Topics in Linguistics (2016), 17(1), pp. 30-44

DE GRUYTER
OPEN

10.1515/topling-2016-0003

Is there an interlanguage speech credibility benefit?

Václav Jonáš Podlipský
Šárka Šimáčková
David Petráž
Palacký University, Czech Republic

Abstract
Some (though not all) previous studies have documented the interlanguage speech intelligibility benefit (ISIB), i.e. the greater intelligibility of non-native (relative to native) speech to non-native listeners as compared to native listeners. Moreover, some studies (again not all) found that native listeners consider foreign-accented statements as less truthful than native-sounding ones. We join these two lines of research, asking whether foreign-accented statements sound more credible to non-native than to native listeners and whether difficult-to-process (less comprehensible) utterances are less credible. In two experiments we measure the intelligibility, comprehensibility and credibility of native and foreign-accented statements for native listeners and non-native listeners matched or mismatched in L1 with non-native talkers. We find an ISIB in both matched and mismatched non-native listeners, and an analogous matched comprehensibility benefit. However, we obtain no evidence of an interlanguage speech *credibility* benefit. Instead, both matched and mismatched non-native listeners tend to trust native statements more (i.e. statements produced by their target-language models). For native listeners, we do not confirm the tendency to mistrust non-native statements, but we do find a moderate correlation between the comprehensibility and credibility of foreign-accented utterances, giving limited support to the hypothesis that decreased perceptual fluency leads to decreased credibility.

Key words
second-language acquisition, foreign accent, speech perception, intelligibility, comprehensibility, credibility, perceptual fluency

Introduction

Foreign accent is a well-known and much studied phenomenon which still remains an intriguing subject of scientific inquiry. Although foreign-accentedness is not necessarily always damaging to communication, its effects are frequently negative. A large body of literature is devoted to discovering factors that influence its degree, such as the age of the onset of second language (L2) acquisition, the amount of and quality of L2 input, and the frequency of first language (L1) use (for a review see e.g. Piske et al., 2001). This research is helpful in efforts to minimize foreign accent by eliminating its causes as much as possible. Similarly extensive is the literature examining the consequences of foreign accent. An easily recognized consequence is reduced intelligibility. Importantly, the intelligibility of an L2 utterance is not its inherent property, but it instead arises from the interaction between the utterance itself and the linguistic

experience of the listener (Munro et al., 2006). Several studies document the so-called interlanguage speech intelligibility benefit (Bent and Bradlow, 2003; Hayes-Harb et al., 2008), i.e. the greater intelligibility of non-native (relative to native) speech to non-native listeners as compared to native listeners. In other words, the intelligibility of L2 utterances changes as a function of the listener's linguistic background and experience. However, the consequences of foreign accent reach far beyond a reduction in intelligibility. Foreign-accentedness triggers off attitudinal reactions on the part of the listener which can be both conscious and subconscious (for a review see Gluszek and Dovidio, 2010) and which impact upon the judgment of both the speaker's personality and of the message. For instance, native listeners consider foreign-accented statements as less truthful than native-sounding ones (Lev-Ari and Keysar, 2010). Interestingly, existing literature has not yet investigated whether and how these judgments change as a function of the listener's linguistic background, similarly to the way *intelligibility* does. The present study aims to address this topic, hoping to provide a better understanding of the underlying dynamics of the judgments foreign-accented speech provokes.

1.1 Intelligibility and comprehensibility of foreign-accented speech

Researchers in foreign accent have studied L2 speech in terms of different dimensions. Munro and Derwing (1995; 1999; Munro et al., 2006) defined the *intelligibility* and *comprehensibility* of L2 speech as two differing concepts that need to be considered separately. They defined the intelligibility of an utterance as the extent to which it is actually understood and the comprehensibility of an utterance as the listener's difficulty in understanding it. We adopt these definitions of the terms in the present study.

Literature on the intelligibility of foreign-accented speech is abundant. Early on, researchers realized that intelligibility cannot be measured in absolute terms. Empirical findings emerged confirming informal observations that non-native listeners may find non-native talkers at least as intelligible as native talkers, which is not the case for native listeners (Smith and Rafiqzad, 1979; van Wijngaarden, 2001; van Wijngaarden et al., 2002; Bent and Bradlow, 2003; see also references therein). Bent and Bradlow termed this effect the *interlanguage speech intelligibility benefit* (ISIB) and distinguished between a "matched" ISIB (for listeners with the same L1 as the non-native talkers) and a "mismatched" ISIB (for non-native listeners with different L1 backgrounds). That is, they observed an intelligibility benefit not only for listeners sharing an L1 with the talkers but even for non-native listeners with other L1s. Later, Hayes-Harb et al. (2008) also recognized the difference between an ISIB "for talkers", which is the form of the ISIB that we have described so far, and an ISIB "for listeners", i.e. the situation when non-native listeners find non-native speech more intelligible than native listeners do.

The ISIB has aroused considerable interest in subsequent literature and a number of studies have appeared both in support of it (e.g. Hayes-Harb et al., 2008; Pinet and Iverson, 2010; Song, 2011; Xie and Fowler, 2013) and against it (Major et al., 2002; Munro et al., 2006; Stibbard and Lee, 2006; Smith et al., 2009). One reason for this discrepancy is that the emergence of the ISIB depends on the talker's as well as the listener's L2 proficiency (van Wijngaarden et al., 2002; Hayes-Harb et al., 2008; Pinet and Iverson, 2010). Also, it is a question whether or not a baseline difference between native and non-native speech perception skills (e.g. Lecumberri et al., 2010) should be corrected for when comparing native and non-native listeners' intelligibility scores (cf. Hongyan and Heuven, 2007), which increases the likelihood of finding a non-native benefit, or instead whether a benefit should be defined strictly as an outperformance of native listeners by non-native listeners in absolute terms (Stibbard and Lee, 2006), in which case many of the "benefits" reported in the literature disappear. For the present study though, this is only a secondary question. The basic idea of the ISIB, i.e. that non-native talkers and listeners enjoy a mutual intelligibility advantage, is intuitively appealing and it is a useful inspiration for innovative research.

1.1 Attitudes towards foreign-accented speech

It has been long known that a foreign accent leads to certain evaluations of speakers (Lambert et al., 1960). Besides interpreting the meaning of what is being said, listeners respond to other information conveyed in the spoken message, including information about the speaker's accent and hence their identity. Studies show that foreign-accentedness arouses spontaneous reactions in native listeners, often, though not always, evoking negative attitudes, judgments and biases (e.g. Mulac et al., 1974; Cargile, 1997; Frumkin, 2007). A non-native accent is a clear signal of out-group identity and a stimulus to evaluate the speaker less favourably (than a member of one's own group). Even very young children disfavour a foreign accent given the choice between native and non-native speech (Kinzler et al., 2007; 2009; Souza et al., 2013).

One consequence of such evaluation of foreign-accented speech is that adult native listeners tend to perceive their non-native interlocutors as less trustworthy. Lev-Ari and Keysar (2010) showed that true or false statements produced by non-native speakers (uninformed about the truth value of the statements) sounded less credible to native listeners than statements produced by native speakers. Studies of spontaneous lie-telling found non-native speakers to be more likely labelled as liars and native speakers as truth-tellers by native listeners (DaSilva and Leach, 2013; Evans and Michael, 2014). However, a bias against speakers with a foreign accent is not always found (Evans et al., 2013).

It is possible that the reduced credibility of foreign-accented speech stems from stereotypes against non-native speakers. Lev-Ari and Keysar (2010) tried to eliminate prejudice as a factor in the design of their study: the native listeners knew that the non-native speakers were only "delivering messages", and in their second experiment, listeners had been warned about a potential bias against non-native speakers. Despite that, the listeners could not help mistrusting more strongly foreign-accented statements. This is perhaps because perceivers' evaluations of the incoming stimuli are largely subconscious and automatic (Ferguson and Zayas, 2009).

Lev-Ari and Keysar (2010) interpret their findings in terms of fluency of processing, which is known to affect the evaluation of perceived information. For instance, Reber and Schwarz (1999) found that printed statements that were easier to read were judged as more truthful than those more difficult to read against a white background. Lev-Ari and Keysar propose that a reduction of perceptual fluency due to a foreign accent causes reduced credibility. However, their results do not actually offer any evidence for this proposal, since a stereotype-induced bias could be at play, despite their claims to the contrary. Furthermore, Souza (2012) found no difference in credibility between clean and noise-masked statements, which doesn't support the link between perceptual fluency and perceived credibility.

1.1 Present research questions

As has already been suggested, the present study joins the two lines of research reviewed above, namely research exploring the effects of interactions between talker and listener L1 backgrounds on intelligibility and research on the conscious or subconscious evaluation of accented speech. The first and main research question is derived from a combination of these two areas of research and it is in fact stated already in the title of this paper: is there an interlanguage speech *credibility* benefit? In other words, unlike previous studies, we ask whether foreign-accented statements (relative to native-sounding ones) sound more credible to non-native than to native listeners.

The second research question focuses on the underlying mechanism of this potential effect. Lev-Ari and Keysar (2010) considered processing fluency to be the driving factor of reduced credibility of accented statements for native listeners. However, they provided no evidence that this was the factor that had produced their results and not any other, and Souza's (2012) findings do not confirm this either. Therefore, we explicitly test whether difficult-to-process (less comprehensible) utterances are less credible.

2. Experiment I

The aim of the first experiment was to obtain measurements of the comprehensibility of native-sounding and foreign-accented

statements from both native and non-native listeners. (The non-native listeners were matched in their L1 with the speakers who had read the foreign-accented statements). The purpose of measuring comprehensibility was to enable the addressing of our second research question, namely assessing the connection between processing difficulty (i.e. comprehensibility) of these statements and their credibility, measured in Experiment II.

2.1 Method

2.1.1 Stimuli

We compiled a list of 50 trivia statements (to be used as stimuli in both Experiment I and II), for instance *Listening to music when eating influences your taste*. Most of them were those used by Lev-Ari and Keysar (2010)[1] and the rest we invented ourselves. All the statements are listed in Appendix 1. Content-wise, they were trivia often about the animal world but also about other subjects. The statements were selected with the aim to minimize the chances that listeners would know for a fact whether the statements were true or false. Half of the statements were true and the other half were false but sounded plausible.

The statements were recorded by 8 speakers (all in their twenties): 4 native speakers of American English (2 female and 2 male), who were residing in the Czech Republic at that time, and 4 advanced, mildly foreign-accented Czech learners of English (again 2 female and 2 male), who were graduate students of English philology at Palacký University, Olomouc. During the recording, speakers read the statements in random order, off a computer screen, at least twice each. The recording took place in a soundproof booth using a Zoom H4n digital recorder with a 44.1 kHz sampling rate and 16bit quantization. None of the speakers were told that half of the statements were false until after the recording.

A fluent rendition of each statement was selected, so that the final set contained a single copy of each statement. Out of the 50 statements, 6 were used only as example stimuli during task familiarization (both in Experiment I and II). The remaining 44 statements were the actual target stimuli, comprising 22 statements by the native English speakers (11 of which were true) and 22 statements by the non-native speakers (11 of which were again true).

All the finally selected recordings were scaled to equal intensity and mixed with speech-shaped noise at the signal-to-noise ratio (SNR) of 0 dB (i.e. the signal and the noise were of equal intensity). The noise was constant throughout each statement, but it was created uniquely for that statement (to have the same overall spectral profile as the signal) using a script in Praat (Boersma and Weenink, 2015).

The purpose of the addition of noise was twofold. First, it was to emphasize the difference between native and non-native speech for native listeners and thus increase the likelihood of finding a difference between them in comprehensibility (Experiment I) as well as in credibility (Experiment II). Previous research on intelligibility (e.g. Rogers et al., 2004) has suggested that with the addition of noise the gap between native and non-native utterances widens.

The second purpose had to do with the design of Experiment I. To obtain filler stimuli that would have low intelligibility, we created one copy of each of the actual target stimuli and mixed it with a higher level of speech-shaped noise: the SNR for a given copy was chosen randomly from the interval (−16, −8) dB. If only the fillers were mixed with noise and the target stimuli were noise-free, participants might focus on the absence versus presence of noise, which could bias their responses.

2.1.2 Procedure

We operationalized comprehensibility of each of the 44 target stimuli in terms of reaction times (measured from the offset of a stimulus to the listener's response). This was considered less impervious to response noise than asking listeners to rate the comprehensibility of the speaker's accent explicitly, which requires considerable metacognition and may thus not be tapping into actual comprehensibility reliably.[2] The price to pay for using reaction times as a

[1] We are grateful to Shiri Lev-Ari for sharing the statements with us.

[2] In fact, we had used an explicit comprehensibility rating in a pilot version of the experiment. We found

measure of comprehensibility is that they are likely to reflect other factors besides the accent, such as the length of the statements, their grammatical complexity or speech rate. However, Munro and Derwing (1995) asked native listeners to make explicit ratings of their comprehensibility of native-accented and non-native-accented statements and reported that utterances rated low for comprehensibility took longer to process, confirming that reaction times are a good index of perceptual difficulty caused by accent. In the present study, the overt task for the listeners was to make a binary judgement about the intelligibility of each utterance by simply deciding, as quickly as they could but without sacrificing accuracy, whether they had understood all of the words in a given statement or not. The experiment was implemented as a script for the Praat Demo window (Boersma and Weenink, 2015). A screenshot of the experimental interface is shown in Figure 1.

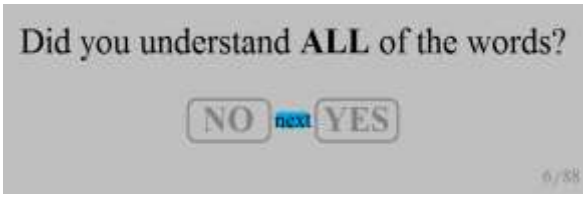

Figure 1: A screenshot of the experimental interface from Experiment I.

In each trial, listeners (tested in a quiet room and wearing circumaural headphones – Sennheiser HD 202) heard a stimulus, then (after the playback of the stimulus finished) the buttons marked "YES" and "NO" turned blue (i.e. became clickable), and listeners indicated their response by clicking on the respective button. When they were ready for the next trial, they clicked on the just-activated "next" button, located centrally to equalize the distance of the mouse cursor to each response button for the next trial. The labelling of the response buttons ("YES" on the left vs on the right) was random but it remained unchanged throughout the experiment for each participant. First, the 6 example statements were presented in 6 task-familiarization trials, and after a short break the experiment started, with 44 (the target stimuli) plus 44 (the fillers) trials. Participants could take a short silent break after the 30th and the 60th trial.

The 88 stimuli were presented in quasi-random order following two rules. First, a target stimulus (i.e. one of the 44 statements with the noise level of 0 dB SNR) was always presented before its filler copy (with a raised level of noise), never after it. This was to ensure that the comprehensibility of the target stimulus could not be increased by repeated presentation of the same statement. Second, there were at least 6 other trials between the trial presenting a target stimulus and the later trial presenting its filler copy.

2.1.3 Participants

There were two groups of listeners. The first group had the same L1 background as the native speakers who recorded our stimuli. They were 12 native speakers of American English, aged between 18 and 32, 7 female and 5 male, and they were tested in the United States by a Czech research assistant with an excellent command of English. They reported normal hearing and had had no or very limited experience with Czech or Czech-accented English.

The second group consisted of listeners who had the same L1 background as the speakers who produced our foreign-accented stimuli. They were 11 Czech advanced learners of English, again students of English philology at Palacký University, Olomouc, aged between 19 and 22, 10 female and 1 male.

2.2 Results and discussion

Responses to fillers were discarded. However, we did not discard the explicit binary ratings of intelligibility. These data were not crucial for the assessment of comprehensibility of the stimuli, which was the aim of this experiment, but they allow us to estimate the intelligibility of the utterances, which is a useful supplementary analysis. For each listener we computed the mean number of "yes" responses to the 22

the responses we obtained too noisy and therefore decided to change the design of the experiment.

native-English target stimuli and the mean number of "yes" responses to the 22 Czech-accented target stimuli (i.e. the cases in which the listener declared to have understood all the words in the stimulus). We submitted all listeners' means to a repeated-measures analysis of variance (RM ANOVA) with Listeners' L1 (English or Czech) as the between-subject factor and Stimulus Accent (native English or Czech-accented English) as the within-subject factor. A significant main effect of Listeners' L1 was found (F[1, 21] = 13.062, p = .0016). Unsurprisingly, the native listeners' overall understanding of the stimuli was better (they understood 21 out of the 22 stimuli on average) than the Czech learners' (who understood 19.14 stimuli on average). No significant effect of Stimulus Accent was found. However, Stimulus Accent interacted with Listeners' L1 (F[1, 21] = 5.891, p = .0243). As shown in Figure 2, and again as could be expected, the native listeners responded "yes" more often for the native-accented stimuli than for the Czech-accented stimuli (a Fisher's post-hoc test found this to be a significant difference, p = .016), whereas for the Czech listeners the reverse pattern could be seen (although for them the difference between the two stimulus accents was not significant). More importantly, the native listeners declared that they understood fully the native stimuli more often than the Czech listeners did (p < .001), while for the Czech-accented stimuli the difference between listener groups was not significant.

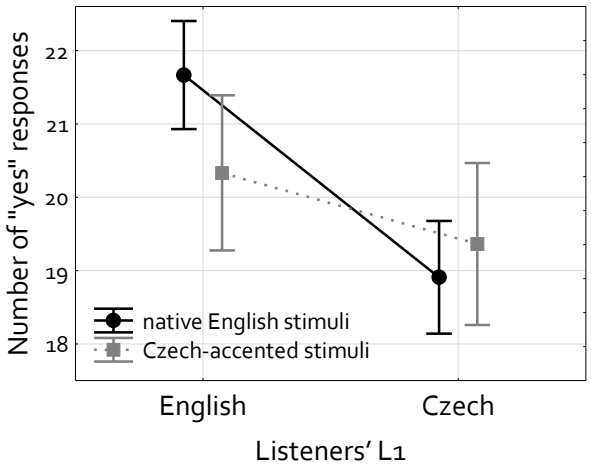

Figure 2: The mean number of "yes" responses (i.e. of the cases in which the listener understood all the words in the stimulus) split by the listener's L1 and stimulus accent. In this and all the figures below, error bars indicate 95% confidence intervals.

Next, we analysed the time listeners took to respond, i.e. the reaction times (RT). This was the actual planned measure of the comprehensibility of the stimuli and so it was the main dependent variable in this experiment. For each listener we computed the mean RT the listener took to respond to the target stimuli produced by native speakers and another mean for the stimuli produced by the Czech learners. All listeners' means were submitted to a RM ANOVA with Listeners' L1 and Stimulus Accent as the factors. Again, a significant main effect of Listeners' L1 was found (F[1, 21] = 4.835, p = .0392). As expected, native listeners' responses were faster (1.0 s on average) than the learners' responses (1.4 s on average). And again, while Stimulus Accent alone was not found to have a significant effect, it did interact with Listeners' L1 (F[1, 21] = 9.612, p = .0054). As confirmed by a Fisher's post-hoc test, and as displayed in Figure 3, even this time the interaction was in the expectable direction: the native listeners' responses to the native English stimuli were faster than their responses to the Czech-accented stimuli (p = .0161), while for Czech learners the difference was not significant (they tended to be slower with native English stimuli). The between-listener comparisons

are more important: for native stimuli, the native listeners responded faster than the Czechs (p = .0054), whereas for Czech-accented stimuli the listener groups did not differ significantly.

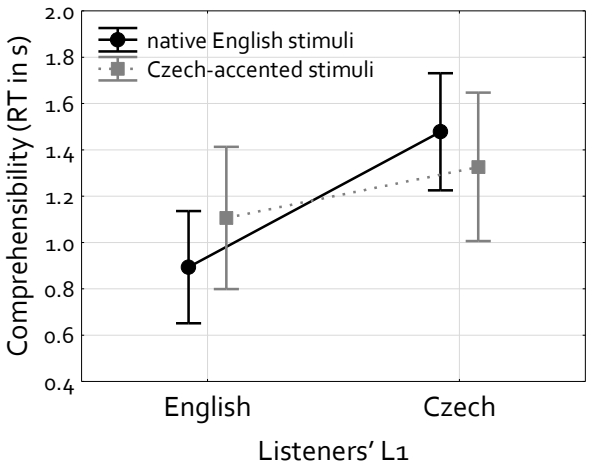

Figure 3: The mean reaction times shown by the listeners' L1 and stimulus accent.

It is noteworthy that the RT results correspond to the numbers of "yes" responses very well. The same effects were statistically significant and the patterns observed in Figures 2 and 3 are inverted reflections of each other. That suggests that the stimuli listeners took longer to respond to were often those they did not fully understand. Similarly, Munro and Derwing (1995) found that utterances explicitly rated low for comprehensibility took longer to process.

To summarize, the native listeners showed lower intelligibility (as shown by the explicit binary ratings) as well as lower comprehensibility (as indexed by higher RTs) of the non-native than of the native stimuli. This was expectable and it confirms previous findings (Munro and Derwing, 1995). The same was not true for non-native listeners in the present experiment, for whom we did not find evidence of an effect of stimulus accent on intelligibility and on comprehensibility either. This pattern of results can be described as a matched ISIB for talkers (using the terminology introduced by Hayes-Harb et al., 2008), such as that reported in previous literature (e.g. van Wijngaarden et al., 2002; Bent and Bradlow, 2003; Xie and Fowler, 2013), and an analogous matched interlanguage speech *comprehensibility* benefit for talkers, which represents a novel finding. Our results do not show an intelligibility or comprehensibility benefit for listeners. This is due to the fact that overall, native listeners outperformed the non-native listeners in both measures. Given that all stimuli were mixed with speech-shaped noise of equal intensity as the signal, this could be expected. In adverse listening conditions, even advanced L2 learners fall short of native-like perceptual performance (for a review, see Lecumberri et al., 2010). As suggested by Hayes-Harb et al.'s (2008) results, it is possible that the overall native-listener advantage in the present experiment would not be found and an ISIB for listeners would emerge if we used utterances produced by more strongly foreign-accented learners.

As noted above, the main purpose of this experiment was to obtain a comprehensibility profile of each stimulus, whose credibility could subsequently be assessed with new listeners in Experiment II,[3] and the potential relationship between the credibility and the comprehensibility of each stimulus could thus be tested. This is why we also computed for each stimulus the mean time all the native listeners took to react to it, as well as the mean RT across the non-native listeners. Importantly, the RTs showed sufficient variability, ranging between 0.63 and 3.99 s. This was necessary to ensure that our second research question (whether statements with lower comprehensibility tend to be rated as less truthful) could be tested. Had the comprehensibility of different stimuli been too similar, any potential differences in credibility due to differences in comprehensibility may be too small to detect.

[3] We had to recruit new listeners for Experiment II because we wanted to avoid repeated exposure to identical stimuli.

3. Experiment II

Having collected data about the comprehensibility of the native and foreign-accented statements, the purpose of this experiment was to measure how credible they sounded to native and to non-native listeners (both matched and mismatched in L1 background with the speakers who had produced the foreign-accented statements).

3.1 Method

3.1.1 Stimuli

The stimuli were the exact same 44 target statement recordings (plus the 6 examples used for task familiarization), including the 0 dB SNR speech-shaped noise, that had been used in Experiment I. The fillers from Experiment I were not used. The stimuli were exactly the same for each participant.

3.1.2 Procedure

Like Lev-Ari and Keysar (2010), we asked listeners to make explicit judgements of the truthfulness of the statements. Lev-Ari and Keysar were at pains to eliminate potential prejudice against non-native speakers. Prior to testing, they recorded their listeners reading statements supposedly for future participants, and so made their listeners aware that the speakers were not the authors of the statements. In the present experiment, we simply emphasized in the instructions that the speakers did not know which statements were true.

Figure 4 is a screenshot of the experimental interface which was again implemented using Praat (Boersma and Weenink, 2015). On each trial, listeners heard a stimulus and then they could click on a scale ranging from "definitely false" on the left to "definitely true" on the right (this orientation was the same for all participants). As seen in the figure, the scale consisted of 18 points with no marking. The first 6 trials presented the example statements and served for task familiarization. Then the 44 target stimuli were presented in random order, each once.

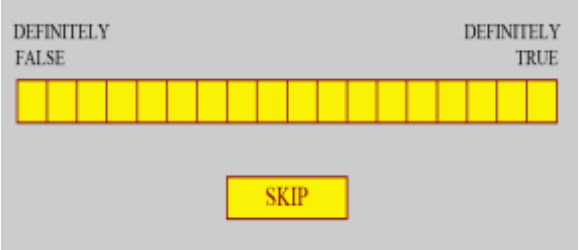

Figure 4: A screenshot of the experimental interface from Experiment II.

To keep the credibility data as noise-free as possible, a "skip" button was provided that listeners were instructed to click only in two cases: when they knew for a fact whether the statement was true or false, or when they did not understand all of the words. Since the topics of the statements were chosen to go beyond the common knowledge of an average person, we assumed the "skip" responses would mostly represent the listener's failure to understand the statements, that is, they could be used as a rough index of intelligibility (just like the "no" responses in Experiment I).

3.1.3 Participants

For this experiment, 43 listeners were recruited, none of whom had taken part in Experiment I. They formed three groups. The first group was comprised of 18 native listeners, i.e. listeners with the same L1 as the native speakers who recorded our stimuli. They were native speakers of American English, aged between 19 and 42, 9 female and 9 male. They reported normal hearing and had had no or very limited experience with Czech or Czech-accented English. They were tested individually in the United States by a Czech research assistant with a high proficiency in English.

The second group were non-native listeners matched for L1 background with the speakers who produced our non-native stimuli. They were 18 Czech advanced learners of English, again students of English philology at Palacký University, Olomouc, aged between 19 and 22, 11 female and 7 male. They were tested in groups in a computer lab at Palacký University.

Finally, the third "mixed" group were 7 non-native listeners mismatched in L1

background with the speakers. They all rated their command of English as advanced, but they were more heterogeneous in L2 proficiency than the second group. Their L1s were Slovenian, Polish, Finnish, French, or Hindi. As they were all tested during their short visits to the Czech Republic, they had had some experience with Czech (though none of them could speak Czech) and Czech-accented English.

3.2 Results and discussion

We first analysed the frequency of "skip" responses. Listeners' numbers of skipped native stimuli and of skipped Czech-accented stimuli were submitted to a RM ANOVA with listeners' L1 background (English, Czech, or "mixed") as the between-subject factor and Stimulus Accent (native English or Czech-accented English) as the within-subject factor. While on average the native listeners skipped fewer stimuli (1.89) than the Czech and the mixed groups did (3.67 and 3.64 respectively), the ANOVA found that the main effect of L1 background narrowly exceeded significance (F[2, 40] = 3.125, p = .0548). There was a significant main effect of Stimulus Accent (F[1, 40] = 4.670, p = .0367), with the native stimuli being skipped more often (3.53 times on average) than the Czech-accented ones (2.61 on average). Clearly, this difference was due to the "skip" responses by the non-native listener groups and not the native group, as confirmed by the significant interaction between L1 background and Stimulus Accent (F[2, 40] = 7.730, p = .0015) which can be observed in Figure 5. A post-hoc Fisher's test found a reversed likelihood of "skip" responses in the native listeners and the non-native listener groups: the native listeners skipped Czech-accented stimuli more often than native sounding ones (p = .0475), whereas the opposite was true for both the Czech listener group (p = .0129) and the mixed group (p = .0153). To describe the results in terms of between-listener differences, the native listeners skipped native stimuli less often than both the Czech listener group (p < .001) and the mixed group (p = .0032), who did not differ significantly from each other, but for the Czech-accented stimuli there were no significant differences between the three listener groups.

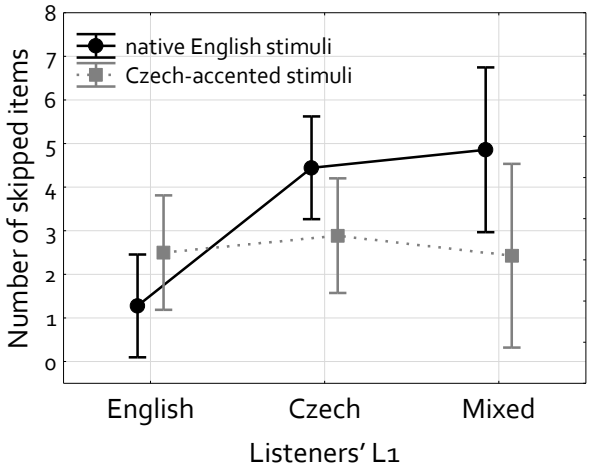

Figure 5: The mean number of "skip" responses split by the listeners' L1 background and stimulus accent.

Although listeners were instructed to skip not only unintelligible statements but also those whose (un)truthfulness they were certain of, these results can still be used to estimate the intelligibility of the two stimulus accents for the different listener groups. The differences between the native and Czech listeners were in the expected direction and corroborate the intelligibility findings from Experiment I (which also tested native English listeners and Czech learners but using a different method and with different participants): even now we can observe a matched ISIB for talkers. However, this time we can also see that the same pattern of results emerged for the non-native listeners in the mixed L1 background group, which constitutes a mismatched ISIB for talkers. This gives support to Bent and Bradlow's (2003) findings, although again an ISIB for listeners did not occur, and following Stibbard and Lee's (2006) stricter view of the "benefit", the pattern shown in Figure 5 (i.e. the lack of difference between native and non-native listeners for the Czech-accented stimuli and a difference between them for the native-sounding ones) is more accurately described as a non-native listeners' intelligibility detriment.

Next, the credibility scores themselves were analysed. For each listener the mean credibility score they gave to the native stimuli and the mean score for the Czech-accented stimuli were computed and

submitted to a RM ANOVA. A significant main effect of L1 background was found ($F(2, 40) = 5.252$, $p = .0094$). A Fisher's post-hoc test showed that overall, listeners in the mixed L1 group were less likely to trust the statements (their mean credibility score was 8.6 on the 1-18 scale) than both the native English ($p = .0024$) and the Czech listeners ($p = .0177$), who were not found to differ from each other significantly (mean credibility 10.44 and 10.02, respectively). Stimulus Accent also had a significant effect ($F[1, 40] = 8.191$, $p = .0067$), with the native English stimuli receiving the mean credibility score of 10.18 across all listeners and the Czech-accented stimuli receiving the mean score 9.25. Both these effects can be ascertained from Figure 6. Crucially, this time no significant L1 background × Stimulus Accent interaction was found. As seen in the figure, the Czech-accented stimuli tended to be judged as less truthful than the native stimuli by all groups of listeners, although a Fisher's post-hoc test found the difference to be significant only for the Czech group ($p = .0226$). For the mixed L1 group, the difference approached significance ($p = .0521$). The post-hoc test found no significant differences between the listeners groups for native stimuli. Czech-accented stimuli then sounded less credible to the mixed group than to the native and the Czechs listeners ($p = .001$, $p = .030$, respectively), who did not differ from each other significantly.

Finally, for the purposes of linking the results of experiments I and II, we also computed the mean credibility score each stimulus received from the native listeners and the mean score it received from the Czech listeners. The results are reported in Section 4.

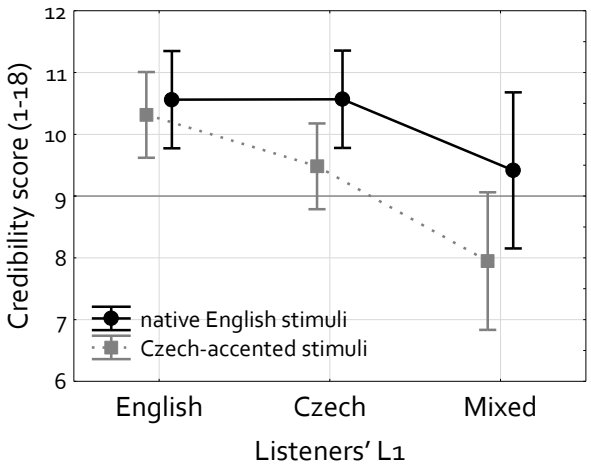

Figure 6: The mean credibility scores split by the listeners' L1 and stimulus accent. The midpoint of the scale, i.e. the value 9, is highlighted.

Let us now discuss the credibility results and relate them to previous literature. First, we found that the listener groups differed somewhat in their baseline level of trustfulness. Whereas the mismatched-L1 listeners' mean score was almost perfectly at the midpoint of the credibility scale (i.e. they were about as likely to choose a score above the midpoint as below it), both the native and the matched-L1 listeners were slightly more trustful. We may speculate that this difference arose because both the native and the Czech listeners had better perceptual skills in English than the mixed-L1 listeners for whom the statements were more difficult to understand and hence less credible. This explanation does not seem to be supported by the number of "skip" responses, which was comparable for the mixed group and the Czechs. However, first, skipping could reflect not only a lack of understanding but also known (un)truthfulness of the statements and second, even if the number of skipped statements for the Czech and the mixed group did not differ significantly, the groups could still have differed in comprehensibility.[4]

More importantly, like Evans et al. (2013), we did not find evidence of lower credibility of non-native compared to native utterances

[4] Note that we did not included a mismatched-L1 group of listeners in Experiment I, since testing a correlation between comprehensibility and credibility would make no sense for such a group, unless they were speakers of identical L1 backgrounds as in Experiment II, which was hard to achieve from a practical point of view.

for native listeners. Our results thus do not replicate the findings of Lev-Ari and Keysar (2010), DaSilva and Leach (2013), and Evans and Michael (2014). Possibly, our non-native speakers' relatively high L2 proficiency (and mild foreign-accentedness) could be the cause. Evans and Michael (2014) proposed that the discrepancy between their results and those of Evans et al. (2013) may have been due to a difference in their speakers' L2 proficiency. However, in the present study, the non-native statements *were* less intelligible for the native listeners (both in the present experiment and, for different native listeners, in Experiment I, where they were also less comprehensible for them) and so this explanation is not likely. Additionally, the fact that our results do not replicate Lev-Ari and Keysar's (2010) could be due to a difference in design: whereas in their study half of the native listeners heard a statement produced by a native speaker and the other half by a non-native speaker, removing the effect of the statement itself, our stimuli were exactly the same for all participants. Thus, we cannot eliminate the possibility that our native listeners did not trust the foreign-accented statements less than the native statements because there was a baseline difference in the credibility of the sentences we chose for native and non-native speakers.

Finally, our non-native listeners (both with the matched and mismatched L1s) were not found to trust non-native stimuli more than the native listeners. They even tended to trust non-native stimuli less than native-sounding ones, which is the opposite of what we had hypothesized based on the existing literature on intelligibility. Therefore, the current results offer no evidence of what could be described as an interlanguage speech credibility benefit. The most accurate interpretation of our results is probably that the native, rather than non-native, speakers enjoyed a credibility benefit from the non-native listeners. We cannot say that this benefit was simply due to a more convincing delivery of the statements by the native speakers, because there was no evidence of such a benefit for our native listeners (a benefit which we had expected). Therefore, we propose that the non-native listeners showed a positive attitudinal bias: perhaps they trusted the native speakers of their L2 more than non-native speakers of that language because they perceive native speakers of their target language as models. At the same time, in light of Šimáčková and Podlipský's (2012) findings that L1-Czech L2-English listeners were aware of and critical of a Czech accent in English, a negative bias against the Czech-accented statements might have affected our Czech listeners' judgments of the credibility of their peers.

4. Correlating credibility with comprehensibility

In order to address our second research question, i.e. whether difficult-to-process (less comprehensible) utterances are less credible, we tested correlations between the comprehensibility of each of our stimuli (indexed by reaction times) measured in Experiment I and their credibility measured in Experiment II. We ran Pearson's correlations separately for each listener group (English native and L1-Czech L2-English listeners) and each stimulus accent (native and Czech-accented English). Only one of the four correlations was significant: native listeners' credibility scores for the Czech-accented statements from Experiment II showed a moderate negative correlation to the times other native listeners in Experiment I took to make binary judgements of the intelligibility of these statements (r = -0.4421, p = .0394). In other words, lower comprehensibility of a non-native statement mildly predicted its lower credibility for native listeners.

This result provides limited support for the hypothesis that decreased perceptual fluency of non-native speech for native listeners results in its decreased credibility (Lev-Ari and Keysar, 2010). However, on average, the non-native listeners tended to trust the native statements more than the non-native statements, despite the fact that they were less intelligible and comprehensible for them (recall that our stimuli contained speech-shaped noise of 0 dB SNR which likely magnified the difference between native and non-native stimuli for the listeners). If the effect of processing difficulty on credibility was robust, we would expect to find correlations between the comprehensibility and credibility for the Czech listeners and for the native stimuli too.

Conclusion

This study found its inspiration in two areas of previous research: studies documenting an interlanguage speech intelligibility benefit (e.g. Bent and Bradlow, 2003; Hayes-Harb et al., 2008; Pinet and Iverson, 2010; Song, 2011; Xie and Fowler, 2013), and studies reporting a reduction of credibility of utterances with a foreign accent for native listeners (Lev-Ari and Keysar, 2010; DaSilva and Leach, 2013; Evans and Michael, 2014). Accordingly, we hypothesized that interlanguage (i.e. non-native) talkers would enjoy a benefit in credibility from non-native listeners.

First, our findings replicate a matched as well as a mismatched interlanguage speech intelligibility benefit for talkers (Bent and Bradlow, 2003; Hayes-Harb et al., 2008) and show an analogous matched comprehensibility benefit. However, we do not replicate Lev-Ari and Keysar's (2010), DaSilva and Leach's (2013) and Evans and Michael's (2014) findings, since our native listeners, as in (Evans et al., 2013), did not trust non-native statements significantly less than native ones. This was perhaps due to our non-native speakers' mild foreign accent or a difference in design.

Second, our results showed that the non-native listeners tended to trust native statements more that non-native ones. Our best interpretation of this result is that the non-native listeners' responses display an attitudinal bias favouring statements produced by their target-language models. Therefore, rather than the hypothesized interlanguage speech credibility benefit, we observed in the non-native listeners' responses a native speech credibility benefit. Interestingly, this benefit occurred only for the non-native listeners, and it was despite the reduced comprehensibility (i.e. reduced perceptual fluency) of native statements for the non-native listeners. The positive bias towards native speech thus overrode any potential detriment caused by reduced perceptual fluency.

Therefore, we did not find a link between the non-native listeners' difficulty to perceive the statements and their ratings of the credibility of the statements. Still, for the native listeners, moderate evidence of such a link was observed: they tended to judge less comprehensible non-native statements as less credible. This lends limited support to Lev-Ari and Keysar's (2010) conclusion that decreased perceptual fluency brings forth decreased credibility.

Acknowledgements:

The work of the first author was supported by the internal grant FPVC2014/23 from the Faculty of Arts, Palacký University, Olomouc.

References

BENT, T. and BRADLOW, A.R., 2003. The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, vol. 114, no. 3, pp. 1600-1610.

BOERSMA, P. and WEENINK, D., 2015. Praat: doing phonetics by computer [Computer program]. Version 6.0.05, [Accessed 8 November 2015] Available at: http://www.praat.org/.

CARGILE, A.C., 1997. Attitudes toward Chinese-accented speech. An investigation in two contexts. *Journal of Language and Social Psychology*, vol. 16, no. 4, pp. 434-443.

DA SILVA, C.S. and LEACH, A.M., 2013. Detecting deception in second-language speakers. *Legal and Criminological Psychology*, vol. 18, no. 1, pp. 115-127.

EVANS, J.R. and MICHAEL, S.W., 2014. Detecting deception in non-native English speakers. *Applied Cognitive Psychology*, vol. 28, no.2, pp. 226-237.

EVANS, J.R., MICHAEL, S.W., MEISSNER, C.A. and BRANDON, S.E., 2013. Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition*, vol. 2, no.1, pp. 33-41.

FERGUSON, M.J. and ZAYAS, V., 2009. Automatic evaluation. *Current Directions in Psychological Science*, vol. 18, no. 6, pp. 362-366.

FRUMKIN, L., 2007. Influences of accent and ethnic background on perceptions of eyewitness testimony. *Psychology, Crime & Law*, vol. 13, no. 3, pp. 317-331.

GLUSZEK, A. and DOVIDIO, J.F., 2010. The way they speak: a social psychological perspective on the stigma of non-native accents in communication. *Personality and Social Psychology Review*, vol.

14, no. 2, pp. 214-237.
HAYES-HARB, R., SMITH, B.L., BENT, T. and BRADLOW, A.R., 2008. The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, vol. 36, no. 4, pp. 664-679.
HONGYAN, W. and van HEUVEN, V.J., 2007. Quantifying the interlanguage speech intelligibility benefit. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken: Universität des Saarlandes, pp. 1729-1732.
KINZLER, K.D., DUPOUX, E. and SPELKE, E.S., 2007. The native language of social cognition. *Proceedings of the National Academy of Sciences*, vol. 104, no. 30, pp. 12577-12580.
KINZLER, K.D., SHUTTS, K., DEJESUS, J. and SPELKE, E.S., 2009. Accent trumps race in guiding children's social preferences. *Social Cognition*, vol. 27, no. 4, pp. 623–634.
LAMBERT, W.E., HODGSON, R.C., GARDNER, R.C. and FILLENBAUM, S., 1960. Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, vol. 60, no. 1, pp. 44-51.
LECUMBERRI, M.L.G., COOKE, M. and CUTLER, A., 2010. Non-native speech perception in adverse conditions: A review. *Speech Communication*, vol. 52, no. 11, pp. 864-886.
LEV-ARI, S. and KEYSAR, B., 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, vol. 46, no. 6, pp. 1093-1096.
MAJOR, R.C., FITZMAURICE, S.F., BUNTA, F. and BALASUBRAMANIAN, C., 2002. The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, vol. 36, no. 2, pp. 173-190.
MULAC, A., HANLEY, T.D. and PRIGGE, D.Y., 1974. Effects of phonological speech foreignness upon three dimensions of attitude of selected American listeners. *Quarterly Journal of Speech*, vol. 60, no. 4, pp. 411-420.
MUNRO, M.J. and DERWING, T.M., 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, vol. 38, no. 3, pp. 289-306.
MUNRO, M.J. and DERWING, T.M., 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, vol. 49, no. 1, pp. 285-310.
MUNRO, M.J., DERWING, T.M. and MORTON, S.L., 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, vol. 28, no. 1, pp. 111-131.
PINET, M. and IVERSON, P., 2010. Talker-listener accent interactions in speech-in-noise recognition: Effects of prosodic manipulation as a function of language experience. *The Journal of the Acoustical Society of America*, vol. 128, no. 3, pp. 1357-1365.
PISKE, T., MACKAY, I.R. and FLEGE, J.E., 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, vol. 29, no. 2, pp. 191-215.
REBER, R. and SCHWARZ, N., 1999. Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, vol. 8, no. 3, pp. 338-342.
ROGERS, C.L., DALBY, J. and NISHI, K., 2004. Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech*, vol. 47, no. 2, pp. 139-154.
ŠIMÁČKOVÁ, Š. and PODLIPSKÝ, V.J., 2012. Pronunciation skills of an interpreter. In: J. Zehnalová, O. Molnár and M. Kubánek, eds. *Teaching translation and interpreting skills in the 21st century*. Olomouc: Palacký University Olomouc, pp. 139-149.
SMITH, B.L., HAYES-HARB, R., BRUSS, M. and HARKER, A., 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics*, vol. 37, no. 3, pp. 257-275.
SMITH, L.E. and RAFIQZAD, K., 1979. English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, vol. 13, no. 3, pp. 371-380.
SONG, J., 2011. The interlanguage speech intelligibility benefit for experienced nonnative listeners: Perception of English lexical stress produced by Korean native speakers. In: *Proceedings of the 17th International Congress of Phonetic Sciences.* Hong Kong, pp. 1882-1885.
SOUZA, A.L., 2012. *Modality-specific effects of processing fluency on cognitive judgments*. (Doctoral dissertation, University of Texas at Austin).
SOUZA, A.L., BYERS-HEINLEIN, K. and POULIN-DUBOIS, D., 2013. Bilingual and monolingual children prefer native-accented speakers. *Frontiers in Psychology*, vol. 4, article no. 953.

STIBBARD, R.M. and LEE, J.I., 2006. Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 433-442.
van WIJNGAARDEN, S.J., 2001. Intelligibility of native and non-native Dutch speech. *Speech Communication*, vol. 35, no. 1, pp. 103-113.
van WIJNGAARDEN, S.J., STEENEKEN, H.J. and HOUTGAST, T., 2002. Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1906-1916.
XIE, X. and FOWLER, C.A., 2013. Listening with a foreign-accent: The interlanguage speech intelligibility benefit in Mandarin speakers of English. *Journal of Phonetics*, vol. 41, no. 5, pp. 369-378.

Author's address and contact details

Mgr. Václav Jonáš Podlipský, PhD.
Department of English and American Studies
Faculty of Arts, Palacký University
Křížkovského 10
77180, Olomouc
Czech Republic
E-mail: vaclav.j.podlipsky@upol.cz

Mgr. Šárka Šimáčková, PhD.
Department of English and American Studies
Faculty of Arts, Palacký University
Křížkovského 10
77180, Olomouc
Czech Republic
E-mail: sarka.simackova@upol.cz

Mgr. David Petráž
Department of English and American Studies
Faculty of Arts, Palacký University
Křížkovského 10
77180, Olomouc
Czech Republic
E-mail: david.petraz@seznam.cz

Appendix 1. The trivia statements used as stimuli in both Experiment I and II. The truth-status and the stimulus accent are shown (nat = produced by a native speaker, fa = foreign-accented).

Statement	Status	Accent
A falcon is the only bird that can see the color blue. [owls]	false	fa
A mosquito has two teeth. [forty-seven]	false	fa
A snake is communicating when it sticks out its tongue. [smelling]	false	fa
Earthworms have five brains. [hearts]	false	fa
Fifteen percent of the water covering the earth is drinkable. [One]	false	fa
Jerusalem is the oldest city in the world. [Damascus]	false	fa
Some crocodiles may eat other crocodiles. [sharks]	false	fa
The first genetically modified food was an apple. [tomato]	false	fa
The Universe is eight billion years old. [fourteen]	false	fa
There are approximately twenty thousand feathers on an eagle. [seven]	false	fa
Women blink four times as much as men. [nearly twice]	false	fa
A snail can sleep for ten years. [three]	false	nat
Almost fifty percent of the Earth's core is composed of iron. [ninety]	false	nat
Men commit suicide twice as much as women. [5 times more often]	false	nat
Only young polar bears hibernate. [pregnant females]	false	nat
People spend approximately one quarter of their lives sleeping. [third]	false	nat
Scientists know of nine kinds of ice. [nineteen]	false	nat
Sharks attack women ten times more often than they attack men. [men; women]	false	nat
The dog was domesticated about nine thousand years ago. [19-32]	false	nat
The heaviest bell in the world weighs ninety-two tons. [193]	false	nat
The koala is the only known animal that never gets sick. [shark]	false	nat
There are more species of butterflies than beetles.	false	nat
An average adult is taller in the morning than in the evening.	true	fa
Birds sing in different dialects in various regions of the world.	true	fa
Chewing bubble gum improves your concentration.	true	fa
Elephants can't jump	true	fa
Listening to music when eating influences your taste.	true	fa
Matches were invented later than the cigarette lighter.	true	fa
Owls swallow food whole because they have no teeth.	true	fa
The can opener was invented 48 years later than the can was.	true	fa
The dog was the first animal to be domesticated.	true	fa
The fur of a polar bear is not white but colorless.	true	fa
The original name for a butterfly was flutterby.	true	fa
A giraffe can go without water longer than a camel can.	true	nat
A hippo can run faster than a man.	true	nat
A human embryo is ninety-four percent water.	true	nat
A kangaroo can only jump if its tail is touching the ground.	true	nat
An ant colony may include millions of individuals.	true	nat
Ants don't sleep.	true	nat
Dogs can smell cancer.	true	nat
Homosexuality is also common in animals like elephants or penguins.	true	nat
Music is older than language.	true	nat
The blue whale is the noisiest animal in the world.	true	nat
The television was invented by a child.	true	nat