021) seemed to do the opposite, ignoring acoustic information probably in favor of an interpretation congruent with the available visual context. In addition, several studies (Kaur, 2011; Mauranen, 2006; Pitzl, 2010) did not find pronunciation problems to be a major source of miscommunication in international interactions (however, cf. Deterding, 2013), suggesting that many nonnative ELF listeners compensate for phonetic-phonological variation by relying on other sources of information besides the acoustic signal.

Clearly, co-text and context have the potential to play a considerable role for intelligibility among nonnative ELF listeners, but empirical findings in this respect are scarce and inconclusive. As previous research suggests that listening proficiency is a crucial variable in the use of top-down processing in nonnative listeners, a more nuanced approach that differentiates more finely between listeners at different proficiency levels seems in order. Obviously, nonnative listeners may differ vastly in their knowledge of the L2 and thus in the availability of linguistic knowledge as a top-down resource in L2 speech perception as well as their ability to effortlessly switch between top-down and bottom-up processing. This might explain the finding that listening proficiency affects listeners' sensitivity to L2 accented speech (Kang, Moran, et al., 2020), suggesting that listeners at different levels differ in their ability to compensate for phonological ambiguity with the help of co(n)textual cues. The present study thus examines the following research questions:

RQ 1: Do nonnative listeners at different proficiency levels (intermediate to low advanced) profit from different types of co-textual and contextual cues when identifying words spoken with another nonnative accent?

RQ 2: If yes, do they do so to a lesser extent than highly advanced nonnative listeners (i.e., those at C2 level)?

As co-text and context have received relatively little attention so far in research on international intelligibility and seem to be somewhat underestimated factors, a further objective of this study is to compare their effects to that of a variable more frequently discussed in the literature: speech familiarity (e.g., Deterding, 2013; Gass & Varonis, 1984; Kennedy & Trofimovich, 2008; Smit, 2010). RQ 3 thus asks the following:

RQ 3: What is the relative influence of co-textual and contextual information on intelligibility to nonnative listeners at different proficiency levels compared to that of speech familiarity?

The types of co(n)textual cues investigated in the present study include syntactic, semantic, and schematic cues. Whereas syntactic and semantic information are here viewed as relating to linguistic co-text, schematic information is regarded as relating to language users' cognitive sociocultural context (Widdowson, 2011). This is because schemata are knowledge structures shaped by previous and influencing new experiences of the world (Gureckis & Goldstone, 2011, p. 725). However, as mentioned above, schematic expectations may sometimes be evoked by certain textual elements (e.g., sentence co-text, topic cues). That is, linguistic *co-text* may be used to trigger a certain cognitive *context* within a listener, which has also been done in the present study.

Method

To attain a sufficiently large and internationally diverse sample of nonnative listeners, an online listening experiment was developed and distributed via the internet. The experiment involved four conditions in which listeners had to orthographically transcribe target words. Three of these conditions supplied listeners with co(n)textual cues. The SYN condition examined the effect of syntactic information on intelligibility. Target words were embedded in short, semantically neutral sentences that merely indicated the part of speech (POS) of the target word (given in square brackets), e.g., "It's quite [flat]."

The SYN+SEM condition examined the effect of syntactic and semantic information on intelligibility. In addition to indicating the POS of the target word, carrier sentences included a semantic cue in the form of a prime word, which was semantically related to the target word, either in terms of a classic meaning relationship (meronomy, hyponymy, or antonymy) or, if target and prime differed in POS, via semantic entailment (i.e., the denotational meaning of the prime entailed an aspect of the target word's meaning). For example, in "He was driving a [van]," *drive* entails [vehicle], which is a semantic property of the target word [van]. The target word was always preceded by the prime.

The SYN+SCH condition examined the effect of syntactic and schematic information and therefore included a schematic cue—that is, a short description of the situation in which the utterance occurred (in small caps in the following example). This cue appeared on screen before the carrier sentence was played. Carrier sentences indicated the POS of the target word but did not include any semantic primes as in the SYN+SEM condition—for example, "WHEN GETTING UP IN THE MORNING: I can't find my [pants]."

The fourth condition was a control (C) condition, where target words were presented without any co-text or context. All target words and words in the carrier sentences were A1–B2 levels according to the *English Vocabulary Profile* (EVP; Cambridge University Press, 2012). Table 1 summarizes all target words and sentences. In each condition, the six[1] target items were intermixed with nine distractor items. Most target words were monosyllabic to increase task difficulty and avoid ceiling effects, but each condition also included two disyllabic target words, as another part of the research project examined word-length effects on international intelligibility.

Target words included either the NURSE or the TRAP vowel (distributed evenly across conditions and mono- and disyllabic words), as another part of the research project compared the international intelligibility of the speaker's realization of these two phonemes (see Thir, 2020). For the monosyllabic TRAP words,[2] each condition also contained two TRAP-DRESS minimal pair (MP) words (e.g., *sand*, *land* in the C condition), as the speaker's realization of TRAP as [e] neutralized this distinction, with the effect of MP status of a word thus being of interest. For NURSE, no MP words were included, as the speaker's realization ([øə]) did not approximate another English

---

[1]As participants were not compensated due to a lack of financial resources, the number of target words per condition was kept at six, although clearly a higher number of target words would have been desirable. However, this would have also necessitated more distractor words, prolonging the online experiment to such an extent that the participant dropout rate would have been unacceptably high.

[2]Only words that contained TRAP in both General American and Received Pronunciation according to the *Longman Pronunciation Dictionary* (Wells, 2008) were included.

Table 1. Full list of target words and carrier sentences

C	SYN
[birth]	There's a [nurse].
[worker]	There's a [curtain].
[rat]	It's quite [flat].
[palace]	It's quite [massive].
[sand]	It's quite [bad].
[land]	It's a [gas].

SYN+SEM	SYN+SCH
They found the <u>feather</u> of a [bird].	WHEN BUYING A NEW BED This one is quite [firm].
His favourite <u>colour</u> is [purple].	ABOUT A RICH OLD MAN He treats his [servant] with respect.
He was <u>driving</u> a [van].	ON THEIR WAY HOME FROM THE PUB Let's get a [cab].
She liked the <u>book's</u> final [chapter].	WHEN HANGING A PICTURE ON THE WALL Could you give me the [hammer]?
He missed his <u>relatives</u>, especially his [dad].	AT THE AIRPORT I need to pick up my [bag].
He <u>fried</u> the vegetables in a [pan].	WHEN GETTING UP IN THE MORNING I can't find my [pants]!

Note. Target words are in square brackets. Semantic cues are underlined, and schematic cues are in bold and small caps. Test items were not presented in the order above to avoid phonological priming effects.

phoneme. Each condition included two types of TRAP-DRESS MP words: one where the words differed in POS (e.g., *bad–bed*) and one where they did not (e.g., *pan–pen*).[3]

Another factor considered in the selection of target words was word frequency, measured as the "contextual diversity" (CD) frequency count in the *SUBTLEX$_{US}$* corpus (Brysbaert & New, 2009; https://www.ugent.be/pp/experimentele-psychologie/en/research/documents/subtlexus/subtlexus3.zip). The eight monosyllabic non-MP words were comparable in terms of word frequency, as determined by phi coefficients calculated with a chi-square test.[4] The same applied to the eight disyllabic target words. However, the monosyllabic non-MP words were more frequent than the disyllabic words, reflecting that word length and word frequency are inversely related (Zipf, 1935/1965). Unfortunately, it was impossible to balance the MP words in terms of frequency, as priority was given to the other selection criteria (occurrence in both American and British Standard pronunciation, adequate level in the EVP), which were applied to *both* words in an MP to increase the likelihood that it would be perceived as MP by L2 listeners. Thus, only a small number of suitable MPs remained.

Carrier sentences in the SYN+SEM and SYN+SCH conditions were developed by the author and tested informally on and discussed with lay persons and colleagues, with a focus on their comprehensibility to L2 listeners and predictability in relation to the target words. The results suggested a satisfactory level of predictability (i.e., sentences pointed listeners to semantically or schematically appropriate words but did not render the additional audio stimulus superfluous). Carrier sentences were then tested (along

---

[3]Sometimes, words in an MP existed as multiple parts of speech. Classification was based on considerations of L2 listeners' likelihood to perceive the MP as involving words of the same or different POS. Thus, *sand-send* was classified as an MP "different POS," as only the noun *sand* is included in the EVP (B1 level), whereas the verb is not, which is moreover rather infrequent apart from certain specialized contexts.

[4]I am grateful to Dr. Andreas Baumann for suggesting this procedure and providing me with the *R* code for this calculation.

with the sentences and words in the C and the Syn conditions) in a pilot study involving 97 L2 listeners form various L1 backgrounds (who were not included in the sample of this study), which further confirmed their appropriateness for L2 listeners.

Participants

A male L1 Austrian German speaker aged 68 years read the stimuli words and sentences. He was recruited due to his typical Austrian accent and his decade-long experience of using English in international contexts. He had learned English formally for 3 years.

The 423 nonnative listeners in this study (male = 142, female = 279, other = 2) came from a data set originally collected for a larger project on international intelligibility involving both native and nonnative speakers of English. They had been recruited via email, social media, and the author's international contacts. Participants were aged 18–70 years, though most (79%) were 18–35 years old (mean age: 29.4 years). They came from numerous different L1 backgrounds, reflecting the diversity of international listeners around the world, though due to the sampling procedure, their L1s were mostly native to the European continent. A quarter (25%) were L1 speakers of Romance languages, 12% of Slavic languages, 11% of Germanic languages other than English, 7% of Finno-Ugric languages, and 4% of Greek. However, the sample also included L1 speakers of Turkish (12%); Chinese (6%); Thai, Arabic, or Iranian languages (4%, respectively); and Japanese (3%). There were no L1 Austrian German listeners in the study.

Participants indicated their listening proficiency in English using the self-assessment scale for listening in a foreign language of the CEFR (Council of Europe, 2018, p. 167; descriptors were slightly adapted for clarity). Each descriptor corresponds to one of the six proficiency levels of the CEFR, allowing straightforward comparison with studies such as Jenkins (2000, 2002). Participants' self-assessed listening proficiency levels are summarized in Table 2. About a third (36%) regarded themselves as highly advanced listeners (C2 level), more than a quarter (29%) thought they were low-advanced listeners (C1 level), and about a third believed to be at intermediate or upper-intermediate level (B1 or B2 level).

Procedure

The experiment was conducted using the questionnaire application *SoSciSurvey* (https://www.soscisurvey.de/). Participants were first asked to adjust the volume of

Table 2. Participants' self-assessed listening proficiency[5]

CEFR level	*n*	%
B1	51	12
B2	94	22
C1	124	29
C2	154	36
Total	*423*	

[5]Data originally collected from A1 and A2 listeners were removed from the sample due to the small *n* for these groups (8 and 11 participants, respectively).

their computers using a test audio spoken in a standard British accent.[6] The use of headphones and participation in calm surroundings was recommended. Participants were asked to provide consent and confirm that they had normal hearing and were not participating multiple times. They were then presented with the four conditions in randomized order. An autoplay function was used to prevent relistening to a test item. In all conditions, they had to type the target word in a designated gap on their screen. In the carrier sentences in the Syn, Syn+Sem, and Syn+Sch conditions (presented in writing[7] in addition to the auditory stimulus), the gap was in place of the target word. The schematic cue in the Syn+Sch condition appeared on screen 5 s prior to the carrier sentence and the onset of the audio.

To reduce guessing and avoid ceiling effects, participants received 11 s (C and Syn conditions) or 12 s (Syn+Sem and Syn+Sch conditions) to submit their answer from the onset of the audio. The time limit was slightly longer in the Syn+Sem and Syn+Sch conditions as here the carrier sentences were slightly longer than in the Syn condition. If participants finished typing before the time limit, they could proceed to the next trial using a button on their screens. Each condition started with a timed practice item containing neither of the two targeted vowels. Participants could control the start of the actual trials themselves. The intelligibility portion of the survey lasted for about 15 min (when using the maximum amount of time).

The experiment concluded with a follow-up questionnaire including 10 items to measure participants' speech familiarity. Participants indicated their prior contact with different types of spoken English ("English spoken with an Austrian accent," "English spoken with an accent similar to the Austrian accent," "English spoken with another non-native accent," "English spoken with a native accent," and "spoken English in general") on a 6-point scale (ranging from *none at all* to *very much*) and their current exposure to the same types of spoken English on a 7-point scale (ranging from *never* to *daily*). Additional data were collected on participants' attitudes toward the speaker but not included in the present analysis due to lack of space.

Data analysis

Intelligibility—specifically, SWR—was operationalized as an exact word match transcription. Misspellings (apart from accidental capitalization or accidentally hitting a numeral key) were not accepted, which inevitably penalized lower proficiency listeners but which was necessary to ensure a satisfactory level of objectivity in data coding.[8] Added punctuation marks and transcriptions of parts of the carrier sentence in addition to the correct target word (e.g., "a van" for *van)* were accepted.

The data were analyzed using a generalized mixed effects model in R (version 4.1.3, https://www.R-project.org/) using *RStudio* (version 2022.2.0.443, http://www.rstudio.

---

[6]Participation via smartphones or tablets was prevented via a filter in the online system.

[7]This was done because the variable of speaker accent in the carrier sentences was impossible to control for precisely, which would have introduced an unwanted confounding factor. That is, listeners might have sometimes been unable to use the available co(n)text not due to a general cognitive inability to do so but because a certain *prime* word was unintelligible to them. The effect of co(n)textual support would thus have been distorted by the variable nonnative pronunciation of the speaker.

[8]No consensus concerning ambiguous orthographic variations was reached when consulting other raters. Though these amounted to only 2.8% of all entries, they were more frequent for certain target words than for others and particularly frequent for the word *servant*, which might have affected the results (see Item Analysis).

com/) and the *lme4* package (version 1.1-28; Bates et al., 2015). Correct word identification (yes/no) was entered as the binomially distributed dependent variable with fixed binomial totals of six (as each condition contained six target words). The model thus predicted the per-trial *probability of a word being correctly identified* (henceforth PCI) in terms of log odds, which can be converted into probabilities (0–100%). A fixed or random effect was included in the model if it decreased the Akaike information criterion (AIC; Akaike, 1974) and significantly improved the model's fit ($p < .05$) as assessed by a log-likelihood ratio test. The optimizer *bobyqa* was used to avoid convergence errors.

In addition to the variables condition and listening proficiency, two measurements of speech familiarity were included in the model. These were component scores derived from a principal component analysis (PCA) involving the 10 items measuring participants' speech familiarity, which was run on the entire data set (including native listeners and A1 and A2 listeners) from which the present subset of nonnative listeners was taken to preempt issues of collinearity in the regression analysis (e.g., Tomaschek et al., 2018). The PCA revealed two dimensions of speech familiarity, a general one (henceforth *general familiarity*) relating to participants' familiarity with other native and nonnative accents in English and with spoken English in general, and a specific one (henceforth *specific familiarity*) relating to the particular accent they were listening to (the Austrian accent) and similar accents in English.

The model's random effects structure included a random intercept for listener as well as a by-listener random slope for condition to account for the fact that participants might react differently to the four conditions (i.e., to the availability of co-text and context). The benefit of this random slope was assessed by comparing two "beyond optimal" models (Zuur et al., 2009), each of which contained fixed effects for condition, the two familiarity variables, and the interaction term Condition × Listening Proficiency but which differed with respect to the inclusion of the random slope. Despite the inclusion of the interaction term, which already captured a substantial amount of by-listener variance with respect to the effect of condition, the random slope still resulted in a significantly better model fit ($p < .001$, decrease in AIC: 49.2) but also a singular fit. Singular fits are not uncommon (Bates et al., 2015, p. 25) but may indicate an overfitted model. However, the sizeable standard deviations for the Syn+Sem and the Syn+Sch slope (≥ 0.85 on the logit scale, Table 3) suggested that the random slope was not a superfluous effect. The singular fit was due to the high correlation of the random effects for the three slopes of condition. That is, participants who had higher slopes in the Syn condition also had higher slopes in the Syn+Sem and Syn+Sch conditions to virtually the same extent. This pattern was most likely a consequence of the experimental design rather than of overparameterization or the model failing to estimate these parameters correctly, with participants who profited particularly from the syntactic cue in the Syn condition also doing so in the Syn+Sem and the Syn+Sch conditions. The random slope was therefore retained, which also constituted the more conservative option as omitting random slopes can increase the

Table 3. Random effects of the beyond optimal model including the by-subject random slope for condition

Groups	Name	Variance	SD	Corr		
Listener	(Intercept)	0.218	0.467			
	Syn	0.038	0.196	−0.47		
	Syn+Sem	0.756	0.869	0.18	0.78	
	Syn+Sch	0.718	0.848	0.02	0.87	0.99

danger of Type I errors (Barr et al., 2013). Moreover, the model evaluation procedure for the model's fixed-effects structure yielded exactly the same results with and without the random slope.

The model's fixed effects structure was then constructed by forward stepwise regression, with condition being entered first, followed by listening proficiency, general familiarity, and specific familiarity. The Condition × Listening Proficiency interaction improved the model's fit further and was therefore also included in the final model.

Results

Effect of condition across different listener groups

The summary of the final model is provided in Table 4. Intelligibility in the C condition for C2 listeners was set as the reference category (intercept). Condition significantly affected the intercept, with intelligibility being significantly higher for C2 listeners in the SYN, the SYN+SEM, and the SYN+SCH condition than in the C condition ($p < .001$, respectively), which is in line with the notion that highly advanced nonnative listeners are able to rely on co-text and context in SWR of another nonnative accent. Moreover, condition and listening proficiency interacted significantly in several combinations

Table 4. Model summary

Fixed effects	Estimate	SE	z	p
INTERCEPT	−0.734	0.083	−8.795	.000***
CONDITION				
SYN	1.872	0.109	17.239	.000***
SYN+SEM	4.144	0.216	19.196	.000***
SYN+SCH	3.290	0.171	19.272	.000***
LISTENING PROFICIENCY				
B1	−0.561	0.195	−2.875	.004**
B2	−0.335	0.143	−2.344	.019*
C1	−0.112	0.122	−0.917	.359
GENERAL FAMILIARITY	0.191	0.047	4.104	.000***
SPECIFIC FAMILIARITY	0.150	0.04	3.79	.000***
CONDITION × LISTENING PROFICIENCY				
SYN × B1	−0.568	0.221	−2.567	.010*
SYN+SEM × B1	−1.686	0.314	−5.365	.000***
SYN+SCH × B1	−1.695	0.285	−5.941	.000***
SYN × B2	−0.025	0.174	−0.142	.887
SYN+SEM × B2	−0.746	0.275	−2.713	.007*
SYN+SCH × B2	−0.811	0.236	−3.441	.001**
SYN × C1	−0.186	0.159	−1.171	.242
SYN+SEM × C1	−0.615	0.265	−2.318	.020*
SYN+SCH × C1	−0.301	0.227	−1.327	.185

Random effects	Variance	SD	Corr		
Listener (Intercept)	0.218	0.467			
SYN	0.038	0.196	−0.47		
SYN+SEM	0.756	0.869	0.1	8	0.78
SYN+SCH	0.718	0.848	0.02	0.87	0.99

Note. Signifiance codes: ***$p < .0005$, **$p < .005$, *$p < .05$.

(ranging from $p < .001$ to $p = .020$), meaning that as expected the effect of condition was regulated by a listener's proficiency level. This variability is visualized in Figure 1, which shows the estimated per-trial PCI across different listener groups and conditions. The PCI values (Table 5) have been obtained using *allEffects()* from the *effects* package (version 4.2-1; Fox, 2003; Fox & Weisberg, 2019).

Figure 1 illustrates that all listener groups profited from the cues in the SYN condition. The size of this effect is very similar for B2, C1, and C2 listeners, who all exhibit a significant gain in intelligibility ($p < .001$, respectively)[9] equaling an increase in PCI of 40%–44%. In fact, the effect of SYN is no different for C1 and B2 listeners than for C2 listeners ($p = .242$, $p = .887$). That is, upper-intermediate and low-advanced listeners seem to benefit from syntactic information when identifying words spoken in another nonnative accent to a very similar extent as highly advanced listeners. However, B1 listeners exhibit a significantly smaller effect of SYN than C2 listeners ($p = .010$), which nevertheless amounts to an increase in PCI of 29% and is significant at $p < .001$.

Moving on to the effect of additional semantic information, all listener groups are predicted to exhibit similar gains in PCI from the SYN to the SYN+SEM condition (21%–
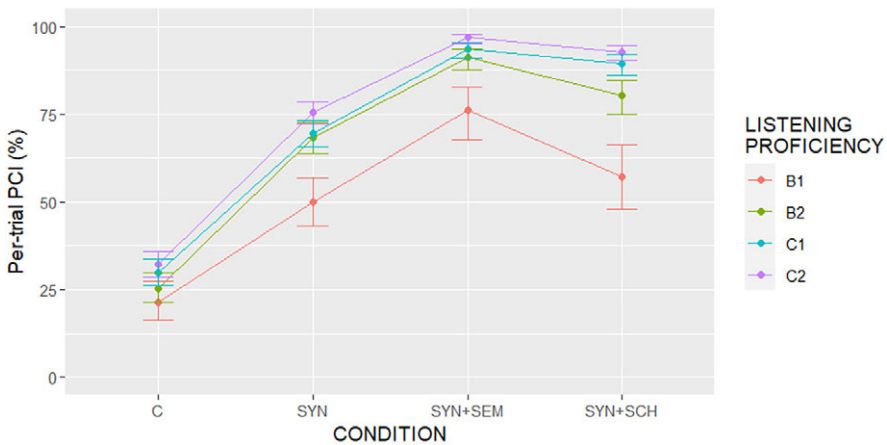

Figure 1. Effect of condition on intelligibility, for listeners at different proficiency levels. Error bars depict 95% CIs.

Table 5. Per-trial PCI (%) for each condition, for listeners at different proficiency levels

	C2	C1	B2	B1
C	32	30	25	21
SYN	76	70	68	50
SYN+SEM	97	94	91	76
SYN+SCH	93	89	80	57

---

[9]Any *p* values not included in Table 4 have been obtained by rerunning the model after adjusting the reference category for condition and/or listening proficiency and adjusted using Holm's (1979) method.

26%). For each of them, the effect of Syn+Sem vis-à-vis the Syn condition is significant at $p < .001$. Interestingly, the estimated gain in PCI is slightly smaller for C2 listeners (21%) than for B2 and C1 listeners (23% and 24%) and highest for B1 listeners (26%). However, this is not due to an increased ability in B1 listeners to profit from semantic cues but a ceiling effect on the probability scale for the other three groups, especially C2 listeners, who are approaching the maximum value of 100% PCI in the Syn+Sem condition. In other words, there is not much more to be gained in terms of intelligibility for these listeners. In terms of odds ratio, the effect of the Syn+Sem vis-à-vis the Syn condition is in fact highest for C2 listeners (OR = 9.67), followed by C1 listeners (OR = 6.31), B2 listeners (OR = 4.69), and B1 listeners (OR = 3.17). The difference between C2 listeners and C1 listeners in terms of this slope is nonsignificant ($p = .298$), but the intelligibility benefit from the Syn to the Syn+Sem condition is significantly greater for C2 listeners than for B2 listeners ($p = .027$) and B1 listeners ($p < .001$).

The additional schematic cue in the Syn+Sch condition is also predicted to result in gains in PCI relative to the Syn condition for all listener groups, albeit consistently lower ones than through the Syn+Sem condition. Moreover, the effect of the additional schematic cue is more variable across the different listener groups. Although it is significant for C2, C1, and B2 listeners ($p < .001$, respectively), it results in a substantially higher gain in PCI for the former two groups (17% and 19%) than for the latter (12%). The difference in slopes between C2 and C1 listeners is nonsignificant ($p = .597$), but the one between C2 and B2 listener is significant at $p = .003$. For B1 listeners, the predicted gain in PCI is quite low (7%) and the effect is nonsignificant ($p = .298$).

Effects of condition and speech familiarity compared

Figure 2 visualizes the model's predictions regarding the simple main effects for general familiarity and specific familiarity. Both familiarity indices are estimated to significantly increase a listener's chance to identify a word correctly ($p < .001$), with the effect being slightly stronger for general familiarity than for specific familiarity.

Figure 3, computed with the *sjPlot* package (version 2.8.10, by Lüdecke, 2021), provides an overview of how these effects compare to that of condition for each group of listeners in terms of OR. An OR below 1 signifies a decrease in the odds to correctly identify a word, whereas an OR above 1 signifies an increase. In each plot, the variables are ordered from greatest to smallest OR. For all four listener groups, the variables follow the same order: the variable with the highest OR (and thus the greatest effect on intelligibility) is the level Syn+Sem of condition, followed by Syn+Sch, Syn, and the two familiarity indices. Whereas the ORs of the three levels of condition vary across the graphs (due to the Listening Proficiency × Condition interaction), the values for the between-subject variables are the same in the four plots. As indicated above, the most important between-subject predictor is general familiarity, followed by specific familiarity. Whereas the effects of these two variables can be compared relatively easily as they constitute regression coefficients operating on similar scales, comparing their effects to that of a categorical variable such as condition is more challenging.

Importantly, the ORs of general familiarity and specific familiarity relate to an increase of 1 on the variable's scale. The range of general familiarity is 6.5 (from -5.4 to 1.1), so an increase from minimum to maximum on the general familiarity scale amounts to an OR of 3.45 (i.e., $1.21^{6.5}$, as with every 1-unit increase, the odds for correct SWR increase by a factor of 1.21). This is the maximum possible effect (MPE) of general familiarity on intelligibility. For specific familiarity (range = 4.3), the MPE
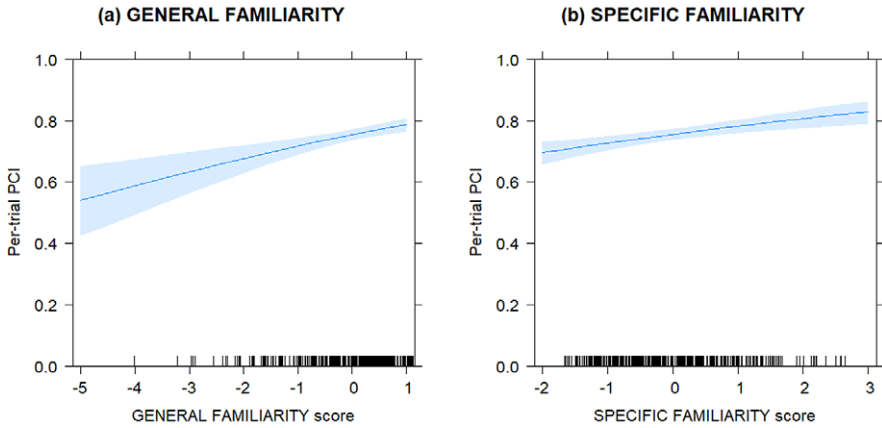
Figure 2. Effects of (a) general familiarity and (b) specific familiarity on intelligibility. Light blue bands depict the effects' CIs.
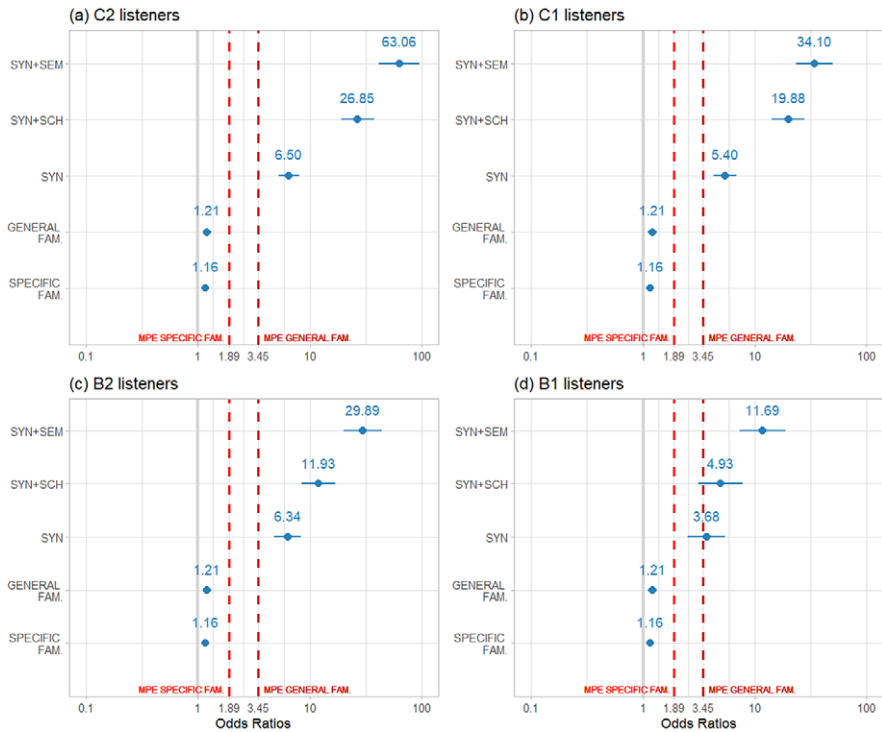

Figure 3. Effect of condition (Syn, Syn+Sem, and Syn+Sch; C is the intercept), general familiarity and specific familiarity on intelligibility in ORs for listeners at different proficiency levels. Error bars show 95 % CIs. Dashed red lines show the MPE of general familiarity (dark red) and specific familiarity (light red).

amounts to an OR of 1.89. These MPEs are visualized in Figure 3 as red dashed lines. For C2, C1, and B2 listeners, the MPE of general familiarity is still considerably lower than the OR of Syn (see Figure 3a–c), which involved the embedding of the target word in a very simple syntactic structure. Thus, for these three listener groups, even the simplest co-textual information still has a greater positive influence on intelligibility than an extreme increase in general familiarity and in specific familiarity. Looking at the ORs of the Syn+Sch and the Syn+Sem condition, which range between 11.93 and 26.85 (Syn+Sch condition) and between 29.89 and 63.06 (Syn+Sem condition) for C2, C1, and B2 listeners, the MPEs of general familiarity and of specific familiarity seem even smaller.

For B1 listeners, the effect of Syn is not necessarily larger than the MPE of general familiarity (note the CIs in Figure 3d) but noticeably larger than the MPE of specific familiarity, with 95% certainty. The OR of the Syn+Sch condition is estimated to be either slightly higher, with 4.93, than the MPE of general familiarity or just about the same (see the lower bound of the CI of Syn+Sch in Figure 3d), whereas the OR of the Syn+Sem condition is estimated to be far greater, with 11.69. Both effects are far greater than the MPE of specific familiarity.

Item analysis

To examine whether the observed differences in intelligibility applied consistently to all six target words within each condition, an item analysis was conducted. As can be seen in Figure 4, the seven most difficult words included all six target words in the C condition as well as one word from the Syn condition, thus corresponding to the finding that words in the C condition were least intelligible to listeners. The difficulty of the Syn word *gas*, which was consistent across all listener groups (same rank as in Figure 4a for all of them), can be explained by its status as a special type of Trap-Dress MP word where both words are of the same POS. Because the simple syntactic co-text in the Syn condition was insufficient to disambiguate *It's a gas/guess*, *gas* was similarly difficult as words in the C condition. In contrast, due to the additional semantic and schematic cues, the same type of MP word in the Syn+Sem and the Syn+Sch conditions (*pan-pen*, *pants-pence*) was much more intelligible (for a full discussion, see Thir, 2020).

The most intelligible target words—that is, the top third, were mostly words in the Syn+Sem condition, occupying the top three ranks in Figure 4a. However, we also find two words from the Syn+Sch and the Syn conditions, respectively. Thus, the trend that Syn+Sem words were most intelligible only holds to some extent. Whereas the special status of the MP word *pan* (see above) might explain its somewhat reduced intelligibility compared with the other Syn+Sem words, it is difficult to explain the comparatively low intelligibility of *van* (though notably, it was still identified correctly in 75% of all cases). Interestingly, *van* underperformed to a similar extent for all listener groups, being in 13th (C1 listeners), 14th (C2 and B2 listeners), or 15th place (B1 listeners).

A somewhat mixed image appears when considering words in places 9–16 in Figure 4a. This middle third is mostly occupied by Syn+Sch words but also includes two words from the Syn and the Syn+Sem conditions, respectively. In particular, the Syn+Sch words *cab* and *servant* seemed to underperform. Regarding *servant*, this might be explained by difficulties in spelling and the strict coding scheme adopted, as it was the word exhibiting the greatest amount of spelling variation in the study. Had the most common variation (*servent) been accepted, an additional 8% of items would
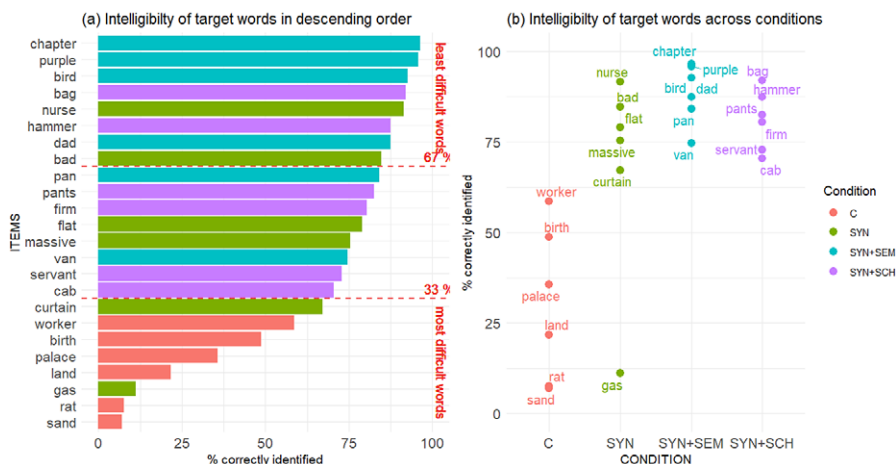
Figure 4. Intelligibility of all 24 target words, (a) in descending order and (b) across conditions.

count as correctly identified. Notably, this spelling was proportionally much more frequent among B1 listeners (18% of all entries) than among all other listener groups (10% for B2 listeners and 5% for C1 and C2 listeners, respectively). A further explanation is imprecision in the schematic cue: Had the plural *servants* been accepted, which seems equally logical on the basis of the available cue, an additional 6% of answers would have been correct. Taking these two adjustments together, the word *servant* would have been correctly identified in 87% of all cases, putting it in the top third in Figure 4. The underperformance of *cab*, might, as one reviewer suggested, have to do with its somewhat outdated status in times of *Lyft* and *Uber*. Notably, this word was particularly tricky for B1 listeners, for whom it was the 5th most difficult word, so lack of familiarity with this word might indeed have been the issue.

Summarizing, the item analysis clearly mirrors the finding that words in the C condition were most difficult to understand. It also roughly reflects the finding that words in the Syn+Sem condition were easiest to understand, followed by words in the Syn+Sch condition, though these trends are less clear (see also Figure 4b). Regarding words in the Syn condition, there is quite some variability: whereas *gas* was particularly difficult for listeners (for the reasons mentioned above), the words *nurse* and *bad* were more intelligible than expected. The observed difference between the Syn and the Syn+Sem (or the Syn+Sch) conditions thus seems partly attributable to the difficulty of identifying a particular type of MP word in a semantically and schematically neutral sentence co-text.

Discussion

RQ1 asked whether nonnative listeners at different proficiency levels (intermediate to low-advanced) profit from different types of co-textual and contextual cues in SWR of another nonnative accent. The current study provided evidence for this assumption regarding syntactic, semantic, and schematic information, operationalized as the Syn, the Syn+Sem, and the Syn+Sch condition, for most proficiency groups examined. The results suggested that nonnative listeners below C2 level are able to exploit simple syntactic cues when listening to another nonnative accent, with the experienced

intelligibility benefit being particularly substantial for low-advanced (C1) and upper-intermediate (B2) listeners. Moreover, listeners at intermediate level (B1) and upward also profited significantly and quite substantially from additional semantic cues (Syn+Sem condition). These findings are in line with other studies that found sentential co-text to benefit intelligibility to L2 listeners (e.g., Mack, 1992; Lagrou et al., 2013; Osimk 2009) but contrast with an earlier claim that listeners below the highly advanced level would be unable to use such information compensatorily when listening to another nonnative accent (Jenkins, 2000, 2002). A beneficial effect of additional schematic information (Syn+Sch condition), however, could only be observed for upper-intermediate and low-advanced listeners in the present study, for whom it was markedly less pronounced than the effect of semantic information. In other words, nonnative listeners at intermediate level might not be able to benefit from additional schematic cues as operationalized in this study. One might argue that schematic information is less accessible to many nonnative listeners from various linguacultural backgrounds because schemata are highly culture dependent and thus necessarily more elusive than semantic or syntactic relationships between linguistic entities. However, part of this finding seems due to two underperforming target words in the Syn+Sch condition, one of which can be partly explained by a slight imprecision in the cue provided as well as spelling difficulties that seem to have affected B1 listeners in particular, whereas the other might be explained by a lack of word familiarity.

RQ2 asked whether nonnative listeners below C2 level benefit from syntactic, semantic, and schematic cues in SWR of another nonnative accent to a lesser extent than listeners at C2 level (i.e., highly advanced ones). The results of the current study suggest that upper-intermediate (B2) and low-advanced (C1) listeners do not differ from highly advanced nonnative listeners with respect to their experienced benefit of syntactic and semantic information when recognizing words spoken with another nonnative accent. Moreover, there was no difference between highly advanced and low-advanced listeners with respect to the benefit of an additional schematic cue. In contrast, upper-intermediate listeners may not be able to exploit such cues to the same extent as advanced listeners. Importantly, although the current study clearly suggests that there are differences between nonnative listeners at different proficiency levels with respect to their ability to profit from co-textual and contextual cues, it also shows that highly advanced nonnative listeners are by no means the only group to effectively exploit syntactic, semantic, and schematic cues when listening to another nonnative accent and that low-advanced and upper-intermediate listeners may differ little or not at all from them in this respect. This contrasts with an early claim regarding highly advanced listeners' superior ability to profit from co(n)textual cues when listening to another nonnative accent (Jenkins, 2000).

RQ3 asked about the relative influence of co-textual and contextual information on intelligibility compared to that of speech familiarity for nonnative listeners at different proficiency-levels. For listeners at upper-intermediate level and above (B2–C2), the current study suggests the superior influence of co-textual and contextual information on intelligibility with respect to two types of speech familiarity (*general familiarity*, i.e., familiarity with different types of native and nonnative English accents and spoken English in general, and *specific familiarity*, i.e., familiarity with the particular accent one is listening to and similar accents). For B1 listeners, the data suggest that co-textual and contextual information are more important than specific familiarity and that a combination of syntactic and semantic information has a greater positive influence on intelligibility than general familiarity. However, whether general familiarity or syntactic

information and syntactic plus schematic information is more important for intelligibility to B1 listeners could not be assessed with certainty.

Overall, co-textual and contextual information emerged as the most important variable for intelligibility to listeners at upper-intermediate level and above and as at least the second most important variable for intelligibility to listeners at intermediate level. These results stand in contrast to the idea that co-textual and contextual information would play a subordinate role for international intelligibility to most nonnative listeners due to their overreliance on acoustic information. However, this finding is reminiscent of Gass and Varonis's (1984) study, which compared different types of familiarity (e.g., familiarity with nonnative speech and familiarity with a particular nonnative accent) with respect to their effect on intelligibility to L1 English listeners. Notably, the type of familiarity most critical for increasing intelligibility was "familiarity with topic"—that is, contextual or background knowledge.

One limitation of the current study that needs to be acknowledged relates to its ecological validity. Clearly, real-world conditions of processing language differ from those in the current experiment, where listeners could focus on the written cues and audio stimuli. In naturally occurring interactions in international contexts, listeners might not be able to draw on co-text and context to the same extent due to processing overload (having to interact and listen at the same time) or because the surrounding co-text of a word is (partly) unintelligible (an issue prevented here by additionally presenting carrier sentences in writing), leading to greater dependency on the acoustic signal. However, the present experiment bore other challenges that listeners typically do not have to face in naturally occurring face-to-face encounters, such as a lack of visual support in the form of gestures or facial expressions or the need to transcribe the target word under a time constraint. It may be argued that the transcription task may have encouraged listeners to focus on a target word's phonological form and thus resulted in a greater amount of bottom-up processing than in naturally occurring language use. From this point of view, the results of the present study appear particularly remarkable.

Another limitation is the fact that listening proficiency was determined via self-assessment, an approximate measure in comparison to, for example, standardized test scores. Consequently, the findings stratified by listening proficiency might only be approximate. This limitation was accepted in favor of obtaining a large, internationally diverse sample including numerous listeners from remote locations and of enabling straightforward comparison of the results to Jenkins' (2000) findings with the help of the CEFR scale.

Finally, the fact that a single L2 talker supplied the stimuli for the present study reduces the generalizability of its results. More research is necessary to examine whether the proficiency-related modulation of co(n)textual effects on international intelligibility persists across L2 talkers of varying proficiency (and L1 background), ideally with L1 talkers as comparators. An inclusion of L1 listeners, which unfortunately exceeded the scope of this paper, would be equally desirable, to provide a more complete picture of international intelligibility.

Conclusion

Until today, co-textual and contextual effects have received relatively little attention in research on international intelligibility. This study has addressed this research gap, expanding the scope of existing intelligibility research concerning co-textual

and contextual cues to nonnative LF listeners. Its results clearly demonstrate the significance of co-textual and contextual information for the intelligibility of L2-accented speech to L2 listeners at various proficiency levels, even when compared with speech familiarity. Its findings are thus in line with the idea that listening in general and SWR in particular are interactive processes in which bottom-up and top-down information are combined to arrive at a meaningful interpretation of speech. However, it also suggests that linguistic proficiency plays a crucial role in nonnative listeners' ability to rely on co-text and context in SWR of another nonnative accent and that its regulatory power with respect to international intelligibility in general (see also Kang, Moran, et al., 2020) and co(n)textual effects in particular deserves greater attention than it currently receives. Future research would thus benefit from adopting a more fine-grained differentiation between listeners at different proficiency levels rather than generalizing across nonnative listeners when studying international intelligibility, as has often been done (e.g., Deterding 2013, Gardiner 2019, Jenkins 2000).

Given the importance of co-textual and contextual effects for international intelligibility, it seems crucial that they receive greater attention both in academic research and in language pedagogy. Regarding the former, the findings of the present study call for a shift in focus, away from exclusively studying the role of particular pronunciation features, to encompass the co-textual and contextual factors that regulate the importance of target-like pronunciation for mutual understanding among nonnative users of English (see also Thir, 2020). Another relevant factor in this respect (and fruitful avenue for future research) may be syllable structure and word length, as the importance of target-like sound production may also depend on how much compensatory phonological material within a word (i.e., word-internal co-text, often termed word-internal phonological context) is available to support listener understanding.

The methodological implications of the present study also need to be recognized, as co-text and context clearly constitute potential confounding factors that need to be taken account of in future research on international intelligibility. In particular, this concerns the choice and construction of stimuli texts and sentences, whose comparability in terms of semantic and schematic predictability needs to be ensured in order to avoid co(n)textually-induced biases. Otherwise, differences in intelligibility between, for example, different sound substitutions might arise not because one feature is actually easier to understand but because the co(n)text in which it occurred gave stronger cues about the intended words than in the case of other features.

Concerning language pedagogy, there are several ways in which the importance of co-textual and contextual information for mutual understanding among nonnative users could be addressed in the English language classroom. Although learners will obviously benefit from pronunciation instruction to attain a certain threshold of international intelligibility and from increasing their familiarity with different accents of English, it seems important to also equip them with communicative skills that will enhance their intelligibility whenever their pronunciation skills may fail them. This involves co(n)textualizing—namely, supporting listener understanding by providing helpful co-textual and contextual cues (e.g., a topic cue or an associated term) to bring them into the right frame of mind to successfully recognize an intended word (for further suggestions, see Kennedy & Trofimovich, 2008). Learners should also be supported in developing co(n)textual sensitivity—that is, an understanding of *when* they will most likely have to engage in co(n)textualizing to preempt loss of intelligibility (e.g., when talking about a topic listeners are unfamiliar with) and when they will have to pay particular attention to their pronunciation because their interlocutor might be largely dependent on the acoustic signal

(e.g., when the possibility to co(n)textualize is limited). As information on listeners' background knowledge is not always available, developing an adaptability to spontaneously engage in contextualizing (and a sensitivity to when this becomes necessary) seems equally important. Such strategies and metalinguistic knowledge might be helpful in preempting miscommunication among interlocutors from different linguacultural backgrounds in the first place, making international communication in a vast number of real-world settings more efficient and less prone to communication breakdown.

References

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

Baese-Berk, M. M., Bent, T., & Walker, K. (2021). Semantic predictability and adaptation to nonnative speech. *JASA Express Letters*, *1*(1), 015207-1–015207-6. https://doi.org/10.1121/10.0003326

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Behrman, A., & Akhund, A. (2013). The influence of semantic context on the perception of Spanish-accented American English. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1567–1578. https://doi.org/10.1044/1092-4388(2013/12-0192)

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Bent, T., Holt, R. F., Miller, K., & Libersky, E. (2019). Sentence context facilitation for children's and adults' recognition of native- and nonnative-accented speech. *Journal of Speech, Language, and Hearing Research*, *62*(2), 423–433. https://doi.org/10.1044/2018_JSLHR-H-18-0273

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(4), 2074–2085. https://doi.org/10.1121/1.427952

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cambridge University Press. (2012). *English vocabulary profile* [online resource]. http://vocabulary.englishprofile.org/staticfiles/about.html

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment.* https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Deterding, D. (2013). *Misunderstandings in English as a lingua franca: An analysis of ELF interactions in South-East Asia*. De Gruyter Mouton.

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, *32*(3), 363–377. https://doi.org/10.1016/j.system.2004.05.002

Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*(15), 1–27. https://doi.org/10.18637/jss.v008.i15

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression*. Sage Publications.

Garcia, J. M., & Cannito, M. P. (1996). Influence of verbal and nonverbal contexts on the sentence intelligibility of a speaker with dysarthria. *Journal of Speech, Language, and Hearing Research*, *39*(4), 750–760. https://doi.org/10.1044/jshr.3904.750

Gardiner, I. A. (2019). Intelligibility of an L2 variety in ELF interactions. *Journal of Second Language Pronunciation*, *5*(2), 247–268. https://doi.org/10.1075/jslp.17049.gar

Gass, S. M., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(1), 65–87. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Goh, C. C. M., & Vandergrift, L. (2021). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge.

Gu, P. Y., Hu, G., & Zhang, L. J. (2005). Investigating language learner strategies among lower primary school pupils in Singapore. *Language and Education*, *19*(4), 281–303. https://doi.org/10.1080/09500780508668682

Gureckis, T. M., & Goldstone, R. L. (2011). Schema. In P. C. Hogan (Ed.), *The Cambridge encyclopedia of the language sciences* (pp. 725–727). Cambridge University Press.

Halliday, M. A. K. (1999). The notion of "context" in language education. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (pp. 1–24). John Benjamins.

Hilpert, M. (2008). New evidence against the modularity of grammar: Constructions, collocations, and speech perception. *Cognitive Linguistics*, *19*(3), 491–511. https://doi.org/10.1515/COGL.2008.019

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6* (2), 65–70. http://www.jstor.org/stable/4615733

Hustad, K. C., & Beukelman, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech, Language, and Hearing Research*, *44*, 497–510. https://doi.org/10.1044/1092-4388(2001/039)

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, *23*(1), 83–103. https://doi.org/10.1093/applin/23.1.83

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

Kang, O., Moran, M., Ahn, H., & Park, S. (2020). Proficiency as a mediating variable of intelligibility for different varieties of accents. *Studies in Second Language Acquisition*, *42*(2), 471–487. https://doi.org/10.1017/S0272263119000536

Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics*, *41*(4), 453–480. https://doi.org/10.1093/applin/amy053

Kaur, J. (2011). Intercultural communication in English as a lingua franca: Some sources of misunderstanding. *Intercultural Pragmatics*, *8*(1), 93–116. https://doi.org/10.1515/IPRG.2011.004

Kennedy, S. (2021). Difficulties understanding L2 speech due to discourse- versus word-level elements. *Journal of Second Language Pronunciation*, *7*(3), 315–342. https://doi.org/10.1075/jslp.18015.ken

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*(3), 459–489. https://doi.org/10.3138/cmlr.64.3.459

Koster, C. J. (1987). *Word recognition in foreign and native language: Effects of context and assimilation.* Foris Publications.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, *16*(3), 508–517. https://doi.org/10.1017/S1366728912000508

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. https://doi.org/10.1097/00003446-199802000-00001

Lüdecke, D. (2021). *sjPlot: Data visualization for statistics in social science* (Version 2.8.10) [Computer software]. https://CRAN.R-project.org/package=sjPlot

Macaro, E., Vanderplank, R., & Graham, S. (2005). *A systematic review of the role of prior knowledge in unidirectional listening comprehension*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/MFL_rv2.pdf?ver=2006-03-02-125000-953

Mack, M. (1992). How well is computer-processed speech understood? A cross-linguistic and cross-dialectal analysis. *World Englishes*, *11*(2–3), 285–301. https://doi.org/10.1111/j.1467-971x.1992.tb00073.x

Mauranen, A. (2006). Signaling and preventing misunderstanding in English as lingua franca communication. *International Journal of the Sociology of Language*, *2006*(177), 123–150. https://doi.org/10.1515/IJSL.2006.008

Mauranen, A. (2018). Conceptualising ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 7–24). Routledge.

Magnuson, J. S. (2017). Mapping spoken words to meaning. In M. G. Gaskell & J. Mirkovic (Eds.), *Speech perception and spoken word recognition* (pp. 76–96). Routledge.

Mirman, D. (2017). Zones of proximal development for models of spoken word recognition. In M. G. Gaskell & J. Mirkovic (Eds.), *Speech perception and spoken word recognition* (pp. 97–115). Routledge.

Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & John M. Levis (Eds.), *The handbook of English pronunciation* (pp. 377–396). Wiley-Blackwell.

Osimk, R. (2009). Decoding sounds: An experimental approach to intelligibility in ELF. *Vienna English Working Papers*, *18*(1), 64–89. https://anglistik.univie.ac.at/fileadmin/user_upload/i_anglistik/Department/Views/Uploads/0901final_1.pdf

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker-listener accent similarity. *The Journal of the Acoustical Society of America*, *130*(3), 1653–1662. https://doi.org/10.1121/1.3613698

Pitzl, M.-L. (2010). English as a lingua franca in international business: Resolving miscommunication and reaching shared understanding. VDM Müller.

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.

Smit, U. (2010). English as a lingua franca in higher education: A longitudinal study of classroom discourse. Mouton de Gruyter.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), The other tongue: English across cultures (pp. 75–90). University of Illinois Press.

Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, *120*(1), 433–442. https://doi.org/10.1121/1.2203595

Stringer, L., & Iverson, P. (2019). Accent intelligibility differences in noise across native and nonnative accents: Effects of talker-listener pairing at acoustic-phonetic and lexical levels. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2213–2226. https://doi.org/10.1044/2019_JSLHR-S-17-0414

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Thir, V. (2020). International intelligibility revisited: L2 realizations of NURSE and TRAP and functional load. *Journal of Second Language Pronunciation*, *6*(3), 458–482. https://doi.org/10.1075/jslp.20012.thi

Thir, V. (2021). The role of co-textual and contextual cues for intelligibility in ELF interactions. In A. Mauranen & S. Vetchinnikova (Eds.), *Language change: The impact of English as a lingua franca* (pp. 267–290). Cambridge University Press.

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, *71*, 249–267. https://doi.org/10.1016/j.wocn.2018.09.004

Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford University Press.

Widdowson, H. G. (2011). Context and co-text. In P. C. Hogan (Ed.), *The Cambridge encyclopedia of the language sciences* (pp. 221–222). Cambridge University Press.

Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition*, *9*(3), 307–326. https://doi.org/10.1017/S0272263100006707

Wells, J. C. (2008). *Longman pronunciation dictionary* (3rd ed). Pearson Education Longman.

Yoneyama, K., & Munson, B. (2010). The influence of lexical factors on word recognition by native English speakers and Japanese speakers acquiring English: A first report. *Journal of the Phonetic Society of Japan*, *14*(1), 35–47. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.183.1265

Zipf, G. K. (1965). *The psycho-biology of language: An introduction to dynamic philology*. MIT Press. (Original work published 1935)

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer.