Listening in the Wrong Language: The Role of Language Dominance and Accent in Cross-language Speech Misperceptions

Mónica A. Wagner[1]
Supervisors: Kristin Lemhöfer[1], James M. McQueen[1]

[1]*Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands*

As has been well-documented in the literature, bilinguals possess a remarkable ability to switch between their languages while interacting with other bilinguals in mixed language contexts. Despite this, sometimes problems can still arise. This study addresses the rare phenomenon of speech misperceptions due to non-target language processing or "listening in the wrong language" (LWL). Our primary goal was to induce LWL states using an auditory sentence verification task with a twist: Participants were led to believe the experiment would be in one of their languages (the base language of the experiment) only to unexpectedly hear speech in their other language (the guest language) one-third of the way through the task. Failures to comprehend were measured by including a response option for when listeners did not understand an utterance. In addition, we investigated how the occurrence of misperceptions was affected by speaker accent (native vs. non-native) and whether the language was the listener's native (L1) or non-native (L2) language. The results revealed more comprehension failures for items in the guest language relative to the base language. Furthermore, the results suggest that language and accent may only initially play a role, with no significant effects observed after the guest language was presented for the first time. Moreover, listener familiarity with an accent may modulate its surprise effect. The results are discussed in the context of theories of bilingual speech comprehension.

Keywords: cross-language speech misperceptions, bilingual speech comprehension, non-native accent, language dominance

Corresponding author: Mónica Wagner; E-mail: M.Wagner@donders.ru.nl

"Selavy."

You have probably heard this phrase a thousand times. Out of context, it may be hard to locate its meaning, but if you were to say it out loud, you might recognise this string of letters as an English homophonic translation of the French phrase "c'est la vie" ("that's life").

Visually, cross-linguistic illusions like this can be induced by manipulating word boundaries and language-specific spelling rules, sending readers on a wild goose chase down the wrong language garden path. Similar phenomena can also occur during speech comprehension: Just imagine that, instead of reading the above word, someone said it to you using English pronunciation. You might not immediately recognize the utterance as French and try to process it as English.

The present study concerns itself with situations like the one described above where bilinguals experience difficulties or failures to comprehend due to non-target language processing or "listening in the wrong language" (LWL). Crucially, in order for a speech misperception to be considered LWL, the listener must be proficient in the target language and/or normally understand the utterance with ease. In a preliminary survey of 402 bilinguals[1] from all over the world, we found that nearly 60% of all respondents and over 80% of Dutch participants (N = 95) reported having experienced LWL at some point in their lives. Furthermore, of those familiar with LWL, most (80%) indicated that they experienced it rarely to occasionally. Therefore, LWL appears to be a rare but real phenomenon.

How exactly LWL states come about is a question researchers are yet to answer. Understanding how speech comprehension normally proceeds in bilinguals may help determine how this process goes awry in LWL states. One model of bilingual word recognition, the Bilingual Model of Lexical Access (BIMOLA; Grosjean, 1988, 1997; Léwy, 2015), was even initially developed to account for LWL. According to this model, phonemes and words are stored separately for each of the bilingual's languages (Léwy, 2015; Shook & Marian, 2013). The language that is accessed at any given moment depends on the global language mode the bilingual is in. Usually, [1]the bilingual will mainly process words in the "base language" or main language of an interaction due to a relatively greater activation of this language. However, when the other language, or "guest language", is also activated, the bilinguals can be said to be in a bilingual mode, and processing ensues in both languages in parallel, although independently (Grosjean, 1988; Léwy, 2015). Thus, following the BIMOLA, LWL could result from selective lexical access in the unintended or "non-target" language.

Many studies have demonstrated that non-target language activation in bilingual word recognition may actually be more common than suggested by the BIMOLA. For example, using a visual world paradigm with eye-tracking, Spivey and Marian (1999) showed that, after hearing instructions to pick up an item in English (e.g., *Pick up the marker*), Russian-English bilinguals fixated more on objects (e.g., a stamp) whose names in the non-target language (*marka*) were similar to the target object than they did on unrelated controls. In light of this and similar findings (Ju & Luce, 2004; Weber & Cutler, 2004), most models of bilingual speech comprehension advocate integrated lexicons with language nonselective lexical access (Li & Farkaš, 2002; Shook & Marian, 2013; Zhao & Li, 2007, 2010). Despite this, bilinguals usually manage to "zoom in" to the target language, as Elston-Güttler, Gunter, and Kotz (2005) call it, and eventually access the meanings of words in the right language. According to integrated lexicon models, this is accomplished in bilinguals, after initially nonselective access, by language-specific patterns of activation. These activation patterns can be explained by words in the same language being more strongly associated to each other as a result of repeated co-activation due to shared language-specific phonology (Li, 1998; Li & Farkaš, 2002; Shook & Marian, 2013). From this perspective, LWL could be explained by non-target language-specific patterns of activation, which would bias the system against the target language.

In some cases, non-target language lexical access may result in meaning, albeit not the one intended by the speaker. This is the case of, for example, near interlingual homophones ("false friends") such as *pet* which means *cap* in Dutch (although phonetically realized differently). Misperceptions of speech resulting in meanings different than the one originally intended are known as "mondegreens" (Hendriks, 2014). The term was coined by writer Sylvia Wright who, as a child, misheard a line from the Scottish song "The Bonny Earl O'Moray" as *They hae slain the Earl O'Moray, and Lady Mondegreen* instead of the original …*and laid him on the green* (Beck, Kardatzki, & Ethofer, 2014; Wright, 1954).

1 Some people prefer to reserve this term for simultaneous and/or balanced bilinguals to distinguish them from unbalanced and/or consecutive bilinguals, often called "second language learners." Here we use the term "bilingual" to refer to all speakers of two or more languages (technically "multilinguals"), clarifying with qualifiers when relevant.

Mondegreens can also occur across languages, due to near interlingual homophones, in which case they are known as Hobson Jobsons (Yule & Burnell, 1903) or Soramimi, when the objects of misperception are song lyrics. Otake (2007) analyzed 194 song lyrics misheard in English as Japanese, the listeners' native language (L1). He found that only 4% were due to purely segmental errors. The rest involved errors extending beyond word boundaries, at the phrasal level. The study of Soramimi can shed light on another way LWL can occur, namely, segmentation.

In contrast to written words, spoken words are not usually separated by pauses (Cole & Jakimik, 1980; McQueen, 1998). Instead, during speech perception, listeners are faced with the task of extracting discrete words from a continuous speech signal (McQueen & Cutler, 2010). In order to accomplish this, listeners make use of different cues that help them detect word boundaries, such as acoustical features (e.g., Gow & Gordon, 1995; Quené, 1992), metrical structure (e.g., Cutler, Dahan, & Donselaar, 1997), and phonotactic information (e.g., McQueen, 1998). The specific cues listeners use, vary depending on the language at hand (Cutler, 2012; Cutler, Mehler, Norris, & Segui, 1986; Tyler & Cutler, 2009). Studies on Soramimi suggest that LWL could be explained by the listener attending to segmentation cues from the wrong language, essentially segmenting the signal at the wrong points, a phenomenon called "juncture misperception" (Kentner, 2015; Otake, 2007).

Most studies on cross-language speech misperceptions have made use of interlingual homophones and Soramimi to demonstrate how bilinguals can end up perceiving words from the non-target language. However, these can be seen as a special case of LWL, where the end product is a (non-target) meaning. More often, though, misperceptions do not result in meaning but rather solely a failure to understand (Bond, 2008). So far, we have discussed two ways in which this might occur: non-target lexical access and segmentation, both intrinsic to the process of speech comprehension. But what extrinsic factors can cause the train to derail, if you will, and proceed in the non-target language?

One likely culprit is a factor briefly touched upon before: The context that the bilinguals find themselves in can lead them to preferentially expect one or the other language. One such contextual factor is the linguistic context, as revealed by the base-language effect (Macnamara & Kushnir, 1971), "a momentary dominance of base-language units (phonemes, syllables, words) at code-switched boundaries…" (Grosjean & Miller, 1994, p. 201). A similar effect was found in an event-related potential (ERP) study on visual comprehension in the L2 (Elston-Güttler et al., 2005). German-English bilinguals performed a lexical decision task on targets (i.e., *poison*) preceded by sentences such as "The woman gave her friend an expensive *gift*" (*poison* in German; control: *item*). Prior to the experiment, participants watched a short film with subtitles in German or English. The results revealed a semantic priming effect behaviourally and modulations in the N200 and N400 ERP components but only following the German version of the film. Moreover, these effects were temporary, disappearing after the first block.

More evidence for a potential role of context comes from work on bilingual speech production. A series of studies suggests that visual cues from the context, such as the speaker's face, whether familiar or unfamiliar, as well as cultural symbols, might bias processing towards the congruent language (Woumans et al., 2015; for a review, see Hartsuiker, 2015). Together these findings suggest that cues from the context might play a role in LWL by augmenting expectations for the non-target language.

Another factor that may contribute to LWL occurrences derives from the speech signal itself: phonetic realization. In particular, activation of the non-target language may increase if the speaker has a non-native accent, especially if the listener speaks the language associated with that accent. Often non-native speakers will even use sounds that only exist in their native language, which could increase expectations for the non-target language. In a series of studies, Grosjean and collaborators (Bürki-Cohen, Grosjean, & Miller, 1989; Grosjean, 1988; Soares & Grosjean, 1984) studied these factors in the recognition of words in the guest language. In a gating study in which participants heard words in increasingly longer fragments, Grosjean (1988) measured how long it took participants to "isolate" (i.e., accurately and consistently identify) guest words in a carrier sentence in the base language. He analyzed the role of three factors: language-specific phonotactics, language-specific phonetics, and the existence of a homophone in the base language. In addition to guest words with guest language-specific phonotactics and guest words that were not homophones, he found an advantage in isolation for guest words pronounced with guest-language phonetics relative to guest words pronounced as in the base language (Li, 1996).

Despite the fact that sentences were used in Grosjean (1988)'s study, it was always the same neutral lead-in phrase: *Il faudrait qu'on… (We should…)*. Consistent with the context effects described above, there is evidence that sentential context can help reduce the amount of non-target language interference to aid selective lexical access during bilingual speech comprehension (for a review, see Fitzpatrick & Indefrey, 2014). This means that phonetic realization may be less important for comprehension in real life than in single-word experiments (Li, 1996). Consistent with this idea, Li (1996), in a replication of Grosjean's (1988) study, found that less of the word was needed for semantically constraining sentences (but see Bürki-Cohen et al., 1989). Lagrou, Hartsuiker, and Duyck (2012) observed similar results using an auditory lexical decision task on the last word in sentences with varying semantic constraint. Critically, the last words were interlingual homophones, which have been shown to cause a delay in processing. They found that the semantic constraint of the sentences, as well as native accents, reduced the effect of interlingual homophones, although not fully eliminating it (see also Chambers & Cooke, 2009; Fitzpatrick & Indefrey, 2010).

Finally, LWL states might also be modulated by a factor pertaining to the listener, that is: whether the target language is the listener's L1 or second language (L2). This is consistent with theories proposing a reduced baseline activation of the L2 (e.g., Pallier, Colomé, & Sebastián-Gallés, 2001). Support for a role of proficiency was provided by a replication of Grosjean (1988)'s study with interlingual homophones in the participant's L1 and L2 (Schulpen, Dijkstra, Schriefers, & Hasper, 2003). There was a disadvantage for the L2, with words in this language being identified less often and, when identified, requiring longer gates. This view is also supported by studies on adverse listening conditions (Bond, 1996), where the difference between the L1 and L2 is found to be exacerbated by adverse conditions (such as noise), causing greater problems for the L2 than the L1.

Furthermore, work on Soramimi have found that the strength of their perception in the L1 correlates positively with verbal fluency in the L1. Moreover, their perception in the L2 was found to not correlate negatively with proficiency in this language. Together, these findings seem to suggest that their occurrence is related to creative solutions to ambiguous acoustic signals, rather than limited linguistic competence (Beck et al., 2014; Beck Lidén et al., 2016).

Studies on switching during language production may also suggest a greater incidence of LWL when the target language is the L1 compared to the L2. In these studies, a common finding is that switching into the L1 is harder than switching into the L2. As the reasoning goes, language-selective access during speech production in bilinguals is accomplished via inhibition. During switching, this inhibition must quickly be lifted and replaced on the non-target language. As the L2's baseline activation is less than that of the L1, speaking in the L2 calls for greater inhibition of the L1 than of the L2 during L1 production. Overcoming this relatively greater inhibition leads to longer reaction times (RTs) for switches into the L1, a phenomenon now known as an "asymmetric switch cost." Following this logic, LWLs should be rarer when the target language is the L1 given the greater amount of inhibition required to keep the L1 at bay (Meuter & Allport, 1999).

In contrast to both of these views on the role of proficiency, yet another possibility is that LWL occurs as often in the bilingual's L1 and L2. An eye-tracking study was conducted in which the effect of semantic constraint of the preceding sentence on L2 auditory sentence processing did not vary with L2 proficiency. This suggests that context effects may play a bigger role than proficiency (Chambers & Cooke, 2009). Similarly, in an ERP study with intra-sentential switching, Fitzpatrick and Indefrey (2014) found no difference between the L1 and L2 in terms of switch costs.

To the best of our knowledge, no studies to date have addressed the occurrence of LWL using a naturalistic experiment. In fact, most studies on bilingual auditory comprehension have focussed on the processing of isolated words, for example in gating or lexical decision tasks (Elmer, Meyer, & Jancke, 2010). Moreover, studies that have looked at sentence-level comprehension have primarily used intra-sentential switching (e.g., Fitzpatrick & Indefrey, 2014). However, these types of tasks are limited in the extent to which they can inform us about how speech comprehension occurs in real life. In addition, many studies have made use of words with form overlap (e.g., cognates or interlingual homophones). These words may be unique, with some studies suggesting that sentential context may better restrict cross-language activation in words without form overlap (Hartsuiker, 2015). Finally, most studies have aimed to evaluate processes resulting in successful comprehension, while here, the interest resides in those cases when comprehension breaks down.

Present study

The present study aimed to induce LWL states in Dutch-English bilinguals. To this end, measures were taken to bias processing towards the nontarget language: No mention was made of the guest language, and the experiment began with a monolingual block entirely in the base language (see Cheng & Howard, 2008 for a similar set-up in visual comprehension). Comprehension was assessed using an auditory sentence verification task with a twist: In order to gauge misperceptions accurately, in addition to the traditional "true" and "false" choices, participants were provided with a third option to indicate utterances they failed to understand. If our manipulation was successful, we would expect more comprehension failures for guest language items than base language items. Furthermore, even in cases where guest language items managed to be perceived accurately, we expected processing to take longer to be resolved (i.e., slower RTs).

The study also aimed to evaluate whether the incidence of LWL differed if the guest language was the bilingual's L1 or L2. Given the dominance of the L1 in unbalanced bilinguals, the L2 may have a reduced baseline activation and thus an L2 guest language item may be more unexpected than one in the L1. On the other hand, if bilingual speech comprehension in the target language is accomplished via inhibition of the non-target language, as has been suggested for bilingual speech production, guest language processing costs should be greater for the L1 than for the L2.

As explained above, in addition to the actual language being spoken, the phonetics of the utterance can influence the language-selectivity of lexical access during speech comprehension. Another goal of the present study, thus, was to evaluate the effect of speaker accent on the incidence of LWL. This was implemented by having listeners hear utterances produced by native and non-native speakers. Only native Dutch and English speakers were used so the non-native speaker would always be a native speaker of the other language of the experiment. We predicted that guest language items produced by a non-native speaker would be misperceived more often and processed more slowly than those spoken by a native speaker, as the accent would increase the expectation for the non-target language. Similarly, native pronunciation would help disambiguate the language being spoken, in the end facilitating comprehension.

In summary, we predicted greater processing costs (in the form of a higher incidence of misunderstandings and slower RTs) for guest language items than base language items, both in the monolingual and bilingual blocks. Moreover, we suspected that this guest language effect might be different for the L1 and L2, although we were not really sure about the direction of the difference. Finally, we expected non-native accent to exacerbate guest language misperceptions and processing costs by increasing expectation for the non-target language.

Methods

Participants

Forty-nine native Dutch speakers (age: M = 23.5, SD = 3.4, range = 18-33; 15 male) participated in the auditory sentence verification study. Participants provided written informed consent before the start of the experiment and afterwards received a €10 voucher for their collaboration.

Design

The present study was different from most studies on speech comprehension in that it aimed to study failures to comprehend and, what's more, a type of failure that occurs only on rare occasions outside the laboratory. Such an infrequent phenomenon called for a unique approach that would maximise the probability of observing these misperceptions during the experiment. To this end, a design was conceived to induce non-target language expectations, essentially tricking the participant. Two important manipulations were introduced to the experimental design. First, the experimental task began with a monolingual block in the base language to establish the expectation for that language and set participants in that language mode. Furthermore, once the guest language was introduced, in an attempt to maintain expectation biased towards the base language, the frequency of guest language items and, thus, code-switches was kept low, specifically 20% of the items.

In terms of analysis, the effect of the guest language could be observed by comparing performance on items spoken in the guest language with those spoken in the base language. This could be accomplished via comparison of base language items in the initial, monolingual block. However, base language items in the monolingual block and guest language items varied in several aspects that could complicate interpretation of the results. First of all, as called for by the design, these base language items always preceded guest language items as they occurred in the first block of the experiment. This meant that any potential effects due to the order of presentation could not be controlled for. In addition to the common concern for effects of fatigue or learning, this was particularly problematic for the present study where participants had to adapt to speakers' voices and accents.

An additional point of contrast was that guest language items, by definition, occurred in a bilingual context where both target languages were activated and the participant was required to switch from one language to the other. Therefore, any processing costs observed for guest language items could also be explained by interference from the increased activation of both of the bilinguals' languages or the fact that language mixing was more effortful. In task-switching studies, this is usually resolved by the introduction of nonswitch trials in the switch blocks, allowing for two comparisons: (1) switch trials (in switch blocks) - nonswitch trials in switch blocks and (2) nonswitch trials in switch blocks - nonswitch trials in nonswitch blocks (Hughes, Linck, Bowles, Koeth, & Bunting, 2014; Koch, Prinz, & Allport, 2005; Weissberger, Wierenga, Bondi, & Gollan, 2012). While the former measures the well known switching cost, the latter provides a measure of the cost of task mixing in general. This method was adopted in the present study, allowing for three critical conditions: (1) base language items when the participant was still in a monolingual context, (2) base language items in a bilingual context, and (3) guest language items (necessarily in a bilingual context). Most studies on task- or language-switching focus on the difference between (1) and (2): the effect of context — monolingual vs. bilingual — and/or (2) and (3): the effect of language status — base vs. guest language. However, those studies also usually conduct by-participant analyses. Given the nature of the present design, comparing conditions within participants would not have been very informative as condition differences were confounded with a change in language. Therefore, by-item analyses were preferred. Since we were interested in misperceptions, no sentence was presented twice to avoid priming effects. Thus, within-item analyses were between-participant and, to increase the power of these analyses, condition was kept as a three-level factor, instead of conducting separate analyses for the effect of context and language status.

Materials

For a comprehensive overview of the materials and pilot studies on the basis of which they were selected, see the complete version of the Materials in the online Supplementary Material.

Auditory sentence verification task (aSVT). The critical stimuli consisted of 64 Dutch and 64 English sentences, selected from an original set of 351 Dutch and 333 English sentences on the basis of two pilot studies. For each language, half of the sentences were true statements and half were false. Sentences were kept short — consisting of only three words — to ensure that participants would not be able to guess the meaning of the statement from the end of the sentence but rather had to understand all of the words. Critical sentences belonged to one of four syntactic structures: (1) noun + verbto be + noun (be + N), (2) noun + verbto be + adjective (be + Adj), (3) noun +verbto have + noun (have), or (4) noun + verbcan + verb (can; as in Collins & Quillian, 1969; for examples, see Table 1). Words whose translations were phonetically very similar (i.e., cognates) and interlingual homophones (i.e., false friends) were avoided by calculating normalized Levenshtein distances between the phonetic transcriptions (DISC; Baayen, Piepenbrock, & van Rijn, 1993) of the first and last words (hereafter "content words") words and their translations (Gooskens & Heeringa, 2004; Schepens, Dijkstra, & Grootjen, 2012). None of the critical content words exceeded .5 phonetic similarity (English sentences: $M = .14$, $SD = .14$, range = 0 - .43; Dutch sentences: $M = .11$, $SD = .12$, range = 0 - .40, $t(254) = 1.776$, $p = .077$).

Since English was not the participants' native language, words thought to be familiar to participants were chosen for the English sentences, resulting in a higher frequency for these, as can be seen in lemma frequency (per million; CELEX: $t(254) = 2.421$, $p = .022$; SUBTLEX: $t(254) = 1.824$, $p = .069$; see Table 2 for averages; Baayen et al., 1993; Brysbaert & New, 2009; Keuleers, Brysbaert, & New, 2010).[2] Nevertheless, given that "the bilingual is not two monolinguals in one person" (Grosjean, 1989), corpus-based frequencies probably do not very accurately reflect subjective frequencies for bilinguals (e.g., Connine, 2004; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008).

The critical items were pretested in a series of off-line sentence verification tasks with native Dutch speakers. Participants read the sentences and were asked to judge, for each, whether they thought the statement was true or false. An additional option was included in the English version for participants to indicate any words they did not know. Each sentence was evaluated by at least 10 raters and the

2 One English word (peels) did not appear in noun form in CELEX, so its frequency value was computed as 0.

Table 1.

Examples of critical sentences per structure, language, and veracity.

	English		Dutch	
Structure	True	False	True	False
be + N	Cars are vehicles	Uncles are women	Tafels zijn meubels *(Tables are furniture)*	Schedels zijn spieren *(Skulls are muscles)*
be + Adj	Sugar is sweet	Deserts are wet	Bergen zijn hoog *(Mountains are tall)*	Schuurpapier is glad *(Sandpaper is smooth)*
have	Rabbits have fur	Shrimp have pearls	Uilen hebben ogen *(Owls have eyes)*	Kwallen hebben botten *(Jellyfish have bones)*
can	Airplanes can move	Fish can walk	Nagels kunnen groeien *(Nails can grow)*	Hanen kunnen brullen *(Roosters can roar)*

veracity of each of the final 128 critical sentences was confirmed by at least 9 raters. A complete list of critical stimuli can be found in Appendix A of the online Supplementary Material. In addition to these critical sentences, 144 sentences (72 true; length: 3 words) and their translations were used as fillers and ten (5 true; 3-5 words) as practice items (Appendix B of the online Supplementary Material).

The number of total critical items and fillers was determined by the number of guest language items in the bilingual blocks: 20%, With a cell size of 32 (16 true, 4 per speaker), this meant a total number of 160 items, of which 64 were critical items (32 guest language and 32 base language items) and 96 fillers. Furthermore, the initial monolingual block consisted of 80 base language items (32 critical and 48 fillers).

The sentences were spoken by two native Dutch and two native English speakers selected on the basis of a series of pilot studies (for full details, see the online Supplementary Material).

Several candidate speakers were recorded reading sentences like those used in the aSVT. An online rating study was conducted with native Dutch speakers to obtain measures of accentedness (the perceived strength of non-native accent of the utterance, on a scale from 1 (*no foreign accent*) to 9 (*very strong foreign accent*) (Munro, 2008), comprehensibility (the perceived ease or difficulty in understanding the utterance, on a scale from 1 (*very easy to understand*) to 9 (*very difficult to understand*), and intelligibility (the degree to which an utterance is actually understood

Table 2.

Frequency, sentence length, utterance length, and speech rate of critical sentences.

	Variables							
	Language				Accent			
	English		Dutch		Native		Non-native	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Word frequency								
CELEX	56.58	136.53	24.60	60.91	37.58	94.41	43.60	118.04
SUBTLEX	95.48	148.33	63.60	130.72	77.02	143.15	82.06	138.19
Sentence length (syllables)	3.89	0.86	5.72	1.05	4.75	1.31	4.86	1.34
Utterance duration (ms)	1137.75	148.17	1317.81	199.27	1207.06	136.38	1248.50	242.19
Speech rate (syllables/s)	3.44	0.73	4.37	0.73	3.93	0.96	3.88	0.76

Note. SUBTLEX-US used for English

by the listener, based on transcription accuracy) in both English and Dutch (Munro, 2008; Saito, Trofimovich, Isaacs, & Webb, 2015). If the raters indicated that the speaker had a native accent (foreign accent = 1), they were then asked to guess the region where the speaker was from/regional accent they had (e.g., a province in the Netherlands or Belgium for Dutch and British/American dialect for English). If, on the other hand, they responded that the speaker had a non-native accent (foreign accent > 1), they were asked to guess what the speaker's native language was. Four speakers (one male and one female per language) were selected based on the following criteria: perceived as (1) native in their mother tongue, (2) free of a strong regional accent, and (3) moderate to strongly accented in their non-native language (English or Dutch), but (4) still understandable. The results of the rating study are shown for the final speakers in Tables 3 and 4.

The final four speakers were recorded reading the final sentences for the aSVT in a sound-attenuated booth using a Sennheiser microphone. Audio files were recorded and saved in Audacity at 44 kHz. Speech was monitored online by the first author and, after reading the entire list, speakers were asked to repeat sentences pronounced with disfluencies or gross pronunciation errors that could hinder understanding. In addition to the four speakers, two different female speakers (one native Dutch speaker and one native English speaker [the experimenter]) recorded the practice items.

Tokens were manually extracted from the audio files by auditory and visual inspection of the waveform and spectrogram in Audacity, removing silence before and after the utterances. The best (i.e., most comprehensible) of the exemplars was chosen. Sentences were quasi-randomly assigned to speakers in such a way as to evenly distribute true and false sentences across speakers.

Audio stimuli were equated in amplitude using the normalize function in Audacity. Overall, utterances averaged 1228 ms in length (SD = 196.87). An analysis of variance (ANOVA) with language and speaker L1 confirmed a main effect of language, $F(1, 124) = 40.8889$, $p < .001$, and speaker L1, $F(1, 124) = 26.928$, $p < .001$ on utterance duration. No significant interaction between language and speaker L1 was revealed, $F(1, 124) = 2.165$, $p = .144$.

An additional measure of speech rate was calculated by dividing the number of syllables in a sentence by the duration of its utterance in seconds. An ANOVA with language and speaker L1 showed a main effect of language on speech rate, $F(1, 124) = 54.651$, $p = .000$, and of speaker L1, $F(1, 124) = 9.722$, $p = .002$, but no significant interaction between language and accent, $F(1, 124) = .151$, $p = .699$.

These analyses revealed two things: (1) that Dutch was spoken faster, and (2) that native Dutch speakers spoke faster than native English speakers. However, these differences are not too problematic for the present design because, as explained above (Design) analyses were within-item and, thus, within-speaker, so any potential difference in guest language effect for native and non-native speakers could not be explained by a difference in audio duration or speech rate.

Sentences spoken by native and non-native speakers did not vary in frequency (CELEX: $t(254) = -.451$, $p = .653$; SUBTLEX: $t(254) = -.287$, $p = .774$) nor did sentences spoken by native English and native Dutch speakers (CELEX: $t(254) = .004$, $p = .997$ SUBTLEX: $t(254) = -1.005$, $p = .316$).

Table 3.

Results of speaker ratings for accentedness, comprehensibility, and intelligibility.

	Selected speakers															
	English male				**English female**				**Dutch male**				**Dutch female**			
	English		**Dutch**		**English**		**Dutch**		**English**		**Dutch**		**English**		**Dutch**	
Rating	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
ACC	1.37	0.56	6.33	2.04	1.60	1.10	5.40	2.43	4.63	2.54	1.57	1.25	6.37	2.43	1.03	0.18
COMP	1.80	1.45	5.47	2.58	2.23	2.03	4.03	2.57	2.63	1.75	1.70	1.74	3.43	2.42	1.07	0.25
INT	.97	.10	.84	.28	.94	.15	.97	.10	.94	.20	.99	.06	.99	.06	1	0

ACC = accentednesss (1 = native, 7 = very strong foreign accent)

COMP = comprehensibility (1 = very easy to understand, 7 = very difficult to understand)

INT = intelligibility (proportion of 3 words correct)

Table 4.

Characteristics of final four speakers.

	English male	English female	Dutch male	Dutch female
Age	32	28	22	36
Originally from	California, U.S.A.	California, U.S.A.	Gelderland, The Netherlands	Noord Brabant, The Netherlands
Years lived in Gelderland	2	2	22	17
Speaking English				
Perceived as native (%)	67	67	20	3
Modal perceived regional dialect	U.S.A. (60%)	U.S.A. (50%)	U.S.A. (20%)	U.K. (3%)
Perceived as Dutch speaker (%)	0	7	43	53
Speaking Dutch				
Perceived as native (%)	0	3	60	87
Modal perceived regional dialect	-	Gelderland (3%)	Gelderland (20%)	Noord-Holland (37%)
Perceived as English speaker (%)	13	40	0	3

Note. Percentages are out of a total of 30 tokens rated.

Language background questionnaire (LBQ). Participants completed a LBQ (in Dutch) with questions about their native language(s) and experience with non-native languages. For each non-native language named, listed in order of proficiency, the following measures were obtained: age of acquisition, frequency of use (1-Never, 2-Rarely, 3-Occasionally, 4-Sometimes, 5-Frequently, 6-Very frequently, 7-Always), and self-rated proficiency for speaking, listening, writing, and reading (1-Very poor, 2-Poor, 3-Fair, 4-Functional, 5-Good, 6-Very good, 7-Fluent).

In addition to the information about languages spoken, participants answered a few questions about their previous exposure to the accents presented during the experiment: (1) were they familiar with the accents of the native Dutch and native English speakers from the experiment, (2) what dialect of English they were most familiar with (options: American, Canadian, British, Scottish, Irish, Welsh, Australian, New Zealand, and South African), (3) how often they heard English-accented Dutch and Dutch-accented English (never, less than once a week, once a week, several times a week), and (4), for each, from how many speakers (0-1, 2-5, 6-10, more than 10; following Witteman et al., 2013). In addition, participants were asked where they were from and how long they had lived in the province of Gelderland.

Finally, a few questions were added to inquire about the incidence of LWL, that is, situations where they did not understand what someone said to them, despite speaking the language, because they were expecting the person to speak another language: (1) had they ever experienced LWL, (2) how often (1-*Never*, 2-*Rarely* [less than once a month], 3-*Occasionally* [once a month], 4-*Sometimes* [more than once a month, less than once a week], 5-*Frequently* [once a week, less than once a day], 6-*Very frequently* [once a day], 7-*Always* [several times a day]), and (3) did they experience LWL during the experiment.

LexTALE. In addition to the self-ratings of English ability provided in the LBQ, the English version of LexTALE (Lemhöfer & Broersma, 2012) was administered as an objective measure of proficiency. This test is a brief lexical decision task which measures English vocabulary knowledge. The test consists of 60 items and scores are calculated by weighing both hit and false alarm rates.

Procedure

Experimental list construction. Considering the length of the aSVT (240 items total), it was considered necessary to split the items into blocks. A first block of 80 items (32 critical) coincided with the monolingual context. To increase comparability, the bilingual context items were divided into two blocks of 80 items, as well, with a sub-set of base and guest language critical items evenly distributed into each block. To this end, critical stimuli were separated into two sets per language and rotated through the three conditions, yielding four experimental lists. The sub-sets were also rotated through experimental block, so each item appeared in each block. Furthermore, care was taken while dividing the stimuli into subsets to ensure each speaker was equally represented in each block and in each language, as well as with an equal number of true and false statements and a similar number of stimuli per verb structure. As mentioned before, controlling the frequency of each manipulation was given such importance since the effect of interest hinged on expectations and, thus, probabilities. Filler sentences were fixed to their blocks and also equally divided in terms of speaker, veracity, and sentence structure.

Speaker's identity changed on every trial, as did the sex of the speaker to avoid having to control for congruence of speaker identity and sex between trials. Similarly, critical items always followed true statements to prevent differential effects from previous statements, as false statements tend to take longer to verify than true ones (Cox, 2005; Gough, 1966). Furthermore, no more than four trials of each condition (veracity, accentedness, speaker L1) appeared consecutively. An attempt was made to make sure critical sentences did not follow sentences with the same verb (be, can, or have), when not possible, care was taken that the sentences did not contain the same conjugation form of the verb (e.g., "is"). Sentences that could be semantically associated were also kept apart. In bilingual context blocks, critical items never immediately followed a guest language item, with three to five intervening trials between guest language items. The order of the variables speaker and veracity were kept constant across the four lists, except for two items so that the first guest language could occur in both native and non-native accent conditions. Base language items in the bilingual blocks always occurred after the first appearance of the guest language item to ensure participants were in bilingual mode.

Testing. Given that the critical manipulation of the study involved an unexpected guest language, special attention was paid to the information participants received about the experiment and several measures were taken to induce a monolingual mode in an effort to maximize the expectation for the base language. Participants were recruited via the Radboud Research Participation System and were prescreened with the following information, provided in the system's general questionnaire: (1) not to suffer from any hearing problems, (2) to be between the ages of 18 and 35, (3) to have Dutch as their native language, (4) to speak Dutch, (5) not to have been raised multilingually, and (6) to speak English. However, only the first requirement was made visible to participants and no mention was made of English or the fact that the study was about language. Recruitment for both base language groups was conducted in Dutch with the premise that separate recruitment in English for the English base language group could result in a differential preselection of the participants (e.g., based on their attitudes towards and confidence in English). Thus, in order to keep English experience constant, no mention of English was made.

On the day of the experiment, participants were assigned to an experimental list and received all information about the study in the corresponding base language, including the informed consent and prescreening forms. The only exception were the instructions received orally from the experimenter, which were always given in English. However, participants of the Dutch baseline group were told that this was a limitation of the experimenter and if they asked (although very few did), participants were led to believe that the experiment would be conducted in the base language. Furthermore, the aSVT instructions were presented visually in the base language and participants were instructed not to talk to experimenter once the aSVT began. After that, a monolingual context was created by presenting the items of the practice session (10 sentences) and entire first block of the experiment (80 sentences) in the base language.

The experimental sessions were conducted in a quiet room where participants were seated in front of a computer, where they read task instructions and filled out the written surveys. After filling out the consent and prescreening forms, participants completed the aSVT task which was administered via PsychoPy (Peirce, 2007). Participants listened to the utterances with headphones and responded by pressing keys on a keyboard. Before the task began, the audio was tested and set at a comfortable listening volume individually for each participant. In order to have a measure sensitive to misperceptions, in addition to the two "true" and "false" options, participants were provided with a separate key (the space bar) to indicate when they failed to understand an utterance (don't understand or DU responses). Furthermore, with the motivation of keeping DU responses as pure as possible, participants were instructed to guess between "true" or "false" if they managed to understand the sentence but were not sure of the correct answer. Assignment of the true and false responses, informed at the beginning of the task, was counterbalanced to the "z" and "/" keys so that half of the participants provided true responses with their dominant hand and half with their nondominant hand. Response keys were signaled with red illumination. To increase RT sensitivity, participants were told to keep their index fingers resting on these two keys during the task and move them to the DU button as needed. Two self-administered breaks were included after blocks one and two. During this time, which never lasted more than a couple of minutes, participants did not speak to the experimenter.

Recordings were set to play 500 ms after each response or, in case no response was given, 5000 ms after utterance offset. Responses were possible at utterance onset, in line with cascaded theories of speech comprehension (Marslen-Wilson, 1987).

After the aSVT, a manipulation check similar to the speaker rating pilot study was conducted. Participants were presented with a sample of each speaker in English and Dutch and asked to rate their accent (on a scale from 1-native to 9-very strong foreign accent). Participants were asked to guess the regional accent of speakers thought to be native and the native language of speakers rated as non-native.

In addition, participants received a list of all the critical English sentences they had heard during the task in order to indicate words they did not know. Items with unknown words were later removed from analyses on an individual basis. This precaution was taken in order to ensure DU responses reflected failures to understand due to speech processing errors and not to unknown words. Then, participants completed the LBQ in the format of an online survey and the LexTALE, administered in PsychoPy (Peirce, 2007). At the end of the session, participants were debriefed and paid for their collaboration. In all, the session lasted between 30 and 60 minutes.

Results

Of the original 49 participants tested, four were immediately discarded based on the following criteria: technical difficulties during the aSVT which prevented their responses from being recorded (1), having participated in one of the pilot studies (1), having been raised bilingually as indicated on the LBQ (Dutch-German: 1; Dutch-English: 1). The remaining participants were from all over the Netherlands, although nearly half were from the province of Gelderland and all had been living in Gelderland for at least a year (M = 12.0 years, SD = 9.1, range = 1-28). Thus, they all had had exposure to the regional accent of Dutch. When specifically inquired about this, 87.5% indicated being familiar with the regional accent of Dutch spoken by the native Dutch speakers in the aSVT. In terms of native English accents, 78% reported an American accent as one of the accents they are most familiar with (it should be noted that British English was also selected by 75% of participants). Furthermore, 70% responded that they were familiar with the accent of English spoken by the native English speakers during the aSVT. Familiarity with the non-native accents was not as comparable, as can be seen in Table 5, with participants hearing Dutch-accented English more often and from more speakers than English-accented Dutch. This is also apparent from the greater difficulty they had in identifying the native language of the native English speakers in Dutch than of the native Dutch speakers when speaking English.

Since the fact that the study involved English was not mentioned during recruitment and the only indication that participants were proficient in English before the experiment was the question on the SONA prescreening questionnaire, "Do you speak English?," some proficiency selection criteria were considered necessary to ensure that any effects found would not be due to differences in English proficiency. To this end, a general English proficiency score was calculated for each

Table 5.

Familiarity with non-native accents by percent of sample.

	Dutch-accented English	**English-accented Dutch**
Frequency heard		
never	2.5	47.5
less than once a week	45	45
once a week	25	7.5
several times a week	27.5	0
Number of speakers heard from		
0-1	10	80
2-5	50	15
6-10	20	5
>10	20	0

participant by averaging self-ratings across the four skills (speaking, listening, writing, and reading). Two participants were removed from further analyses for not providing information about their English language skills and two more for having a general proficiency score of ≤ 3 (out of 7). The previous preprocessing steps resulted in an unequal number of participants per experimental list: 11 for list 4 and 10 for each of the other lists. Therefore, the last participant tested on list 4 was removed, yielding a final sample size of 40 participants (11 males; Age: $M = 23.0$, $SD = 3.1$, range = 18-31). No significant difference ($p > .1$) was found between the lists for any of the measures of English experience, except self-rated reading proficiency, $F(3, 36) = 3.314$, $p = .031$. A post-hoc test revealed that this difference was driven by a difference between lists 1 and 2, both Dutch base language groups, with participants of list 2 presenting greater proficiency than list 1. All participants noted English as their most fluent non-native language except for one who indicated being more proficient in German. The English experience of the final participants is summarized in Table 6.

The data were preprocessed in the following way. First, as explained before, English items containing words that participants indicated not knowing were removed for each participant. Remaining incorrect (true for false statements or false for true statements) responses were counted as errors. Items with ≤70% accuracy in the monolingual block were discarded, as this was the baseline of the experiment. Fourteen items (English: Four items spoken by a native, three by a non-native, Dutch: three by a native, four by a non-native) were removed based on this criterion. The final data set was composed of: 138 errors (4.08% of the observations), 122 DUs (3.61%), and two missing responses (0.06%).

RTs (measured from sentence onset) were processed by first subtracting the duration of the corresponding audio stimuli to adjust them to sentence offset in order to control for differences in sentence duration. RTs for errors and DUs were removed from subsequent RT analyses. Only a participant-based criterion was used for outlier detection. To compensate, a stricter threshold of 3 *SD*s was employed. RTs were considered outliers if they deviated more than 3 *SD*s from factor mean, which factor being determined by language status (base vs. guest), actual language of the item, and context (monolingual vs. bilingual). This resulted in the exclusion of 39 RTs (1.14% of the RT data).

For reasons explained before, here by-item analyses were preferred over by-participant analyses with the main variable of interest being between-

Table 6.

English experience for final participant sample by list.

	List									
	1 base language Dutch		2 base language Dutch		3 base language English		4 base language English		Overall	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age of acquisition	10.00	2.58	10.20	1.81	10.90	3.63	11.00	2.21	10.53	2.58
Frequency of use	4.10	1.10	5.00	0.67	4.60	0.70	4.60	0.84	4.58	0.87
LexTALE score	78.88	11.26	82.38	10.73	82.00	9.99	77.13	13.88	80.09	11.32
Word knowledge	98.28	3.72	96.07	7.03	97.89	3.49	96.49	5.73	97.18	5.09
Self-rated English proficiency										
Speaking	4.60	1.35	5.60	0.70	5.30	1.06	5.30	1.16	5.20	1.11
Listening	5.40	1.17	6.00	0.82	5.60	0.84	5.70	0.48	5.68	0.86
Writing	4.70	0.82	5.60	0.97	5.40	1.07	5.00	0.82	5.18	0.96
Reading	5.70	0.67	6.70	0.48	6.10	0.99	6.00	0.67	6.13	0.79
General *(average across skill)*	5.10	0.94	5.98	0.67	5.60	0.83	5.50	0.68	5.54	0.82

Note. Percent of English items with all words known was calculated after removal of items with > 30% errors.

participant. This means that performance on the same item (sentence) produced by the same speaker (and, thus, with the same accent) was compared when it occurred in each of the different critical conditions: in the base or guest language of the experiment and in a monolingual or bilingual context. Considering the impossibility of having guest language items in a monolingual block, the distribution of the factors context and language status were uneven: Base language items could occur in the monolingual and bilingual context, but guest language items only occurred in the bilingual context. Because of this, these two factors were combined into one of three levels which we called "condition." For simplicity's sake, the three levels will from here on be referred to as base monolingual (base language items in the monolingual context), base bilingual (base language items in the bilingual context), and guest language (guest language items necessarily in the bilingual context). Differences between base monolingual and base bilingual items give the effect of context and those between base bilingual and guest language items give the effect of language status. Therefore, repeated measure ANOVAs were run with condition (base monolingual vs. base bilingual vs. guest) as a within-item independent variable and language (Dutch and English) and accent (native vs. non-native) as between-item independent variables and error rate, DU rate, and RT as dependent variables.

On average, participants made 4.10% errors (SD = 3.36) and responded DU 4.96% (SD = 5.17) of the time. RTs averaged 788.19 (SD = 223.18) across all variables. Similarly, items averaged 4.09% errors (SD = 5.59), 3.63% DUs (SD = 7.18), and RTs of 805.35 ms (SD = 223.60). Averages of error rates, DU rates, and RTs by item per factor can be found in Table 7.

An ANOVA on DU rates with condition, language, and accent as independent variables yielded a main effect of condition, $F(2, 220) = 9.050$, $p = .001$, $\eta^2 = .07$. Planned comparisons revealed a difference between the base monolingual ($p = .003$) and bilingual ($p = .001$) conditions, on the one hand, and the guest language condition, on the other. Furthermore, a significant interaction between language and accent on DU rates was also observed, $F(1, 110) = 4.718$, $p = .032$, $\eta^2 = .04$. With planned comparison it was possible to see that this was due to a difference between native and non-native Dutch, $p = .012$. No other significant interactions were observed.

An ANOVA on RTs revealed a main effect of language, $F(1, 110) = 13.476$, $p < .001$, $\eta 2 = .99$, with Dutch sentences being processed faster than English ones. No significant main effect of condition was observed for RT. However, there was also an interaction between condition and language, $F(2, 220)$[3] $= 3.298$, $p = .029$, $\eta^2 = .03$. Planned comparisons indicated that this was due to a difference between the guest condition and the base monolingual and bilingual conditions, but only for Dutch (monolingual-bilingual: $p = .035$. guest-bilingual: $p = .004$; see Table 7 for all values). However, across languages the tendency ($p = .076$) was for the guest language condition to be slower than both monolingual and bilingual conditions. However, as will be seen below, a look at each language individually revealed a different pattern of results.

An ANOVA on error rates revealed no significant main effect nor interactions (all p values $> .1$). Error rates were included in all of the subsequent analyses but consistently yielded no significant effects. Therefore, they will not be discussed further.

Analysis of the First Guest Language Item

Although much effort was made to make the guest language unexpected, its surprise value probably largely wore off once it began to appear regularly. It follows that the first guest language item was inherently different from the rest and the trial where we thought LWL was most likely to occur. Because of this, the first guest language item was inspected separately. During the construction of the experimental list we made sure that the first guest language item occurred in all of the critical accent-language combinations: native accent in Dutch, non-native accent in Dutch, native accent in English, and non-native accent in English. In Table 8 you will find a summary for these first items in all three conditions. An inspection of DU frequencies revealed that the native English, non-native English, and non-native Dutch first guest language items had the highest DU rates of all sentences in any condition. This was in stark contrast to the native Dutch condition with 0% DU rate for the first guest language item. A chi squared test revealed that there was an association between these groups and DU responses, $\chi^2(3, N = 39) = 17.598$, $p = .001$ (Likelihood ratio). Follow-up analyses (with α Bonferroni-adjusted to .008 to account for the number of comparisons) indicated that there was a significant difference between native

3 Uncorrected degrees of freedom are reported here for aesthetic reasons. However, in actual analyses, degrees of freedom were corrected for violated assumptions.

Table 7.

Average DU rates, error rates, and RTs per variable before and after removal of the first guest language item.

	Variable													
	Language				**Accent**				**Condition**					
	English		**Dutch**		**Native**		**Non-native**		**Monolingual**		**Bilingual**		**Guest**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
Before first guest language item removal														
Error rate (%)	4.6	6.1	3.6	5.1	4.6	5.0	4.1	6.2	3.8	6.0	4.1	8.7	4.4	9.6
DU rate (%)	3.6	6.0	3.6	8.2	3.6	6.0	4.6	8.1	2.5	6.6	2.2	7.5	6.2	13.3
RT	884	236	726	180	884	234	799	214	790	258	798	276	839	304
After first guest language item removal														
Error rate (%)	4.5	6.1	3.7	5.1	4.5	4.9	4.1	6.3	3.7	6.0	4.1	8.9	4.4	9.7
DU rate (%)	3.0	5.0	3.1	7.1	3.0	5.3	3.8	6.8	2.4	6.5	2.0	7.4	4.7	9.5
RT	879	238	728	182	879	238	797	212	792	258	796	276	831	299

Dutch and non-native Dutch, $\chi^2(1, N = 20) = 10.208$, $p = .001$ (Continuity correction). The difference between native Dutch and native English was also notable, but not significant at the adjusted α, $\chi^2(1, N = 19) = 4.947$, $p = .026$ (Continuity correction). Analysis of RTs was not possible for all conditions since, due to the large number of DU responses, insufficient data points remained for RTs (only 2-5 for all conditions except native Dutch). An ANOVA for the native Dutch conditions showed that there was a main effect of condition, $F(2, 29) = 11.933$, $p < .001$, $\eta^2 = .88$, with Tukey post-hoc tests revealing that RTs were slower for the first guest language item than the same item in the monolingual ($p < .001$) and bilingual conditions ($p = .014$), while there was no significant difference between these last two ($p = .195$).

Analysis Without First Guest Language Item

Given the relatively high DU rates overall for the first guest language items, it is possible that these items were mainly responsible for the effects observed. What is more, these rates indicate that first guest language items are very different from the rest of the items and thus may involve different processes. Therefore, it was considered prudent to re-conduct the analyses without these sentences to ensure the condition effects reported above remained intact and did not only occur for the first unexpected item (see averages in Table 7) ANOVAs without the first guest language items and their counterparts in the other two conditions yielded, once again, a main effect of language on RTs, $F(1, 106) = 13.105$, $p < .001$, $\eta^2 = .11$, with Dutch sentences being processed faster ($M = 742.76$, $SE = 31.16$) than English ones ($M = 902.26$, $SE = 31.16$). The main effect of condition on DU rates observed before was also found here, $F(2, 212) = 6.030$, $p = .004$, $\eta^2 = .05$. Planned comparisons showed a significant difference between base monolingual ($p = .014$) and base bilingual conditions ($p = .003$), on the one hand, and guest language, on the other, while the difference between base monolingual and bilingual conditions was not significant ($p = .575$). Therefore, the main effects observed with the first guest language item included remained significant. However, the interactions between condition and language for RTs and language and accent for DU rates were only marginally significant here ($p = .078$

Table 8.

Comparison DU rates, error rates, and RTs for first guest language sentences and same sentences in other conditions.

	Condition											
	Base monolingual				Base bilingual				Guest			
	DU (%)	Err (%)	M RT *(SD)*	*N*	DU (%)	Err (%)	M RT *(SD)*	*N*	DU (%)	Err (%)	M RT *(SD)*	*N*
Dutch												
Native	0	0	381 (166)	10	0	0	686 (394)	10	0	0	1206 (505)	10
Nonnative	20	0	699 (390)	10	10	0	483 (314)	10	80	0	2157 (1827)	10
English												
Native	0	11	709 (157)	9	20	10	1217 (507)	10	56	11	863 (112)	9
Nonnative	0	10	1158 (826)	10	0	0	1037 (423)	9	50	10	1742 (417)	10

Note. Err = error, *N* = total cases.

and p = .066, respectively). No other significant interactions were observed.

Given their volatility, in order to further test the reliability of these interactions, an additional analysis was run after removing items with a small cell size (< 5 out of a maximum of 10) to ensure these were not biasing the results. This led to the exclusion of three items, with a total of 53 and 54 items remaining for English and Dutch, respectively. Following analyses, the main effects of condition and language remained. However, the interactions between language and condition, on the one hand, and language and accent, on the other, were not significant (p = .087 and p = .092, respectively).

Discussion

The aim of the present study was two-fold: On the one hand, given the scarcity of the literature on the topic, we wanted to see if it was possible to induce LWL states in a laboratory setting. To this end, we came up with a novel design that would bias bilinguals towards expecting one or the other of their languages, called the base language of the experiment. Our inclusion of a "don't understand" response option allowed us to more precisely measure comprehension and, our main interest, failures to comprehend.

Our second aim was to evaluate the role different factors play in the occurrence of LWL. In particular, we thought speech misperceptions would occur differentially in the bilingual's two languages and that they would be augmented when listening to the speech of a non-native speaker.

Analysis of the first guest language item revealed that the manipulation was indeed successful. After a monolingual block in the base language, participants were surprised with an item in their other language, the guest language of the experiment. Failures to comprehend, as indexed by DU rates, those first guest language items were the greatest in the experiment.

First of all, this study brings to light the importance of the sensitivity of your measurement tool. Specifically, many studies in cognitive science, especially those using button presses where RT is vital, do not provide participants with an option to indicate insecurity. This may lead to inaccurate responses and RTs, which actually reflect misunderstandings. This problem is even more serious in studies on non-native speech comprehension, where comprehension is more taxed. Here we hope to have demonstrated the value of data commonly piled together with other error rates.

One concern when deciding to include the DU button was that participants would use it as a "don't know" option when they did actually understand the utterance, but did not know the answer. Several precautions were taken in order to ensure this did not happen, such as telling participants to guess in case they understood the answer but did not know whether the statement was true or false. In addition, items with words participant were not familiar with (in English) were removed. The strongest evidence against this explanation, however, stems from the fact that DU rates, for the same item, were higher when that item occurred in the guest language condition. This, together with the observation that the final data do not reveal a difference in DU rates for English and Dutch, is strong support for the claim that misperceptions cannot be (entirely) attributed to a lack of knowledge or low L2 proficiency.

On the other hand, DU rates were not null in the other conditions. Furthermore, as a result of our manipulation check, we know that only 70% of participants claimed to have experienced a LWL situation during the experiment. While this could be viewed as a high success rate for a new experimental manipulation, it begs the question: can we be sure that the failures to comprehend that we observed were really due to LWL or could they just be explained as a language switching cost? Indeed, it may be difficult to disentangle these two concepts, and that is because LWL is a form of language switching. It could be defined as a failure or delay in speech comprehension due to non-target language processing. Effectively, resolution of a LWL state requires a language switch in speech perception mechanisms. The effect of condition found here was, by design, caused by the participant having to change from the base language of the experiment to the less frequent guest language. This does not mean that all DU responses were necessarily caused by LWL, but the fact that these rates were higher for the guest language condition suggests that making participants respond to an item in a different language than the previous one increases the likelihood of a comprehension failure. Of course, not unlike many other psycholinguistic processes, the occurrence of LWL does not presuppose consciousness. In fact, people misperceive speech in the same language all the time, without necessarily knowing why. Still, stronger evidence of LWL could be found with other designs, for example using target competitors, to demonstrate activation of the non-target language.

Returning to our predictions, we expected that guest language items would result in more DUs and slower RTs. As concerns the first, guest language items did produce more comprehension failures than base language items and this effect did not differ for the L1 or L2, nor for native vs. non-native speech[4]. While prior studies have demonstrated increased difficulties in processing gated guest words (e.g., Grosjean, 1988; Schulpen et al., 2003) or intra-sentential code-switches (Fitzpatrick & Indefrey, 2014), we are not aware of any previous evidence of complete breakdowns in comprehension during bilingual sentence comprehension. Methodologically, the fact that DU rates were observed at all is promising for future studies on LWL or similar bilingual speech misperceptions. Theoretically, this suggests that bilingual speech comprehension can at least partially proceed in a language-selective manner, for example, when strongly biased by the context, both local (previous item) and global (entire experiment). While the present study cannot be said to support any particular model of bilingual speech comprehension, since they all concern word recognition, the present findings could hint at the fact that comprehension proceeds in a more language-selective fashion during sentence processing, as has been suggested by some researchers (Fitzpatrick & Indefrey, 2014; Lagrou et al., 2012; Li, 1996).

Interestingly, no significant difference was observed between the monolingual and bilingual base language conditions. Studies on non-selective lexical access have been finding a cost for the mere activation of the non-target language since seminal studies (e.g., Kolers, 1966; Macnamara & Kushnir, 1971) showing a processing cost for reading mixed language passages compared to monolingual passages. However, RTs to sentences, measured here, may not have been precise enough to reflect these differences. In addition, as mentioned before, the comparison made here is not ideal as the monolingual block always preceded the bilingual blocks. Further study would be needed to confirm these observations.

Regarding the second part of our prediction on guest language processing costs, we found that DU rates were not accompanied by RT differences. One possible explanation is that DUs and RTs reflect the use of different strategies, with participants using the DU option when deciphering utterances proves too daunting. However, analyses of RTs to DU responses seem to suggest that participants do try to understand these utterances before "giving up," taking, numerically, on average longer than for correct responses (DU: M = 1480.89, SD = 776.16; correct responses: M = 790.55, SD = 516.13).

Another possible explanation for the lack of relationship between DU rates and RTs has to do with the nature of the task, which was meant to be challenging. With short words and sentences, and utterances presented shortly after responses were given, the task was designed to make sure participants had to tune in to the utterances quickly. In cases where LWL has to do with a speech segmentation problem, taking longer to process the utterance might not present a real benefit: once a part of the utterance was misperceived there was little chance of recovering it (except, perhaps, via a phonological loop mechanism).

4 It should be noted that these were still rare, as suggested by our initial survey on LWL incidence.

In addition to inducing speech misperceptions, we were interested in examining the role of two factors on its occurrence: language and accent. As regards language, we expected that whether the guest language was the L1 or L2 would influence the incidence of speech misperceptions. In terms of the direction, we equally envisioned the two possible directions. Findings of a reduced baseline activation for the L2 would seem to predict a greater incidence of DU responses for items in this language, in line with Schulpen et al. (2003)'s study. On the other hand, studies demonstrating an asymmetric switch in bilingual speech production would have one believe that during L2 processing, the L1 is heavily inhibited, which could manifest itself as a greater processing cost for L1 guest language items.

Regarding our second variable of interest, accent, we thought that guest language utterances produced with non-target language pronunciation would hinder comprehension by tilting the system towards the non-target language. Support for this idea comes from gating studies showing that guest words pronounced with guest language phonetics are identified faster than those pronounced in base language phonetics (Grosjean, 1988; Li, 1996). However, there is also evidence that sentential context can help reduce the amount of non-target language interference during auditory comprehension in bilinguals (see Fitzpatrick & Indefrey, 2014 for a review).

In line with previous findings, participants were faster to respond in their L1 than in their L2 (Proverbio, Leoni, & Zani, 2004; Schulpen et al., 2003). Furthermore, analysis of the first guest language item revealed that not all conditions were equally surprising. Listeners never failed to understand utterances produced by native speakers of their L1, although RTs revealed a delay in comprehension relative to the same item in the base language. This was in stark contrast to non-native speech in Dutch, despite it being participants' L1. Speaker accent did not modulate comprehension of speech in the L2. These differences are likely due to familiarity with the accents, with Dutch speakers being less exposed to non-native Dutch than native Dutch and both native and non-native English. Some studies have found that previous experience with a particular accent can increase comprehension of that accent (Witteman et al., 2013).

When the first guest language item was excluded from the analyses, language and accent no longer played a role. This suggests that language and accent may only initially play a role, when other information is not available. These findings are in line with studies showing that bilinguals are able to quickly adapt to unfamiliar accents (Witteman et al., 2013). A similar effect was observed in the study on the role of facial cues on bilingual production (Woumans et al., 2015). In that study, prior to the task, bilinguals interacted with speakers in one of their languages. Later, during a noun-verb association task, speakers' faces were presented while producing noun stimuli to elicit participant responses. Crucially, speech could be produced in the same language spoken by that speaker before (congruent trials) or a different language (incongruent trials). The results revealed a difference between congruent and incongruent trials for the first six trials only, with slower RTs observed for incongruent trials. This difference was not evident, however, in later trials. The authors interpreted these findings by proposing that faces are used as cues for language production as long as they are considered reliable. If expectations are violated, however, the association between speaker face and language can be weakened. These findings are also consistent with Elston-Güttler et al. (2005)'s finding that priming effects from prior exposure to a film in one language decreased throughout the experiment.

Extended to the present study, a claim could be made for the surprise effect of guest language productions decreasing after this language made its first appearance in the experiment. Rather than completely dissipating, however, a guest language processing cost remained, although this no longer differed for the L1 and L2 nor for utterances produced by native and non-native speakers. Nonetheless, given the small number of items available for the first guest language analysis, additional studies are necessary with a great deal more items and participants, in order to properly test for an effect of an entirely unexpected guest language, as well as a difference between early and later trials in that guest language.

Analyses without the first guest language item also showed that, after the first guest language item, accent did not exacerbate the effect of the guest language. This lends support to the claim that phonetic information may be more relevant when processing isolated words than sentences, where other information can help decipher meaning. These results are similar to those of Fitzpatrick and Indefrey (2014) who reported symmetrical switch costs in ERPs. However, more conclusive evidence could be provided by future studies, particularly where the monolingual block could be counterbalanced and allowing for within-participant analyses and an equal amount of switching into the L1 and into the L2.

It should be noted that, despite not finding evidence for the role of these factors here, it is likely that LWL is so rare that the roles of these factors are hard to assess. In fact, this is why much psycholinguistic research relies on types of stimuli that are not very common in everyday life, like interlingual homophones. Although we did not find evidence for a role of language or accent in guest-language induced speech misperceptions, we do not deny the possibility that these aspects, and others, can modulate the occurrence of LWL outside the laboratory. Here we have developed a paradigm that has proven effective in inducing speech misperceptions. Further studies can explore ways to increase the likelihood of LWL states in an experimental setting, such as increasing cognitive load or adverse listening conditions.

Here we were interested in a rare failure in comprehension that occurs in bilinguals. Misperceptions are valuable to research in that, by highlighting what can go wrong, they can provide insight into the processes underlying normal speech comprehension. However, it goes without saying that, outside of the laboratory, in the real world, they are probably a lot less detrimental than research would lead one to believe. Indeed, in everyday life, bilinguals manage to communicate and code-switch without major problems.

References

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical data base on (CD-ROM).* Philadelphia, PA: Linguistic Data Consortium.

Beck, C., Kardatzki, B., & Ethofer, T. (2014). Mondegreens and Soramimi as a method to induce misperceptions of speech content - Influence of familiarity, wittiness, and language competence. *PLoS ONE, 9*(1).

Beck Lidén, C., Krüger, O., Schwarz, L., Erb, M., Kardatzki, B., Scheffler, K., … & Ethofer, T. (2016). Neurobiology of knowledge and misperception of lyrics. *NeuroImage, 134,* 12–21.

Bond, Z. S. (2008). Slips of the Ear. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 290–310).

Brodkey, D. (1972). Dictation as a measure of mutual intelligibility: A pilot study. *Language Learning, 22*(2), 203–217.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera 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.

Bürki-Cohen, J., Grosjean, F., & Miller, J. L. (1989). Base-language effects on word identification in bilingual speech: Evidence from categorical perception experiments. *Language and Speech, 32*(4), 355–371.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1029–1040.

Cheng, Y.-L., & Howard, D. (2008). The time cost of mixed-language processing: An investigation. *International Journal of Bilingualism, 12*(98568), 209–222.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 136–163). Hillsdale, New Jersey: Erlbaum.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 240–247.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review, 11*(6), 1084–1089.

Cox, E. A. (2005). *Second language perception of accented speech.* (Doctoral dissertation, University of Arizona, 2005). Retrieved from Digital Dissertations, (UMI No. 3177525).

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

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*(4), 385–400.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 19*(1), 1–16.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15*(4), 850–855.

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research, 25*(1), 57–70.

Fitzpatrick, I., & Indefrey, P. (2010). Lexical competition in nonnative speech comprehension. *Journal of Cognitive Neuroscience, 22*(6), 1165–1178.

Fitzpatrick, I., & Indefrey, P. (2014). Head start for target language in bilingual listening. *Brain Research, 1542*, 111–130.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America, 9*(1), 370–389.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1276–1293.

Gooskens, C., & Heeringa, W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using

Norwegian dialect data. *Language Variation and Change, 16*(3), 189–207.
Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language, 122*(2), 92–102.
Gough, P. B. (1966). The verification of sentences: The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior, 5*(5), 492–496.
Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21*(2), 344–359.
Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes, 3*(3), 233–274.
Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3–15.
Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspecitives* (pp. 225–254). Mahwah, New Jersey: Lawrence Erlbaum Publishers.
Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science, 5*(4), 201–202.
Hartsuiker, R. J. (2015). Visual cues for language selection in bilinguals. In R. Kumar Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and Vision in Language Processing* (pp. 129–145). New Delhi: Springer India.
Hendriks, P. (2014). *Asymmetries between Language Production and Comprehension*. Dordrecht: Springer.
Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods, 46,* 702–721.
Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science, 15*(5), 314–318.
Kentner, G. (2015). Rhythmic segmentation in auditory illusions- evidence from cross-linguistic Mondegreens. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 0–4.
Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643–650.
Koch, I., Prinz, W., & Allport, A. (2005). Involuntary retrieval in alphabet-arithmetic tasks: Task-mixing and task-switching costs. *Psychological Research, 69*(4), 252-261.
Kolers, P. A. (1966). Reading and talking bilingually. *American Journal of Psychology, 79*, 357–376.
Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2012). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition, 16*(2013), 1–10.
Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343.
Léwy, N. (2015). *Computational psycholinguistics and spoken word recognition in the bilingual and the monolingual.* Université de Neuchâtel.
Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language, 35*(6), 757–774.
Li, P. (1998). Mental control, language tags, and language nodes in bilingual lexical processing. *Bilingualism: Language and Cognition, 1*(2), 92–93.
Li, P., & Farkaš, I. (2002). A self-organizing connectionalist model of bilingual processing. In R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 59–85). North Holland: Elsevier Science Publisher.
Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior, 10*(5), 480–487.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71–102.
McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*(1), 21–46.
McQueen, J. M., & Cutler, A. (2010). Cognitive processes in speech perception. In W. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 489–520). Wiley-Blackwell.
Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*, 25–40.
Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 193–218).
Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289–306.
Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The Role of Speaking Rate. *Studies in Second Language Acquisition, 23*(4), 451–468.
Nelson, K. E., & Kosslyn, S. M. (1975). Semantic retrieval in children and adults. *Developmental Psychology, 11*(OCTOBER), 807–813.
Otake, T. (2007, August). Interlingual near homophonic words and phrases in L2 listening: Evidence from misheard song lyrics. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, 777–780.
Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science, 12*(6), 445–449.
Pecher, D., De Rooij, J., & Zeelenberg, R. (2009). Does a pear growl? Interference from semantic properties of orthographic neighbors. *Memory & Cognition, 37*(5),

541–546.
Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.
Piske, T., Mackay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191–215.
Proverbio, A. M., Leoni, G., & Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia, 42*(12), 1636–1656.
Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20*, 331–350.
Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia, 85*, 245–255.
Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2015). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In *Interfaces in second language pronunciation assessment: Interdisciplinary perspectives.*
Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition, 15*(01), 157–166.
Schepens, J., Dijkstra, T., Grootjen, F., & Van Heuven, W. J. B. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLoS ONE, 8*(5).
Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1155–1178.
Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism: Language and Cognition, 16*(2), 304–324.
Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition, 12*(4), 380–386.
Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281–284.
Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America, 126*(1), 367.
Van Der Haagen, M. J. (1998). *Caught between norms: The English pronunciation of Dutch learners.* The Hague: Holland Academic Graphics.
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1–25.
Weissberger, G. H., Wierenga, C. E., Bondi, M. W., & Gollan, T. H. (2012). Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging, 27*(4), 959–974.
Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics, 75*(3), 537–556.
Woumans, E., Martin, C. D., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a language? *Psychological Science, 26*(9), 1343–1352.
Wright, S. (1954). The death of Lady Mondegreen. *Harper's Magazine, 209*(1254), 48–51.
Yule, H., & Burnell, A. (1903). *Hobson-Jobson: a glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive.* London: John Murray.
Zhao, X., & Li, P. (2007). Bilingual lexical representation in a self-organizing neural network model. *Proceedings of the 29th Meeting of the Cognitive Science Society, 1*, 755–760.
Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13*(5), 505–524.