

Published in final edited form as:

J Phon. 2011 January 1; 39(1): 1–17. doi:10.1016/j.wocn.2010.10.006.

Word Durations in Non-Native English

Rachel E. Baker^a, Melissa Baese-Berk^b, Laurent Bonnasse-Gahot^c, Midam Kim^b, Kristin J. Van Engen^b, and Ann R. Bradlow^b

^aEF Education First 22 Chelsea Manor Street London SW3 5RL England, U.K.

^bNorthwestern University Department of Linguistics 2016 Sheridan Road Evanston, IL 60208-4090 USA

^cCentre d'Analyse et de Mathématique Sociales (CAMS, UMR 8557 EHESS-CNRS) Ecole des Hautes Etudes en Sciences Sociales 54, bd Raspail F-75270 Paris Cedex 06 France

Abstract

In this study, we compare the effects of English lexical features on word duration for native and non-native English speakers and for non-native speakers with different L1s and a range of L2 experience. We also examine whether non-native word durations lead to judgments of a stronger foreign accent. We measured word durations in English paragraphs read by 12 American English (AE), 20 Korean, and 20 Chinese speakers. We also had AE listeners rate the 'accentedness' of these non-native speakers. AE speech had shorter durations, greater within-speaker word duration variance, greater reduction of function words, and less between-speaker variance than non-native speech. However, both AE and non-native speakers showed sensitivity to lexical predictability by reducing second mentions and high frequency words. Non-native speakers with more native-like word durations, greater within-speaker word duration variance, and greater function word reduction were perceived as less accented. Overall, these findings identify word duration as an important and complex feature of foreign-accented English.

Keywords

second language acquisition; foreign accent; lexical frequency effects; second mention reduction; function word reduction; Korean; Chinese

1. Introduction

Spoken word durations can vary dramatically depending on several lexical features aside from number of phonological segments or syllables. These may be features of words in the lexicon (e.g. frequency and part of speech), or in the discourse (e.g. discourse status). The effects of these features on word duration have been thoroughly examined for native English

© 2010 Elsevier Ltd. All rights reserved

Corresponding Author Rachel Baker: Rachel.Baker@ef.com Phone +44 207 341 8580 Fax +44 207 341 8501.

Melissa Baese-Berk: m-baese@northwestern.edu

Midam Kim: midam-kim@northwestern.edu

Kristin J. Van Engen: k-van@northwestern.edu

Ann R. Bradlow: abradlow@northwestern.edu

Laurent Bonnasse-Gahot: lbg@ehess.fr

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

speakers, providing valuable insights into these speakers' psycholinguistic processes (e.g. Anderson & Howarth, 2002; Aylett & Turk, 2004; Baker & Bradlow, 2009; Bard et al., 2000; Bell et al., 2002; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Fowler & Housum, 1987; Gahl, 2008; Jurafsky, Bell, Gregory, & Raymond, 2001; Shi, Gick, Kanwisher, & Wilson, 2005). However, very little is known about how these features affect word duration in the speech of language learners.

The effects of lexical features on word duration may differ in native and non-native speech for a number of reasons. One possibility is that a non-native speaker's first language (L1) differs from their second language (L2) in the effects of these lexical features. For example, a speaker's L2 may have a large set of function words which are frequently reduced. Their L1 may be an agglutinative language which uses affixes in place of function words, and therefore does not have a class of words that are typically reduced. As a result, this speaker may not reduce function words in their L2. Another possibility is that the added cognitive demands associated with speaking a second language mean that these subtle effects are not realized in non-native speech. A third possibility is that non-native speakers have such a different experience with their second language than native speakers that lexical features (such as word frequency) have different effects on their word durations. For example, non-native English speakers who work in America but speak their native language at home may hear English words related to cooking or child-care with a lower frequency than native English speakers. Research on how non-native speakers differ from native speakers in this respect can provide information about language learners' mental lexicons and their abilities to process discourse. Of course, if word-level duration affects how accented a non-native speaker sounds to native listeners, then it becomes especially important to determine what differences exist between native and non-native speakers, and which ones influence native perception the most. Such information could be used to help English learners develop more native-like accents, thereby protecting them from the negative stereotypes, discrimination, and reduced employment opportunities that can face English speakers with non-native accents (e.g. Munro, 2003).

In addition to differing from native English speakers in terms of word duration, non-native speakers may also differ from each other. Such durational diversity among English learners could arise because they have different L1s, as well as different levels of experience, proficiency and fluency in their L2. The three possibilities laid out in the preceding paragraph also make predictions about whether non-native speakers should differ from each other in their word durations. If differences between native and non-native speakers are due to differences in the effects of lexical features between a non-native's L1 and L2, we would expect non-native speakers with the same L1 to behave similarly, and speakers with different L1s to behave differently. If the differences are due to added cognitive demands, we would expect similarities across speakers with different L1s, but differences across speakers with varying proficiency levels. If the differences are due to the unique experience that language learners have with their L2, we would also expect speakers with different L1s to behave similarly, but for differences between non-native speakers to arise based on their experience with their L2. For example, students studying different subjects might behave differently because they use different words in their daily life. Determining how much non-native speakers differ from one another (both within and across language backgrounds) can help researchers determine the extent to which they can apply their results to the general population of language learners. Research into whether non-native speakers with different L1s behave differently can also shed light on whether particular features of non-native speech are due to transfer from these speakers' L1s.

In this study, we examine how native English speakers and native Chinese and Korean speakers learning English differ in their word durations. We specifically explore three

issues: (1) differences in word-level durations and lexical effects on duration between native and non-native English speakers, (2) differences in word-level duration and lexical effects on duration across non-native speakers, and (3) the relationship between non-native word durations and the perceived accentedness of a non-native speaker. These three questions are discussed in detail below.

- (1) Are there differences between native and non-native English speakers in terms of word-level duration? If so, can these differences be explained by lexical features of English?

Non-native speakers' slower speech and reduced within-speaker durational variance are two of the strongest findings in research on non-native speech duration. Non-native English speakers, including Mandarin, Korean, and Italian learners of English, produce slower utterances than native English speakers (Guion, Flege, Liu, & Yeni-Komshina, 2000; Munro & Derwing, 1995). Non-native speech rate is influenced by age of acquisition (Guion et al., 2000), proficiency (Anderson-Hsieh & Horabail, 1994), and time spent in an English-speaking country (Lennon, 1990). Studies on syllable and vowel duration have also shown that non-native English speakers produce less within-speaker durational variance than native speakers. At the syllable level, low proficiency Chinese learners of English produced a smaller durational difference between stressed and unstressed syllables than native English speakers (Anderson-Hsieh & Horabail, 1994). In addition, non-proficient Japanese learners of English produced less syllable reduction than native English speakers as the number of syllables in a foot increased (Mochizuki-Sudo & Kiritani, 1991). At the vowel level, non-native speakers of English with five different language backgrounds (Farsi, Japanese, Spanish, Hausa, and Chinese) (Fokes & Bond, 1989), Korean learners of English (Lee, Guion, & Harada, 2006), and Spanish learners of English (Flege & Bohn, 1989; Shah, 2004) reduced the durational differences between stressed and unstressed vowels, relative to native English speakers. However, in a counterexample to these findings, Japanese speakers in Lee et al. (2006) produced native-like durational patterns for stressed and unstressed vowels. While the majority of these studies demonstrate reduced syllable and vowel duration variance in non-native speech, none examine word duration variance.

We are examining word durations because they allow us to explore the effects of lexical features on non-native speech production. We are specifically studying the effects of word frequency, previous mention, and word type on word duration. All of these effects have been demonstrated in native English speech, and some have been found in other languages. Some lexical effects on duration can be described as predictability effects; more predictable words tend to be phonetically reduced. Words can be predictable in the language as a whole (e.g. more frequent), or within a discourse (e.g. previously mentioned in the discourse). The related phenomenon of function word reduction is tied to function words' higher frequency relative to content words, but also to the unique role that function words play in language.

Higher frequency words tend to have shorter durations than lower frequency words in English, even when features such as number of phonemes have been controlled. Frequency effects on duration have been found in spontaneous speech by native American English speakers (Bell et al., 2002; Bell et al., 2009; Jurafsky et al., 2001) and native Glaswegian English speakers (Aylett & Turk, 2004). These effects have also been found in read speech by native American English speakers in both clear and plain speech styles (Baker & Bradlow, 2009). Studies involving a variety of predictability factors have shown that lexical frequency is the strongest or one of the strongest factors influencing word duration (Bell et al., 2002; Bell et al., 2009). Gahl (2008) demonstrated that frequency effects extend to more and less frequent homophones, proving that frequency effects are lexical effects, and not tied solely to the phonological form of a word. Less frequent words are more likely to receive a pitch accent¹, which may contribute to frequency effects on duration (Pan & McKeown,

1999). However, these effects have been shown to appear independently of accent status (Bell et al., 2002; Bell et al., 2009).

Languages other than English also show frequency effects on speech production. Dutch affixes attached to higher frequency words were reduced compared to the same affixes attached to lower frequency words (Pluymaekers, Ernestus, & Baayen, 2005). High-frequency Spanish words had shorter naming latencies than lower frequency words (Navarrete, Basagni, Alario, & Costa, 2006), although these researchers did not study word duration. Most relevant to our own study, higher frequency Cantonese words were produced in a reduced lexical tone space relative to lower frequency words (Zhao & Jurafsky, 2009).

To explore the accuracy of non-native perceptions of word frequency in English, Schmitt and Dunham (1999) asked native and non-native English speakers about their word frequency intuitions. Although native language did affect these intuitions, education was a stronger predictor of a person's success at this task. This suggests that non-native English speakers do have some knowledge of English word frequencies. However, it is unclear whether such intuitions translate into frequency effects in language production. As far as we know, no one has examined frequency effects on duration in non-native English.

In second mention reduction, speakers reduce words that are more predictable because they have already appeared in the discourse. When English speakers repeat a word in a discourse, its second mention tends to be shorter and less intelligible than its first mention (Aylett & Turk, 2004; Bell et al., 2009; Fowler & Housum, 1987). Such reduction is not simply articulatory priming, as it does not appear for words primed by a homophone, or for repeated words in word lists (Fowler, 1988). Words that are repeated in a word list differ from second mentions within a discourse because there is no expectation that a word will appear more than once in a list, but it is normal for a word to appear multiple times within a coherent discourse. Second mention reduction appears to be quite robust: it has been found when the first and second mentions are produced by different speakers (Anderson & Howarth, 2002), when the listener has changed between the first and second mentions (Bard et al., 2000), and in hyperarticulated clear speech (Baker & Bradlow, 2009).

Because first mentions of words are more likely to receive a pitch accent than second mentions (Baker & Bradlow, 2009; Hawkins & Warren, 1994), there is some disagreement over whether pitch accent placement is the only mechanism driving second mention reduction, or whether such reduction is at least partially gradient. Hawkins and Warren (1994) found that words with pitch accents were more intelligible than words without pitch accents, but the intelligibility of first mentions with pitch accents did not significantly differ from second mentions with pitch accents. They concluded that second mention reduction is due to second mentions being less likely to receive a pitch accent. In contrast, Bell et al. (2009) and Baker and Bradlow (2009) found second mention reduction effects on duration after controlling for pitch accent status. These studies provide evidence that second mention reduction is not completely dependent on differences in pitch accented placement. Such contradictory results may be due to the fact that Hawkins and Warren examined intelligibility, while the other researchers examined duration.

As with frequency effects, the majority of work on second mention reduction has focused on American and British English. Still, this phenomenon appears in other languages and dialects, even those with prosodic systems that differ from these dialects of English. The

¹English pitch accents are local intonational events (e.g. pitch peaks) associated with stressed syllables, which can be marked with longer durations and higher amplitudes. English pitch accents are often placed on new or important words or phrases (see discussion in Ladd, 1996).

Korean prosodic system does not include pitch accents (Jun, 1993), and Indian English is less likely to involve second mention deaccenting than American or British English (Gumperz, 1982). Despite these differences, second mentions are significantly shorter than first mentions in Korean and Indian English² speech (Baker & Bradlow, 2007). There is, however, a scarcity of research on second mention reduction in non-native English. Although Baker and Bradlow (2007) studied Indian English, Indian English speakers may differ from typical non-native English speakers because of the prominent role that English plays in the Indian education system, government, and media (Sailaja, 2009).

A number of causes have been proposed for predictability effects on duration, some applying only to frequency, and some applying more generally. Lindblom (1990) suggested that a speaker's desire to provide his listener with maximum information using minimal effort leads the speaker to hyper-articulate less predictable words, and reduce more predictable words. Aylett and Turk's (2004) Smooth Signal Redundancy Hypothesis builds on Lindblom's theory, but adds that speakers are trying to maintain smooth signal redundancy, or a roughly equal chance that each element will be understood. They claim that speakers use prosodic prominence to regulate smooth signal redundancy. Others have suggested more automatic causes, such as articulatory practice, speed of lexical access, and priming effects (Shields & Balota, 1991). Bybee (2001, 2006) suggests that frequency effects on duration are due to articulatory practice. Bell et al. (2009) instead describe lexical access as a likely source of predictability-related reduction effects. They claim that a mechanism which slows down articulation for words that take longer to access allows the speaker to coordinate lexical access and articulatory planning. Gahl (2008) points out that more predictable words might be shortened for any or all of these reasons, but exemplar models of speech (e.g. Bybee, 2001; Pierrehumbert, 2001, 2002) offer an explanation for the persistence of frequency effects. Exemplar models hold that words are represented by multiple examples of the word that the speaker has heard, complete with fine phonetic detail. More frequent words may have relatively shorter examples than less frequent words for reasons discussed above (e.g. they may be more predictable, speakers may have more articulatory practice producing them, or they may be accessed more quickly in the mental lexicon). Under an exemplar account, hearing so many short examples of frequent words would influence a speaker's mental representation of these words. A mental representation with multiple short examples could lead the speaker to produce the word with a shorter duration (Gahl, 2008).

Like predictability, word class can also influence duration. Function words in English tend to be shorter than content words, even after statistically controlling for the phonological forms of the words, their pitch accent status (accented or unaccented), and position in the phrase (Bell et al., 2009). This effect was found even after excluding the highest frequency words, which were mostly function words. Relative to content words with similar frequencies, function words have shorter and less intense vowels, and are more likely to have their final /t/s and /d/s dropped (Shi et al., 2005). These studies indicate that function word reduction is not simply a frequency effect. Function words are members of closed classes, have a low semantic load, and tend to be predictable in context (Shi, Morgan, & Allopenna, 1998). They are treated differently from content words within the prosodic system of English, in that they are often prosodic clitics, which are included in the same prosodic word with a content word, and are realized with a reduced phonological form (Selkirk, 1996). They also seldom receive pitch accents (Lavoie, 2002). In addition to these

²The Indian English speakers in this study reported that they were either native English speakers or learned English between 4 and 6. All of them had also learned at least one Indian language from birth (Hindi, Marathi, Telugu, Tamil, or Bengali). They had all grown up in India, and were recorded after moving to the U.S. to go to high school or university.

prosodic differences, function words may be accessed through different psychological processes than content words (Garrett, 1980; Lapointe & Dell, 1989).

Function word reduction has been demonstrated in languages other than English. Mandarin function word syllables were approximately half as long as content word syllables, and function words had smaller relative amplitudes than content words (Shi et al., 1998). Similar function word reduction effects were found for Turkish, a language with a rich morphological system that is typologically different from English and Mandarin (Shi et al., 1998). Dutch speakers produced function words with vowel qualities similar to unstressed syllables in content words (Van Bergen, 1993). Despite the reduction of function words in a variety of languages, there is evidence that at least some non-native speakers do not reduce function words like native speakers do. In particular, Japanese learners of English did not reduce English function words as much as native English speakers (Aoyama & Guion, 2007; Ueyama, 2000).

(2) Do non-native English speakers, with or without a shared language background, produce similar word-level durational features?

The studies discussed above provide some evidence for variability among non-native speakers, both within and across language backgrounds. Proficiency and experience with a non-native language affect speech rate and vowel duration variance in that language. Non-native speakers who acquired English early and who were more proficient in English spoke more quickly than other non-native speakers (Anderson-Hsieh & Horabail, 1994; Guion et al., 2000; Lennon, 1990). Advanced Chinese learners of English, like native English speakers, also had greater variance in their vowel durations than lower proficiency Chinese learners (Anderson-Hsieh & Horabail, 1994). Therefore, we expect to see similar differences within our set of non-native participants, even within language groups, as these language groups include speakers with different amounts of English experience.

We might see differences between non-native speakers with different L1s because non-native speakers speak an interlanguage, in which features (including prosodic features) are transferred from their L1 to their L2 (see reviews in Rasier & Hiligsmann, 2007, and Ueyama, 2000). While such transfer could lead to similar behaviors across speakers who share a first language, it is also possible that these speakers could transfer different features of their L1 to their L2, leading to diversity within a group of non-native speakers with the same L1. Unfortunately, it is difficult to assess the role of native language on speech sound durations produced by non-native speakers, because few studies examine multiple speakers from more than one language background, and even fewer directly compare speakers with different L1s. The results we do have are mixed. Shi et al. (1998) found similar English function word reduction for native speakers of Mandarin and Turkish. In contrast, Lee et al. (2006) found that native Japanese speakers produced native-like durations for stressed and unstressed English vowels while native Korean speakers did not.

In our study, we are somewhat limited in the predictions we can make about lexical effects in the speech of English learners with different L1s, because of incomplete information about these effects in Korean and Chinese dialects. Frequency effects on speech production have been found in Cantonese (Zhao & Jurafsky, 2009), and function word shortening has been found in Mandarin (Shi et al., 1998), but it is unknown how frequency and word type affect speech production in Korean. Similarly, second mention reduction has been found in Korean, but it is unknown whether it appears in any Chinese dialects. In the absence of further research, we have no reason to believe that Korean and Chinese dialects differ in the effects of predictability on speech production.

However, there are differences between these languages in their treatment of function words, and such differences could lead to differences in English function word production by Korean and Chinese speakers. Ueyama (2000) suggests that the native Japanese speakers in her study may have reduced English function words less than native English speakers because of transfer from Japanese. She explains that in Japanese, grammatical functions are typically performed by suffixes rather than independent function words, so Japanese speakers are not used to treating function words as prosodic clitics, which are phonologically reduced. Korean is morphologically similar to Japanese, in that both languages are agglutinative (Iwasaki, 2002; Lee & Ramsey, 2000). This means that both languages form words by attaching affixes to base words. Korean and Japanese, like English, have both function words and grammatical suffixes (Iwasaki, 2002; Lee & Ramsey, 2000). The independent function words in these languages include pronouns and conjunctions, and the grammatical suffixes include plural markers and derivational suffixes (Iwasaki, 2002; Lee & Ramsey, 2000). Both languages also make extensive use of grammatical particles, which fall between independent function words and affixes (Lee & Ramsey, 2000); these include conjunctions, and case, topic, and discourse markers (Iwasaki, 2002; Lee & Ramsey, 2000). In contrast to Korean and Japanese, Mandarin is an isolating language (Li & Thompson, 1989). This means that Mandarin words generally contain only one morpheme. Mandarin has a large range of independent function words, and has few grammatical affixes (Li & Thompson, 1989). Like English, Mandarin uses independent function word for pronouns, prepositions, conjunctions, and auxiliary verbs, as well as a variety of other functions (Li & Thompson, 1989). Mandarin speakers also durationally reduce function words relative to content words (Shi et al., 1998). If the role of function words in a speaker's L1 does influence the production of such words in their L2, then we would expect to see some differences between Korean and Chinese speakers' English function word durations. Specifically, we would expect native Korean speakers to produce less function word reduction than native Chinese speakers, because there are fewer independent function words in Korean, so Koreans might not be used to reducing such words.

(3) Are non-native word-level durational features in English associated with the perception of a stronger non-native accent?

Durational features have been shown to play an important role in non-native speech intelligibility, comprehensibility³, and accentedness. Adjusting the durations of segments to match those of a native English speaker significantly increased the intelligibility of a phrase spoken by a Chinese learner of English (Tajimi, Port, & Dalby, 1997). Similarly, adjusting the durations of segments spoken by a native English speaker to match those of a Chinese learner of English significantly decreased a phrase's intelligibility. Rhythmic errors were the greatest detriment to the intelligibility of Nigerian learners of English (Tiffin, 1992). When native English speakers were asked to list the factors they considered to be important when judging non-native speakers' accentedness and comprehensibility, 23% responded that prosodic features (such as rhythm and intonation) affected their judgments (Derwing & Munro, 1997). Prosodic features were listed as important factors more often than fluency, volume, or vocabulary, and for judgments of comprehensibility, prosodic features were listed more than segmentals. In the same study, native English speaking listeners rated the comprehensibility, accentedness, and intelligibility of non-native speech recordings, and the experimenters assigned prosodic ratings to low-pass filtered versions of these recordings. The prosodic ratings assigned by the experimenters were correlated with comprehensibility ratings for 35% of native listeners, with accentedness ratings for 27% of listeners, and with

³Intelligibility refers to how accurately a native listener can understand an utterance. Comprehensibility refers to a native listener's perception of the utterance's intelligibility, or how easy the utterance was to understand (Derwing & Munro, 1997).

speaker intelligibility for 8% of listeners. Similarly, Anderson-Hsieh, Johnson, and Koehler (1992) found that pronunciation ratings for non-native English speakers had a stronger correlation with prosody (as rated by ESL teachers) than with segments or syllable structure. Differences between Spanish-accented English and American English on word duration, unstressed vowel duration, and stressed-unstressed vowel duration ratios were related to native English speakers' perceptions of accentedness (Shah, 2004). These findings lead us to predict correlations between at least some word-level durational features and non-native speakers' perceived accentedness.

2. Methods

2.1. Participants

We analyzed recordings from 20 native Korean speakers, 20 native Chinese speakers⁴, and 12 native American English speakers. All were recorded as part of the Wildcat Corpus (for details see Van Engen et al., in press). The native Korean speakers (eleven males, nine females) ranged in age from 25 to 33 years old (mean = 27.85), the native Chinese speakers (ten males, ten females) ranged from 22 to 31 (mean = 24.4), and the native English speakers (six males, six females) ranged from 18 to 33 (mean = 21.58). None of the speakers reported having any speech or hearing impairments.

Most of the non-native speakers were recruited from the Northwestern University International Summer Institute (ISI), an intensive English program for incoming Ph.D. students. In addition to the ISI students, twelve of the native Korean speakers and four of the native Chinese speakers were recruited from the Northwestern community by word of mouth and advertisements posted on campus. All speakers were paid for their participation. Table 1 provides information about the Korean and Chinese groups' English experience, including the age at which they began studying English, the length of their English studies, and the length of time they had spent in English speaking countries. Out of the 20 participants in each group, 12 Korean participants and 11 Chinese participants were formally studying English through either tutoring or classes at the time of the experiment.

2.2. Materials

We analyzed recordings of two paragraphs. The first was the 'Stella' paragraph from The Speech Accent Archive at George Mason University (Weinberger, <http://accent.gmu.edu/>). The paragraph includes many difficult words for non-native speakers of English (e.g. words containing /θ/, /ð/, /ʌ/, /l/, and consonant clusters). It is four sentences long and was designed to contain most of the consonants, vowels, and consonant clusters of standard American English. Out of the 42 words in the Stella paragraph that were used in the analysis of lexical effects, 57% are content words, and 43% are function words.

The second paragraph was the 'Gina's Pizza' paragraph, which was designed by one of the experimenters (RB) to examine word reduction as a function of status in the discourse (first vs. second mention). It contains eleven repeated words, spread across ten sentences. The distance between first and second mentions ranged from five to 132 words. The paragraph was designed to ensure that the repeated mentions of words appeared in similar phonetic and prosodic contexts. Three phrases (*Gina's Pizza Shop*, *Johnson Expressway*, and *blue steeple*) were repeated, so both mentions of the words contained in these phrases appeared in partially identical contexts. For example, both mentions of the word *Johnson* were followed by the word *Expressway*. As punctuation marks are often accompanied by prosodic phrase

⁴One of the Chinese participants self-identified as a Cantonese speaker; three participants self-identified as Mandarin speakers; the rest listed Chinese as their native language.

breaks (Taylor & Black, 1998), both mentions of each word appeared in identical positions relative to periods. The Gina's Pizza paragraph was part of a set of paragraphs analyzed in Baker and Bradlow (2009). Both paragraphs are provided in Appendix A.

2.3. Recordings

All participants were recorded in sound-treated booths. Their speech was recorded using an AKG C420 headset microphone and a Marantz PMD 670 flash recorder, with a sampling rate of 22.05 kHz. All of the scripted materials in the Wildcat Corpus, of which the Stella and Gina's Pizza paragraphs form a part, were read off of a computer screen, and participants used the mouse or a keystroke to advance from item to item.

2.4. Measurements

The recordings of the Stella paragraph were first automatically time-aligned with the paragraph transcription using a new alignment interface, NUaligner, which was developed for this project. This program utilizes the SONIC speech recognition toolkit developed by the Center for Spoken Language Research (CSLR). It can take as input a transcription, which has been segmented into a series of short phrases (e.g. "Please call Stella"), and a recording of the paragraph, which has also been hand-segmented into these phrases. It generates time-aligned word-level transcriptions for the recordings, which can be opened in Praat. NUaligner is more accurate than systems that automatically align an entire transcription to a recording because errors cannot propagate beyond a phrase. This is absolutely necessary for non-native speech, which deviates substantially from the standard expected by the speech recognition software.

Because the automatic aligner is easily misled by extraneous sounds, mispronunciations, and distortions, hand correction is necessary for both native and non-native speech recordings. As part of the hand correction process, human aligners annotated instances when the recordings deviated from the text of the Stella paragraph, leading to word additions or deletions. If a word was repeated because of disfluency, the duration of the second production was annotated and used in our analysis. During hand correction, both the waveform and spectrogram were used, and boundaries were placed at the nearest zero-crossing on the waveform. Hand correction conventions were developed by a group of five trained linguists after aligning several recordings of both native and non-native speakers. The conventions describe which acoustic features mark the start and end of each word. We treated some sets of words as single units because of the great difficulty associated with locating the boundary between adjacent phonemes that were identical or acoustically similar (e.g. with the same manner of articulation), and stops followed by /h/. These sets of words are: *ask her*; *these things* (occurring twice in the passage); *six spoons*; *fresh snow*; *maybe a*; *we also*; *big toy*; *for the*; *three red bags*; *we will*; *meet her Wednesday at*.

All but one of the Stella paragraph recordings were hand corrected by a single aligner (KM). The remaining recording was corrected by four aligners separately, to check the reliability of the duration measurements. These four aligners then met and decided upon a consensus version to be used in the analysis. Although the average length of a word in this paragraph was 334.17 ms, the average standard deviation between the four aligners was only 9.91 ms, and the average range was 21.15 ms.

The Gina's Pizza paragraph was only used to examine the differences in duration between first and second mentions of words. Therefore, only the repeated words were measured for this paragraph. Unlike the alignment process for the Stella paragraph, all measurements for the Gina's Pizza paragraph were done by hand. For the Gina's Pizza paragraph measurements, both the waveform and spectrogram were used and boundaries were placed

at the nearest zero-crossing on the waveform. The measurement conventions for this paragraph were developed by two trained linguists after making measurements on several recordings of both native and non-native speakers. For each word, the conventions describe the acoustic features that mark its start and end. Because of the known effects of disfluencies on duration (Bell et al., 2003), we removed all disfluent words. Any target word that contained or was adjacent to a major disfluency (e.g. a long pause in the middle of the word, stuttering, or repetition) or that was strongly mispronounced (e.g. half of the word was not produced) was marked, and was not included in the analysis. Whenever one mention of a word was removed from the analysis due to disfluency, the other mention of the word in the recording was also removed. All recordings of the Gina's Pizza paragraph were measured by the same researcher who aligned the Stella paragraph (KM). A second researcher (RB) measured a subset of the recordings to check the reliability of the duration measurements. While the average target word duration in the paragraph was 395.98 ms, word measurements by the two aligners differed by an average of 19.98 ms.

2.5. Accent Ratings

2.5.1. Read Speech Accent Ratings—We conducted an accent rating experiment on Stella paragraph readings produced by 13 native and 52 non-native speakers of English. These readings ranged from 14 to 32 seconds long. The 52 non-native speakers included all 40 of the Chinese and Korean speakers analyzed in this study, plus 12 speakers with a variety of other language backgrounds. The native English speakers' readings (20% of the stimuli), served as the native anchor for the accent ratings. Fifty native English speaking Northwestern undergraduate students (ranging in age from 19 to 34) participated in the Stella paragraph accent rating experiment. They received course credit for their participation. None of the listeners reported any speech or hearing impairment. Each listener listened to the 65 Stella paragraph readings (one for each speaker) and rated them on a scale of 1 to 9, with 1 being “native” and 9 being “foreign”. The Stella paragraph readings were presented to each listener in a random order with no speaker repetitions, using SuperLab Pro software. The average accent rating for the native speakers was 1.27, for the Chinese learners of English was 6.63, and for the Korean learners of English was 6.31. In order to overcome the different ranges of accent ratings used by different listeners, we converted the accent ratings for each listener into z-scores. We then took the mean of all the z-scores for each speaker, to get their normalized read speech accent rating.

2.5.2. Spontaneous Speech Accent Ratings—As part of a separate study (Kim, in prep), accent ratings were collected on spontaneous speech samples from 21 task-oriented conversations between native and non-native (Chinese or Korean) English speakers. The spontaneous speech accent ratings used samples from 18 of the 20 Chinese participants in this study and 16 of the 20 Korean participants in this study. We therefore include the data from these accent ratings for comparative purposes in this report. The conversations were recorded during the ‘Diapix’ task in the Wildcat Corpus, described in detail in Van Engen et al. (in press). Nine short speech samples (between 1 and 2 seconds long) were included from each speaker's utterances during the conversations (three from the first third of the conversation, three from the second third, and three from the last third). Each speech sample consisted of one intonational phrase. When selecting the samples, the experimenter made an effort to minimize speech disfluency, such as the use of fillers like *um* and *uh*. In total, 378 speech samples were selected from the 42 speakers (34 non-native and 8 native). The 378 speech samples were divided into three blocks (126 samples per block) for accent rating tests, and were presented in a random order for each listener, using Inquisit 2.0 (<http://www.millisecond.com/>). Each listener rated the degree of accentedness of each speech sample on a scale of 1 (“native”) to 9 (“foreign”). Fifteen native English speaking Northwestern undergraduate students (ranging in age from 19 to 22) participated in the

spontaneous speech accent rating experiment. They received course credit for their participation. None of the listeners reported any speech or hearing impairment. The average accentedness score for the native speakers was 1.28, for Chinese the learners of English was 6.19, and for the Korean learners of English was 6.08. Like the read speech accent ratings, all spontaneous speech accent ratings were normalized by converting them to z-scores and averaging the z-scores across listeners.

2.5.3 Accent Rating Correlation—There are a number of differences between the samples used in the read and spontaneous speech accent ratings. In addition to the inherent differences between read and spontaneous speech, the read speech accent rating samples were longer. The read speech samples were also the same for every speaker and contained pauses and disfluencies. In contrast, the spontaneous speech samples were different for every speaker (and listeners heard multiple samples from each speaker) and were chosen for their fluency. Because of these differences, we examined the relationship between the two accent ratings with a Pearson correlation. There was a significant correlation between the two accent ratings ($r=0.469$ $p=0.01$). However, this correlation was driven by two Chinese speakers. If we removed these two speakers, the correlation would no longer be significant ($r=0.168$ $p=0.36$). Due to the tenuous relationship between the two sets of accent ratings, we compared both sets to our durational measures.

3. Results

3.1. Analysis of Overall Speech Duration (Stella Paragraph)

We first examined the differences between the three speaker groups' total speech durations in the Stella paragraph, using a between-subjects ANOVA. The ANOVA showed a significant effect of speaker group ($F(2, 49)=29.64$, $p<0.001$). Follow-up t-tests showed that native speakers had significantly shorter total speech durations than both Chinese and Korean speakers (English vs. Chinese: $t(30)=-5.68$, $p<0.001$; English vs. Korean: $t(30)=-9.45$ $p<0.001$). There was no significant difference between the total speech durations for Chinese and Korean speakers ($t(38)=-1.65$, ns). The mean and standard deviation of the total Stella paragraph speech duration for the English group was 17.32 sec. ($SD=1.39$), for the Korean group was 24.57 sec. ($SD=2.42$), and for the Chinese group was 23.05 sec. ($SD=3.31$). This result indicates that our non-native participants were behaving like non-native speakers in other studies, in that they were speaking more slowly than native speakers (e.g. Anderson-Hsieh & Horabail, 1994; Guion et al., 2000; Lennon, 1990; Munro & Derwing, 1995). To further explore the variation within the non-native group, we examined whether English experience measures affected non-native speech durations. We ran correlations between non-native speech duration and three measures of language experience: age at which English study began, years of English study, and months spent in an English speaking country. We also ran a t-test comparing total speech durations for participants who were formally studying English at the time of the experiment to those of participants who were not studying English. None of these analyses showed a significant effect of language experience on non-native speech durations.

3.2. Analysis of Within-Speaker Word Duration Variability (Stella Paragraph)

The analysis of word duration variability in the Stella paragraph was based on 'relative' durations, rather than absolute durations, in order to control for differences in speech rates between the participants. Relative durations were calculated by dividing each word's duration by the sum of the durations of all words for that speaker, resulting in the proportion of speech time that the participant spent on that word.

Within-speaker word duration variability was quantified as the variance of the relative word durations for a participant, resulting in a single score for each participant. An ANOVA showed a significant effect of speaker group ($F(2, 49)=6.13, p<0.005$). Follow-up t-tests showed that native speakers had significantly greater relative word duration variance than Chinese and Korean speakers (English vs. Chinese: $t(30)=3.64, p<0.005$; English vs. Korean: $t(30)=2.55, p<0.05$). There was no significant difference between Chinese and Korean speakers ($t(38)=-0.98, ns$). These data can be seen in Figure 1. In addition to L1, the effect of non-native speakers' English experience on their within-speaker variability was examined. Correlation tests were used to evaluate the relationship between non-native within-speaker variability and the age at which the non-native's English study began, their years of English study, and the number of months they had spent in an English speaking country. A t-test compared within-speaker variability for non-native speakers who were and were not studying English at the time of the experiment. None of these tests of English experience were significant.

Because the variance differences between native and non-native speakers may reflect differences in the treatment of function words (discussed in Section 3.3), we also examined the variance for content words alone. To calculate within-speaker variability in content word durations, we examined the 24 'isolated' content words in the paragraph, i.e. all content words that were not combined with adjacent words in the alignment process. For each participant, we computed the variance of the relative durations of these content words. An ANOVA on within-speaker variability for content words showed no significant effect of speaker group ($F(2, 49)=2.3, p=0.11$). This null result could be due to a lack of power when we examine content words alone. As Figure 2 shows, the non-native speakers are trending towards reduced variance for content words, relative to native speakers. However, the null result for content word variance supports the idea that the variance differences found for all words are at least partially driven by native English speakers producing greater differences between their content and function words. We explore this possibility in Section 3.3.

3.3. Analysis of Word Type, Word Frequency, and Language Effects on Word Duration (Stella Paragraph)

3.3.1. Regression Method—We analyzed the effects of word type, word frequency, native language, and L2 experience on word duration in the Stella paragraph with linear mixed effects regressions. The regressions were run in R version 2.9.1, using the `lmer` function. The regression analyses allowed us to examine multiple lexical effects simultaneously for the language groups while controlling for other factors influencing a word's duration. The dependent variable in all regression models was raw single word duration.

It is important to note that if a regression model includes an interaction, then the estimates for the individual effects that are included in the interaction actually describe conditional effects, rather than main effects. For example, an analysis may include an interaction between a categorical variable (e.g. word type: content vs. function) and a continuous variable (e.g. years of English study). The results for the individual effects when an interaction is not included reflect how word type relates to the dependent variable and how years of English study relates to the dependent variable. However, when the interaction is included, each of these individual effects is calculated as if the other variable were set to zero. So the coefficient for the word type variable applies only to participants who have spent zero years studying English (Aiken & West, 1996). Because zero is not a very informative value for most independent variables, all continuous variables involved in interactions were centered by subtracting the mean for all scores from each score. All categorical variables in these analyses were contrast coded. This transformation and coding

make the conditional effects more interpretable (Aiken and West 1996). In addition, we began each analysis by first building a model without any interactions, to determine the main effects across all values.

In the Stella paragraph regressions, we included random variables for participant and word. We also included three control variables: total speech time, number of phonemes in each word, and number of syllables in each word. Total speech time for the paragraph controlled for variation in speech rate across different speakers. Number of phonemes and number of syllables per word controlled for the inherent lengths of different words. The phoneme and syllable counts for each word were taken from the MRC psycholinguistic database (Wilson, 1988). Following Bell et al. (2009) we included a control variable in the regression if it improved the fit of the control model (at $p < 0.20$). All three of our control variables improved the fit of the control model at $p < 0.1$ (number of phonemes: $p < 0.001$, number of syllables: $p < 0.1$, total speech time: $p < 0.001$). All three control variables also significantly contributed to the control model ($p < 0.05$). We examined previous word duration as a possible fourth control variable, in order to control for the potential non-independence of words that are adjacent in the recordings.⁵ However, this variable did not improve the fit of the control model ($p = 0.34$), so it was excluded.

Our variables of interest that related to lexical properties were log word frequency, and word type (content vs. function). For the frequency variable, we took the log of the frequency, so as to work with a less skewed distribution of frequencies. We used frequency counts from BYU-BNC: The British National Corpus (Davies, 2004-), a 100 million word corpus consisting of samples of written and spoken British English from a variety of sources. We examined word type by separating all isolated function words (18 words) from all isolated content words (24 words). The function words included articles, prepositions, conjunctions, pronouns, the infinitival marker *to*, and the modal verb *can*. The content words included nouns, verbs, adjectives, numerals, and the interjection *please*. All isolated words in the Stella paragraph, along with their word type (content or function), BNC frequency, part of speech, and phoneme and syllable counts, are listed in Appendix B. We also included variables of interest related to participants' language experience. All participants were put in one of two nativeness categories (native vs. non-native English speakers). In addition, for non-natives, we included language experience variables for L1 (Chinese vs. Korean), and four factors relating to L2 experience: age at which English study began, years of English study, months spent in an English speaking country, and whether participants were formally studying English at the time of the experiment.

3.3.2. Regression Results—The native main effect Stella regression was used to establish whether we could replicate the frequency and word type effects found for native speakers in previous studies with our recordings of the Stella paragraph. This regression used only data from the English group. The regression included the two variables of interest (log frequency and word type) as well as the control variables (total speech time, number of phonemes in each word, and number of syllables in each word). The estimates, confidence intervals, and p -values for all the variables of interest in the native main effect Stella regression are presented in Table 2. Out of the variables of interest (log word frequency and word type), only log word frequency had a significant main effect in this regression. Log word frequency had a negative effect on word duration: more frequent words tended to be shorter than less frequent words.

The non-native main effect Stella regression was used to examine whether the frequency effect found for native English speakers in the first regression also appears for non-native

⁵This method for controlling for the relationship between adjacent words is discussed in Baayen (2008).

speakers. It also examined whether non-native speakers show a word type effect (which did not appear for native speakers in the first regression). Finally, it examined whether features of the non-native speakers' language experience directly affected their word durations. We included variables for L1 (Chinese vs. Korean), and four factors relating to L2 experience: age at which English study began, years of English study, months spent in an English speaking country, and whether participants were formally studying English at the time of the experiment. The estimates, confidence intervals, and *p*-values for all the variables of interest in the non-native main effect Stella regression are presented in Table 3. As with the first regression, the only significant effect on duration was log frequency.

Although L1 and L2 experience factors did not have significant effects on word duration, it is possible that they influenced the size of the word frequency effect. For instance, English learners who had spent more time in English speaking countries might show a stronger frequency effect than those who had spent less time in these countries. To test this possibility, we built a non-native two-way interaction Stella regression model, which includes all interactions between the two lexical variables of interest (log frequency and word type) and the five L1 and L2 experience variables (L1, age at which English study began, years of English study, months spent in an English speaking country, and whether participants were formally studying English at the time of the experiment). The estimates, confidence intervals, and *p*-values for all the variables of interest in the non-native interaction Stella regression are presented in Table 4. Once again, log frequency has a negative effect on word duration, while word type does not have a significant effect on word duration, although the effects in this regression are conditional rather than main effects. Only one interaction between lexical and experiential variables was significant: the interaction between word type and years of English study. As Figure 3 demonstrates, this interaction is due to the fact that participants who have spent more time studying English have longer function words relative to their content words.

The native and non-native Stella regressions revealed that log word frequency has a significant effect on word duration for both native and non-native speakers, while word type does not have a significant effect (in models that include a frequency variable). In order to compare the size such effects across native and non-native participants, it is necessary to use a single model with data from both groups. In the native/non-native two-way interaction Stella regression model, the variables included log frequency, word type, and nativeness (native vs. non-native English speakers), as well as two-way interactions between these three variables. The estimates, confidence intervals, and *p*-values for all the variables of interest in the native/non-native two-way interaction Stella regression are presented in Table 5. This regression showed a significant interaction between log word frequency and the native vs. non-native contrast. As illustrated in Figure 4, the effect of lexical frequency on word duration was stronger for the English language group than the Chinese and Korean groups. There was also a significant interaction between word type and the native vs. non-native contrast. There was a greater difference between function and content words for the English group than for the Chinese and Korean groups, as shown in Figure 5.

To complete our understanding of the relationship among lexical factors and native status, a final native/non-native regression model was built which included a three-way interaction among log frequency, word type, and nativeness. The estimates, confidence intervals, and *p*-values for all the variables of interest in the native/non-native three-way interaction Stella regression are presented in Table 6. The model shows that this three way interaction is significant.

To explore the three-way interaction in the native/non-native regression more fully, we ran separate regressions for content and function words. These separate regressions are related

to the ANOVA analysis of within-speaker variability for content-words alone (Section 3.2), in that both the regressions and ANOVA remove the effects of word type from the analysis. The difference between the within-speaker variability content word ANOVA and the content word regression is that the variability ANOVA captures durational variance from all sources, while the content word regression focuses on the role of lexical frequency in word durations. The results of the content and function word regressions can be seen in Tables 7 and 8, respectively. In the content word regression, the interaction between log frequency and the native vs. non-native contrast approached, but did not reach, significance. However, in the function word regression, the interaction between log frequency and the native vs. non-native contrast was highly significant. Non-native English speakers showed a stronger frequency effect than natives for function words. The frequency effects for content and function words in each of the three language groups can be seen in Figure 6. The significant effects of frequency in the content and function word regressions should be interpreted with caution because they are simply conditional effects. However, they do suggest that the frequency effects that were seen in the native and non-native regressions are not simply word type effects (high frequency function words vs. low frequency content words). The downward slopes of all the regression lines in Figure 6 support this idea.

3.4. Analysis of Mention and Language Effects on Word Duration (Gina's Pizza Paragraph)

3.4.1. Regression Method—We analyzed the effects of mention, native language, and L2 experience on word duration in the Gina's Pizza paragraph with linear mixed effects regressions. These regression models were similar to those used for the Stella paragraph, with some changes to the variables included. The dependent variable was once again word duration in seconds.

In the Gina's Pizza paragraph analyses, we again used random variables for participant and word. We included one control variable: total duration of all target words for a speaker. This allowed us to control for variation in speech rate across different speakers. We did not control for inherent word length (e.g. number of phonemes) because the variable of interest in these analyses was the effect of mention. Therefore, identical sets of words were included in the first mention and second mention groups. When a word was removed from analysis (e.g. due to disfluency), the other mention of that word for that speaker was also removed from analysis. Our only variable of interest related to lexical features was mention (first mention vs. second mention). We also included variables of interest related to participants' language experience. All participants were put in one of two nativeness categories (native vs. non-native English speakers). In addition, for non-natives, we included language experience variables for L1 (Chinese vs. Korean), and four factors relating to L2 experience: age at which English study began, years of English study, months spent in an English speaking country, and whether participants were formally studying English at the time of the experiment. The categorical variables (nativeness, L1, and whether participants were studying English at the time of the experiment) were contrast coded, and the continuous variables (age at which English study began, years of English study, and months spent in an English speaking country) were centered.

3.4.2. Regression Results—As in our analysis of the Stella paragraph, we began by testing whether we could replicate previously found lexical effects – in this case, second mention reduction – with the current materials. The estimates, confidence intervals, and *p*-values for the variables of interest in the native main effect Gina's Pizza regression are presented in Table 9. There was a significant main effect of mention on word duration. This was a negative effect, indicating that second mentions tended to be shorter than first mentions.

The non-native main effect Gina's Pizza regression examined whether our non-native participants reduced second mentions of words relative to first mentions, and tested whether L1 or L2 experience factors (L1, age English study began, years of English study, months in an English speaking country, and whether the participant was studying English at the time of the experiment) affected their word durations. The estimates, confidence intervals, and *p*-values for all the variables of interest in the non-native main effect Gina's Pizza regression are presented in Table 10. Like the native participants, the non-native participants showed significant second mention reduction. As in the previous regression, this was a negative effect, showing that non-native speakers produced shorter second mentions than first mentions. However, none of the L1 or L2 experience factors directly affected word durations.

Just as the L1 and L2 experience factors could influence frequency or word type effects in non-native speech, they could also influence second mention reduction. To explore this possibility, a non-native two-way interaction Gina's Pizza regression model was built. The estimates, confidence intervals, and *p*-values for all the variables of interest in this model are presented in Table 11. This regression model shows that none of the L1 or L2 experience factors tested significantly affected non-native speakers' second mention reduction.

The native and non-native main effect regressions have demonstrated that second mention reduction appears in both native and non-native speech. In order to test whether there are differences between native and non-native speakers in the size of this effect, we built a native/non-native two-way interaction regression model, which includes the interaction between nativeness and mention. The estimates, confidence intervals, and *p*-values for all the variables of interest in the native/non-native two-way interaction regression are presented in Table 12. There is no significant interaction between mention and nativeness, indicating that the degree of second mention reduction is roughly comparable for the native and non-native participants in this study. The second mention reduction effect for all three language groups is illustrated in Figure 7, which shows the ratios of first mention mean duration divided by second mention mean duration for each word, within each language group. All ratios greater than 1 indicate reduction.

3.5. Analysis of Between-Speaker Variability (Stella Paragraph)

Between-speaker variability was quantified as the variance in duration for each word across participants (within each language group). The non-parametric Kruskal-Wallis test and Wilcoxon signed-rank tests were used to analyze these data because of their non-normal distribution. The Kruskal-Wallis test showed a significant effect of language group on between-speaker variability ($H=30.28$ (2, $N=159$) $p<0.001$). Follow-up Wilcoxon signed-rank tests indicated that the native English speakers were less variable than the Chinese ($W=166$, $p<0.001$) and the Korean ($W=147$, $p<0.001$) speakers. In other words, the natives formed a more homogeneous group. Again, there was no significant difference between Chinese and Korean speakers, ($W=717$, ns). In order to examine the internal cohesion of each of the two non-native language groups, we compared the variance within each of the non-native language groups to the variance of non-native speakers as a whole. We chose this statistical method because there is no reason to expect that one group of non-native speakers will have greater between-speaker variance than the other. However, it is certainly possible that differences between non-native speakers with different L1s would lead the between-speaker variance for the combined non-native group to be greater than the variance for a group of non-native speakers with a shared L1. We found no significant difference between the between-speaker variance for the Chinese speakers and the entire non-native group ($W=685$, ns), or between the between-speaker variance for the Korean speakers and the entire non-native group ($W=735$, ns). This suggests that sharing a language background did

not make these non-native English speakers more similar to one another, in terms of word-level durations.

3.6. Analysis of Accent Ratings

Surprisingly, the spontaneous speech accent ratings correlated better with non-native word duration measures than the read speech accent ratings, even though most of these measures were based on data from the very recordings used for the read speech accent ratings. Within-speaker word duration variance was negatively correlated with both types of accent ratings (read speech: $r = -0.39$, $p < 0.05$ ⁶; spontaneous speech: $r = -0.55$, $p < .001$). These results indicate that non-native speakers with greater variance in their word durations received lower (more native-like) accent ratings. None of the other word duration measures correlated with the read speech accent ratings. Relative duration of function words was correlated with spontaneous speech accent ratings ($r = 0.43$, $p < 0.05$). This means that non-native speakers who produced shorter function words received more native-like accent ratings. We also analyzed the correlation between accent ratings and similarity to the native centroids. The native centroid for a word is its mean relative duration, averaging across all native speakers. We calculated a non-native speaker's similarity to the native centroids as the Spearman correlation between her word durations and the native centroids for the words. Similarity to the native centroids was negatively correlated with spontaneous speech accent ratings ($r = -0.39$, $p < 0.05$). This shows that non-native speakers who produced relative durations that were similar to the means for native speakers received more native-like accent ratings. Total Stella paragraph duration, second mention reduction ratio, and within-speaker content word variance did not significantly correlate with either set of accent ratings.

Although the focus of these analyses was to determine the effects of durational features on accent ratings, it is also interesting to determine how the English experience factors that we have used in the preceding analyses relate to accentedness. We ran correlations between both type of accent ratings and 1) the age at which English study began, 2) years of English study, and 3) months spent in an English speaking country. We also ran t-tests to compare accent ratings for non-native participants who were formally studying English at the time of the experiment to accent ratings for those who were not. Of these eight analyses, the only significant result was the correlation between the number of months spent in an English speaking country and both sets of accent ratings (read speech: $r = -0.53$, $p < 0.005$; spontaneous speech: $r = -0.50$; $p < 0.005$). These correlations suggest that the length of time spent in an English speaking country is a good predictor of accentedness.

4. Discussion

These results reveal a number of interesting similarities and differences between native and non-native English word durations. Native and non-native English speakers were similar in that they both exhibited predictability-related effects on word duration. Specifically, both frequency and mention influenced word durations in native and non-native speech. However, native and non-native speakers differed in that native speakers produced shorter words and greater relative variance in their word durations. Native speakers also had a greater difference between content and function words, stronger effects of frequency over all words, and weaker effects of frequency over function words than non-native speakers. Interestingly, we did not find any differences between the native Chinese and Korean speakers. Still, native speakers had more similar word durations to each other than non-native speakers did. Some word-level durational features were correlated with the perceived

⁶The read speech accent rating correlation was driven by two Chinese speakers. If we removed these participants from the analysis, the correlation with the read speech accent ratings would disappear ($r = -0.20$, $p = 0.27$), but the correlation with the spontaneous speech accent ratings would remain ($r = -0.45$, $p < 0.01$).

accentedness of the non-native speakers. Non-native speakers who had native-like relative word durations, with greater variance and more reduced function words, were judged to be less accented. In the discussion below, we answer the questions posed in the introduction.

- (1) Are there differences between native and non-native English speakers in terms of word-level duration? If so, can these differences be explained by lexical features of English?

Like non-native participants in previous studies, the Chinese and Korean learners of English in our study spoke more slowly and had less within-speaker durational variance than native English speakers. That is, the native speakers generally produced shorter short words and longer long words than the non-natives. Previous research (Anderson-Hsieh & Horabail, 1994; Fokes & Bond, 1989; Lee et al., 2006; Mochizuki-Sudo & Kiritani, 1991; Shah, 2004) has shown that reduced within-speaker durational variance is a common feature of non-native English vowels and syllables. Our study extends this finding to non-native words. Our analyses suggest that non-native speakers have reduced within-speaker word duration variance, at least in part, because they reduce function words less than natives. When we examined content and function words together, we found significantly greater word duration variance for native English speakers than non-native speakers. However, when we examined content words alone, there was only a trend towards greater variance for the native speakers. Research on learners of other languages is needed to determine whether such reduction in word duration variance is a feature of non-native speech in general, or specific to non-native English.

In both native and non-native English speech, lexical properties of words influenced their durations. Words that were more predictable, either because they were more frequent or because they had already been mentioned in the paragraph, were shorter than less predictable words in native and non-native English. We found frequency effects on duration for both content and function words. However, as Bell et al. (2009) point out, the frequency effect we found for function words may actually reflect a split between high and low frequency function words. Ten of the 18 function words analyzed in this study are among the ten most frequent English function words. High-frequency function words may have idiosyncratic features, like more commonly used reduced forms (e.g. /ə/ for *a*), which could lead to greater reduction of these words.

The fact that, like native English speakers, non-native speakers showed predictability effects on duration suggests that at least some of the processes leading to predictability-related reduction in native speech are also at work in non-native speech. The possible processes include a listener-focused desire for clarity on less predictable words (Lindblom, 1990), more articulatory practice for frequent words (Bybee, 2001, 2006), faster lexical retrieval for frequent words (Bell et al., 2009), and shorter stored exemplars of more frequent words (Bybee, 2001; Pierrehumbert, 2001, 2002). The frequency effect on duration that we found for non-native speakers supports Schmitt and Dunham's (1999) finding that non-native English speakers can have relatively accurate representations of English word frequencies. This could be due to their knowledge of the frequencies of words representing similar concepts in their L1, rather than being solely based on English exposure. Such a possibility might be tested by examining the frequencies of L1 translations of each word, to determine whether these translation frequencies show as strong a relationship with word duration as the English frequencies. The second mention reduction effect that we found demonstrates that these non-native speakers were able to track whether a word had already been mentioned in the discourse. It also shows that their speech production was influenced by changes in word predictability within a discourse. One avenue of future research is to try to determine whether the same set of processes are at work in native and non-native speech, using experiments that distinguish between the possibilities outlined above. For instance,

replicating Gahl's (2008) results examining homophones with different frequencies using non-native speakers would indicate that frequency effects in non-native speech, like those in native speech, are due to more than articulatory practice.

We found no main effect of word type in regression models that also included word frequency. Our failure to replicate Bell et al.'s (2009) word type effect is not surprising, however, given the distribution of word frequencies across function and content words in the Stella paragraph. The highest frequency content word (*go*; frequency: 84 845) had a lower frequency than the lowest frequency function word (*her*; frequency: 100 352), meaning that there was no overlap in frequency across the two groups. In contrast, Bell et al. (2009) examined a subset of the Switchboard Corpus, which had considerable overlap in the frequencies of content and function words.

Although both native and non-native speech showed predictability effects, native speakers produced greater differences between content and function words and between high and low frequency words than non-native speakers. The interaction we found between word type and native status extends earlier research on Japanese learners of English to Chinese and Korean learners. Like Ueyama (2000) and Aoyama and Guion (2007), we found that non-native English speakers reduced function words less than native speakers. The greater frequency effect we found for native speakers is largely due to their greater reduction of function words, which are the highest frequency words in the paragraph. When we examined frequency effects in content and function words separately, we did not find significantly greater frequency effects for native speakers in either set of words. While there was a trend towards greater frequency effects for natives in the content words, the only significant difference between the natives and non-natives was actually a significantly greater frequency effect for non-natives in the function words. Examining Figure 6 and the relative duration per segment (RDPS) values for words in each of the three language groups reveals that this is because the non-native speakers have a much wider spread of function word durations. Instances of two of the three most frequent words in the passage (*the* and *and*) were noticeably reduced by non-native speakers, but there is a great deal of overlap between RDPS values for function and content words in non-native speech. In contrast, native English speakers had function word RDPS values that generally clustered together tightly, below the RDPS values for most content words. It is possible that native English speakers' stronger delineation between function and content word durations often puts an upper limit on function word durations. This reduces the range of function word durations and weakens the correlation between function word duration and frequency. However, non-native English speakers have greater overlap between function and content word durations, and therefore have a much wider range of possible function word durations. This allows greater distinctions between the durations of low and high frequency function words, and therefore a stronger correlation between function word duration and frequency.

While native English speakers reduced function words and frequent words more than non-native speakers, the two groups produced similar levels of second mention reduction. This difference could be partially due to methodological challenges inherent in studies of frequency effects. Specifically, lexical frequency estimates that were derived from a corpus of native English language use might not match the frequency with which non-native speakers are exposed to these words, while exposure to previous mentions in a read paragraph is the same for natives and non-natives. The difference between word type and frequency effects on the one hand, and second mention reduction effects on the other, could also reflect the separate sources of these lexical effects. Function word reduction and frequency effects on duration both depend on knowledge of how words are used in the English language. Function words play a special role in language, and word frequency reflects how often a word is used in the language as a whole. In contrast, second mention

reduction is due to the increased predictability of a word within a specific discourse. Non-native speakers may be more influenced by the context in which a word appears than its broader usage in the language. Future research should explore this possibility by examining predictability effects in non-native speech at a wider range of levels, such as predictability in the language (frequency effects), predictability in the discourse (e.g. second mention reduction), predictability in the sentence (e.g. verb-bias effects; Gahl & Garnsey, 2004), and predictability in the phrase (e.g. joint or conditional probability based on surrounding words; Bell et al. 2003; Bell et al., 2009).

(2) Do non-native speakers, with or without a shared language background, produce similar word-level durational features?

Native English speakers formed a more homogenous group than both non-native speakers overall and non-native speakers who shared a first language. In other words, native English speakers produced more similar English word durations than native Chinese speakers and native Korean speakers. This is not surprising, as native English speakers share a prosodic system, and have essentially the same degree of proficiency in English. In contrast, non-native English speakers use an interlanguage which may combine aspects of the prosodic systems of their native language and English (Rasier & Hiligsmann, 2007). It is likely that these individual interlanguages vary in which aspects of English they incorporate. In addition, non-native speakers vary in their proficiency, so some will incorporate more English features than others. Finally, non-native speakers' word durations are influenced by their difficulties with English, such as problems with particular segments or clusters of segments, or uncertainty about how to pronounce a particular word. Such difficulties may arise because of transfer from their L1, but individual speakers who share an L1 may still differ in the problems they have in their L2.

Interestingly, we did not find that the Chinese and Korean groups were more homogeneous within their subgroups than the non-native group as a whole. We also did not find that these two groups differed significantly in terms of their total speech durations or their within-speaker duration variances. This similarity across L1s may be because these languages are durationally similar enough that transfer from these L1s results in similar English word durations. It may also mean that the challenges associated with speaking a non-native language have a greater effect on non-native word durations than L1 to L2 transfer. Unfortunately, we do not have enough equivalent information on word duration effects in Korean and Chinese dialects to tease these possibilities apart. For instance, there is evidence for frequency effects on duration in Cantonese (Zhao & Jurafsky, 2009), and for function word shortening in Mandarin (Shi et al., 1998), but no equivalent data on Korean. In the future, we plan to examine these effects in Korean using the recordings of Korean read speech in the Wildcat Corpus (for details see Van Engen et al., in press). Such data would show us whether there are differences between Chinese dialects and Korean, which are not being realized in the English spoken by native speakers of these languages. Determining the relative importance of each of the factors discussed above requires research on language learners with a range of L1s and L2s and a variety of proficiency levels. These studies must also explore whether lexical effects on duration, like those discussed in this paper, appear in these learners' L1s. Only then can we determine which non-native durational features are due to transfer from a speaker's L1, and which are due to the cognitive challenges involved in speaking a non-native language.

Although we did not expect to see any differences in word predictability effects between the Chinese and Korean language groups, we did expect that the two groups might differ in their function word reduction. Given that Mandarin makes extensive use of function words, and these function words are phonetically reduced in Mandarin (Shi et al., 1998), we predicted that the Chinese group would reduce function words more than the Korean group. However,

we did not find any significant difference between the two groups in the amount of function word reduction: both groups reduced function words less than the native English speakers. These results do not support the hypothesis that weaker function word reduction in non-native speech is due to transfer of function word treatment from a non-native speaker's first language. The results instead support the idea that non-native speakers' weaker function word reduction arises, at least in part, from the challenges involved in speaking a second language, or from the non-native speakers' lack of mastery over the English prosodic system.

Ueyama (2000) suggested that non-native speakers might reduce function words less than native speakers because they produce more, smaller, prosodic units, making it more likely that a function word will form its own prosodic unit. The substantial differences between the English prosodic system and the prosodic systems used in Korean and Chinese dialects may add to this problem. Future work in this area should examine how the prosodic status of words in non-native English influences their durations. Such work would determine whether these differences between native and non-native speech are due to differences in prosodic organization (e.g. the locations of prominent words and phrase boundaries), or due to differences in gradient features like duration, independent of higher-level prosodic categories. Function words rarely receive pitch accents in native English speech (Lavoie, 2002), and are often included in the same prosodic word with a content word (Selkirk, 1996). Therefore, difficulties with prosodic phrasing and pitch accent placement would strongly affect function word reduction.

One suggested source of variance between non-native speakers of a language is their amount of experience with their L2. We examined four measures of L2 English experience: age that English study began, years of English study, months in an English speaking country, and whether the participant was studying English at the time of the experiment. In general, these measures did not significantly predict durational features in non-native speech. None of them correlated with total speech duration in the Stella paragraph, or with within-speaker word duration variance. They also were not significant predictors of word duration in either paragraph, and did not mediate frequency or second mention reduction effects. However, there was a significant interaction between word type and years of English study. Because native speakers tend to reduce function words more than non-native speakers, we might expect that more experienced non-native speakers also reduce function words more. Surprisingly, the opposite was true. Non-native English speakers who had been studying English for a greater number of years actually had longer mean function word durations relative to their mean content word durations. This could be because all of the non-native participants were well practiced at producing the function words in the Stella paragraph, but could be less familiar with some of the content words. They might produce less familiar words more slowly because they are being more careful or because they are producing more disfluencies. As the non-native speakers gain experience with English, they become more familiar with lower frequency content words, and could therefore produce them more quickly. This extra experience, however, does not seem to help them replicate native speakers' word duration patterns such as their greater reduction of high frequency words and function words. This may be one reason that length of English study did not significantly correlate with accentedness ratings.

Our findings of heterogeneity in our non-native participants' durational features have implications for researchers examining language acquisition. Participants in this study had relatively similar experiences with English: for instance, many had recently moved to the U.S. from Korea or China, and had passed standardized English tests with scores high enough for admission to an American graduate school. The fact that we still found such variance between participants suggests that researchers should be careful when applying results from one non-native speaker to another. This variance also illustrates the need for

studies on non-native speech to have a large sample size, in order to more fully capture the variation in the population.

(3) Are non-native word-level durational features associated with the perception of a stronger non-native accent?

We found that non-native English speakers who had greater variance in their relative word durations, greater function word reduction, and more native-like word durations were judged by native speakers to have more native-like accents. As discussed above, greater function word reduction was one cause of the native speakers' increased word duration variance. Therefore, the correlations involving word-duration variance and function word reduction are strongly linked. These experiments have demonstrated a correlation between non-native word durations and perceived accentedness. The next step is to determine whether these durational features actually cause the higher accentedness ratings, by using experiments that manipulate the durations of otherwise identical stimuli. If such causation is found, it would further highlight the importance of teaching language learners to produce native-like durations.

The very different results we found for analyses using the read speech accent ratings and those using the spontaneous speech accent ratings show the importance of choosing accent rating materials carefully. Our read and spontaneous speech accent ratings were significantly correlated, but this correlation was driven by two speakers, so the relationship between the two types of accent ratings did not hold for all speakers. In this study, we found significant correlations between accent ratings of spontaneous speech and three durational features of non-native speech, while we only found one correlation (driven by two participants) with the read speech accent ratings. There are a number of differences between the read and spontaneous speech samples used in these accent ratings. The read speech samples are relatively long (14–32 second) stretches of read speech, including pauses and disfluencies, and all speakers produced the same utterances. In contrast, the spontaneous speech samples are short (1–2 second) spontaneous utterances, chosen for their fluency and lack of pauses. The spontaneous speech samples are also different for every speaker, and listeners heard multiple samples for each speaker. We were surprised to find that the spontaneous speech accent ratings correlated better with our durational measures, because those durational measures were based on the very same recordings used in the read speech accent ratings. We may have found better correlations with the spontaneous speech ratings because they contain fewer disfluencies, allowing raters to focus on more subtle indicators of accentedness, like duration. It is also possible that the raters were more attentive when rating the spontaneous speech samples than the longer and more repetitive read speech samples. A final possibility is that we got more fine-grained accent ratings for the spontaneous speech samples because raters heard multiple samples for each speaker, so a single listener's accent rating for a speaker could fall between the whole numbers used in the rating scale. More research is clearly needed on the effects of passage length, variety, fluency, and spontaneity on accent ratings. Researchers should carefully consider these factors when designing and interpreting studies involving accent ratings.

In addition to their implications for second language teaching and research, these accent rating results may also be used to improve computer assisted language testing systems. The durational measures in this study which correlated with native English speaker accent ratings might be used to automatically evaluate the accentedness of a non-native speaker using speech recognition technology. These features might join existing features like phoneme-based duration scores (Neumeyer, Franco, Digalakis, & Weintraub, 1999) in language testing systems. As word duration variance was the measure that was most highly correlated with accent ratings, this is the most promising measure examined in this paper.

5. Conclusions

This study has shown that word durations in both native and non-native English are affected by a word's predictability, based on its frequency in the lexicon and whether it has already been mentioned in the discourse. However, we did find some differences between the native and non-native speakers. Like earlier researchers, we found that native speakers produced shorter durations and more within-speaker duration variance than non-native speakers. In addition, native speakers exhibited stronger function word reduction and frequency effects. Finally, there was less variance among native speakers than among non-natives.

Our results have both practical and theoretical implications for future research on second language acquisition. From a practical standpoint, the variance we found across our non-native participants illustrates the importance of large sample sizes in research on non-native speech. In addition, the differences we found between our two types of accent ratings demonstrate the large effect that features of recordings like length, variety, fluency, and spontaneity can have on accentedness judgments. From a theoretical standpoint, our results suggest that similar psycholinguistic processes are leading to the reduction of more predictable words in native and non-native speech. Despite the cognitive challenges involved in speaking a second language, non-native speakers are able to track which words are discourse-old, and acoustically reduce these words. Non-native speakers are also influenced by a word's frequency in the language as a whole, whether this is because they are storing reduced exemplars of more frequent words, retrieving more frequent words faster, or making an effort to produce less frequent words clearly. Although all lexical effects that appeared in native speech also appeared in non-native speech, some of the effects were stronger than others in non-native speech. Native speakers showed greater function word reduction and frequency effects than non-natives, but the two groups showed similar amounts of second mention reduction. This suggests that non-native productions are more influenced by the local discourse context than the usage of words in the language in general. Finally, the similar productions of function words by Chinese and Korean speakers provides evidence against the hypothesis that the role of function words in a speaker's L1 influences function word production in their L2. General cognitive difficulties or problems with the English prosodic system are more likely to influence non-native function word production.

Our results also have implications for language teaching. We have revealed links between word duration and accentedness, with non-native speakers who produce greater durational variance, greater function word reduction, and more native-like durations being perceived as less accented. These results highlight the importance of researching the factors that influence word duration in non-native speech and determining whether training language learners on word durations can lead to accent reduction.

Acknowledgments

We would like to thank Kelsey Mok for her excellent measurements, Page Piccinini for assistance with running the experiment, and both of them for their assistance in developing the alignment conventions. We thank Matt Goldrick for his advice on our statistical analyses. We also gratefully acknowledge Chun Liang Chan and Janet Pierrehumbert for their roles in the technical innovations that facilitated this work. These data were presented at the Acoustical Society of America meeting in Portland, OR, in May 2009; we thank the participants for their comments. This work was supported by grant R01-DC005794 from NIH-NIDCD.

Appendix A

Stella paragraph (Weinberger, Speech Accent Archive)

Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Gina's Pizza paragraph (Baker and Bradlow, 2009)

First mentions of target words are underlined, second mentions are italicized

If you want to go to Gina's Pizza Shop, I can tell you the best way to get there. Go straight down this street and follow the signs for the Johnson Expressway. However, don't actually go onto the *Johnson Expressway*. When you get to the on-ramp, take a left onto Cleveland Street, the main street in town. You'll go past a big school called *Cleveland High School*, right between a church with a yellow door and a *church* with a blue steeple. There is a small alley just past the church with the *blue steeple*. Take this *alley* for several blocks, and turn left on the third road you come to. Eventually, the *road* will split in two. Take Fillmore Boulevard, which is the one on the right. A block and a half later you'll see the sign for *Gina's Pizza Shop*, also known as the best pizza place in town.

Appendix B

Word	Word Type	BYU-BNC Frequency	Part of Speech	Num. Phonemes	Num. Syllables
scoop	c	156	verb	4	1
snack	c	328	noun	4	1
slabs	c	404	noun	5	1
frog	c	493	noun	4	1
peas	c	612	noun	3	1
Stella	c	645	proper noun	5	2
snake	c	726	noun	4	1
cheese	c	2526	noun	4	1
store	c	3510	noun	3	1
Bob	c	3709	proper noun	3	1
plastic	c	3923	noun	7	2
kids	c	4202	noun	4	1
thick	c	4392	adjective	3	1
train	c	6262	noun	4	1
brother	c	8277	noun	5	2
blue	c	8805	adjective	3	1
station	c	9899	noun	6	2
Please	c	12 804	adverb	4	1
call	c	12 961	verb	3	1
bring	c	15 007	verb	4	1
need	c	38 025	noun	3	1
five	c	39 985	cardinal number	3	1
small	c	42 179	adjective	4	1
go	c	84 845	verb	2	1
her	f	100 352	personal pronoun	2	1
into	f	157 627	preposition	4	2
her	f	203 367	determiner	2	1
can	f	231 720	verb	3	1
she	f	352 837	personal pronoun	2	1
from	f	404 306	preposition	4	1

Word	Word Type	BYU-BNC Frequency	Part of Speech	Num. Phonemes	Num. Syllables
with	f	639 913	preposition	3	1
for	f	824 359	preposition	2	1
to	f	1 585 779	infinitival marker `to'	2	1
a	f	2 109 903	article	1	1
a	f	2 109 903	article	1	1
and	f	2 615 087	conjunction	3	1
and	f	2 615 087	conjunction	3	1
and	f	2 615 087	conjunction	3	1
of	f	2 886 056	preposition `of'	2	1
of	f	2 886 056	preposition `of'	2	1
the	f	6 046 768	article	2	1
the	f	6 046 768	article	2	1

References

- Aiken, LS.; West, SG. Multiple regression: Testing and interpreting interactions. Sage; Thousand Oaks, CA: 1996.
- Anderson, AH.; Howarth, B. Referential form and word duration in video-mediated and face-to-face dialogues. In: Bos, J.; Foster, ME.; Matheson, C., editors. sixth workshop on the semantics and pragmatics of dialogue (EDILOG 2002); Edinburgh, U.K.. 4–6 September 2002; Edinburgh: Cognitive Science Centre, University of Edinburgh; 2002. p. 13-20.
- Anderson-Hsieh J, Horabail V. Syllable duration and pausing in the speech of Chinese ESL speakers. TESOL Quarterly. 1994; 28(4):807–812.
- Anderson-Hsieh J, Johnson R, Koehler K. The relationship between native speaker judgements of nonnative pronunciation and deviance in segmental, prosody, and syllable structure. Language Learning. 1992; 42:529–555.
- Aoyama, K.; Guion, SG. Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In: Bohn, O-S.; Munro, M., editors. Language experience in second language speech learning. John Benjamins Publishing Co.; Amsterdam: 2007. p. 281-297.
- Aylett M, Turk A. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech. Language and Speech. 2004; 47:31–56. [PubMed: 15298329]
- Baayen, RH. Analyzing linguistic data. A practical introduction to statistics using R. Cambridge University Press; Cambridge: 2008.

- Baayen RH, Davidson DH, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Baker RE, Bradlow AR. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*. 2009; 52(4):391–413. [PubMed: 20121039]
- Baker RE, Bradlow AR. Second mention reduction in Indian English and Korean (A). *Journal of the Acoustical Society of America*. 2007; 122(5):2993.
- Bard EG, Anderson AH, Sotillo C, Aylett M, Doherty-Sneddon G, Newlands A. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*. 2000; 42:1–22.
- Bell, A.; Gregory, ML.; Brenier, J.; Jurafsky, D.; Ikeno, A.; Girand, C. Which predictability measures affect content word durations?. In: Byrne, W.; Fosler-Lussier, E.; Jurafsky, D., editors. *Workshop on Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology (PMLA)*; Estes Park, CO, USA. 2002. p. 65-70.
- Bell A, Jurafsky D, Fosler-Lussier E, Girand C, Gregory M, Gildea D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*. 2003; 113:1001–1024. [PubMed: 12597194]
- Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*. 2009; 60:92–111.
- Bybee, J. *Phonology and language use*. Cambridge University Press; Cambridge: 2001.
- Bybee J. From usage to grammar: The mind's response to repetition. *Language*. 2006; 82:711–33.
- Davies, M. BYU-BNC: The British National Corpus. 2004-. Retrieved June 15, 2009, from <http://corpus.byu.edu/bnc>
- Derwing TM, Munro MJ. Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*. 1997; 19:1–16.
- Flege JE, Bohn O-S. An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition*. 1989; 11:35–62.
- Fokes J, Bond ZS. The vowels of stressed and unstressed syllables in nonnative English. *Language Learning*. 1989; 3:341–373.
- Fowler CA, Housum J. Talkers' signaling of “new” and “old” words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*. 1987; 26:489–504.
- Fowler CA. Differential shortening of repeated context words produced in various communicative contexts. *Language and Speech*. 1988; 31:307–319. [PubMed: 3271893]
- Gahl S. Time and Thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*. 2008; 84(3):474–496.
- Gahl S, Garnsey SM. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*. 2004; 80:748–775.
- Garrett, MF. Levels of processing in sentence production. In: Bower, GH., editor. *Language production*. Vol. 1. Academic Press; London: 1980. p. 177-220.
- Guion SG, Flege JE, Liu SH, Yeni-Komshina GH. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*. 2000; 21:205–228.
- Gumperz, JJ. *Discourse strategies*. Cambridge University Press; Cambridge: 1982.
- Hawkins S, Warren P. Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*. 1994; 22:493–511.
- Iwasaki, S. *Japanese*. John Benjamins Publishing Co.; Amsterdam: 2002.
- Jun, S-A. *The Phonetics and Phonology of Korean Prosody* (Doctoral dissertation). The Ohio State University; Columbus, OH: 1993.
- Jurafsky, D.; Bell, A.; Gregory, M.; Raymond, WD. Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J.; Hopper, P., editors. *Frequency and the emergence of linguistic structure*. John Benjamins Publishing Co.; Amsterdam: 2001. p. 229-254.
- Kim M. Discourse markers in conversations between native and non-native speakers. In prep.
- Ladd, DR. *Intonational Phonology*. Cambridge University Press; Cambridge, UK: 1996.

- Lapointe, SG.; Dell, SG. A synthesis of some recent work in sentence production. In: Carlson, GN.; Tanenhaus, MK., editors. *Linguistic structure in language processing*. Kluwer; Dordrecht: 1989. p. 107-156.
- Lavoie L. Some influences on the realization of for and four in American English. *Journal of the International Phonetic Association*. 2002; 32:175–202.
- Lee B, Guion SG, Harada T. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*. 2006; 28:487–513.
- Lee, I.; Ramsey, SR. *The Korean language*. State University of New York Press; Albany: 2000.
- Lennon P. Investigating fluency in EFL: A quantitative approach. *Language Learning*. 1990; 40:387–417.
- Li, CN.; Thompson, SA. *Mandarin Chinese: A functional reference grammar*. University of California Press; Berkeley: 1989.
- Lindblom, B. Explaining phonetic variation: A sketch of the H and H theory. In: Hardcastle, W.; Marchal, A., editors. *Speech production and speech modeling*. Kluwer; Dordrecht: 1990. p. 403-439.
- Mochizuki-Sudo M, Kiritani S. Production and perception of stress-related durational patterns in Japanese learners of English. *Journal of Phonetics*. 1991; 19:231–248.
- Munro MJ. A primer on accent discrimination in the Canadian context. *TESL Canada Journal*. 2003; 20(2):38–51.
- Munro MJ, Derwing TM. Processing time, accent, and comprehensibility in the perception of foreign-accented speech. *Language and Speech*. 1995; 38:289–306. [PubMed: 8816082]
- Navarrete E, Basagni B, Alario FX, Costa A. Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology*. 2006; 59(10):1681–1690. [PubMed: 16945853]
- Neumeyer L, Franco H, Digalakis V, Weintraub M. Automatic scoring of pronunciation quality. *Speech Communication*. 1999; 30:83–93.
- Pan, S.; McKeown, KR. *Proceedings of the EMNLP/VLC'99*. University of Maryland; College Park (MD), USA: 1999. Word informativeness and automatic pitch accent modeling.
- Pierrehumbert, JB. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J.; Hopper, P., editors. *Frequency and the emergence of linguistic structure*. John Benjamins Publishing Co; Amsterdam: 2001. p. 137-57.
- Pierrehumbert, JB. Word-specific phonetics. In: Gussenhoven, C.; Warner, N., editors. *Laboratory phonology 7*. Mouton; Berlin/NewYork: 2002. p. 101-139.
- Pluymaekers M, Ernestus M, Baayen RH. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*. 2005; 118(4):2561–2569. [PubMed: 16266176]
- Rasier L, Hiligsmann P. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de linguistique Française*. 2007; 28:41–66.
- Sailaja, P. *Indian English*. Edinburgh University Press; Edinburgh: 2009.
- Schmitt N, Dunham B. Exploring native and non-native intuitions of word frequency. *Second Language Research*. 1999; 15(4):389–411.
- Selkirk, E. The prosodic structure of function words. In: Morgan, JL.; Demuth, K., editors. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Lawrence Erlbaum Associates; Mahwah, NJ: 1996. p. 187-214.
- Shah, AP. Production and perceptual correlates of Spanish-accented English. In: Slifka, J.; Manuel, S.; Matthies, M., editors. *From sound to sense: 50+ years of discoveries in speech communication*. 2004. p. 79-84. Retrieved 19 April, 2009 from <http://www.rle.mit.edu/soundtosense/conference/starthere.htm>
- Shi R, Gick B, Kanwischer D, Wilson I. Frequency and category factors in the reduction and assimilation of function words: EPG and acoustic measures. *Journal of Psycholinguistic Research*. 2005; 34(4):341–364. [PubMed: 16142587]

- Shi R, Morgan JL, Allopenna P. Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language*. 1998; 25:169–201. [PubMed: 9604573]
- Shields LW, Balota DA. Repetition and associative context effects in speech production. *Language and Speech*. 1991; 34:47–55. [PubMed: 1819682]
- Tajimi K, Port R, Dalby J. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*. 1997; 25:1–24.
- Taylor P, Black AW. Assigning phrase breaks from part-of-speech sequences. *Computer Speech and Language*. 1998; 12:99–117.
- Tiffen, B. A study of the intelligibility of Nigerian English. In: van Essen, A.; Burkart, EI., editors. *In Homage to W. R. Lee: Essays in English as a foreign or second language*. Foris; Berlin: 1992. p. 255-259.
- Ueyama, M. *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese* (Doctoral Dissertation). University of California; Los Angeles, L.A., CA: 2000.
- Van Bergen DR. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*. 1993; 12(1):1–23.
- Van Engen KJ, Baese-Berk M, Baker RE, Choi A, Kim M, Bradlow AR. The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*. In press.
- Weinberger, SH. The Speech Accent Archive. George Mason University; <http://accent.gmu.edu/>
- Wilson MD. The MRC Psycholinguistic Database: Machine readable dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*. 1988; 20(1):6–11.
- Zhao Y, Jurafsky D. The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*. 2009; 37:231–247.

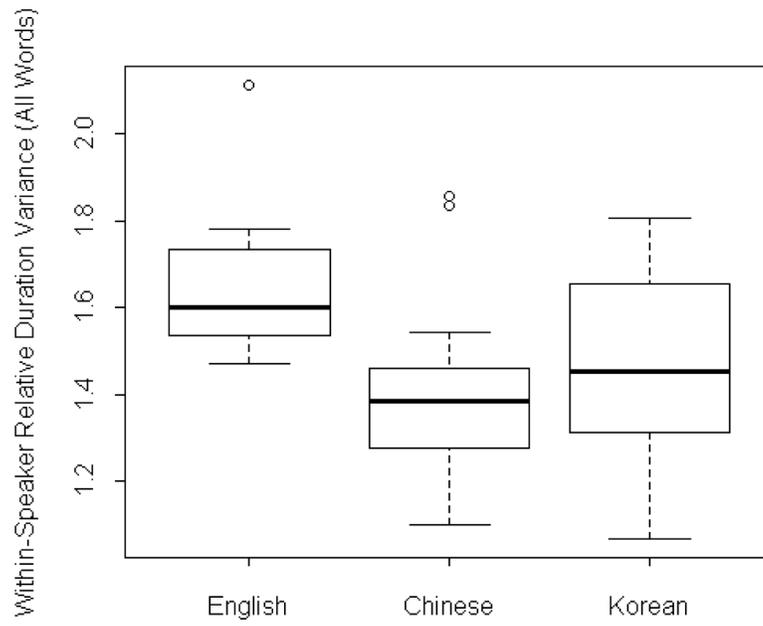


Figure 1. Comparison of within-speaker relative word duration variance for all words in the English, Chinese, and Korean language groups

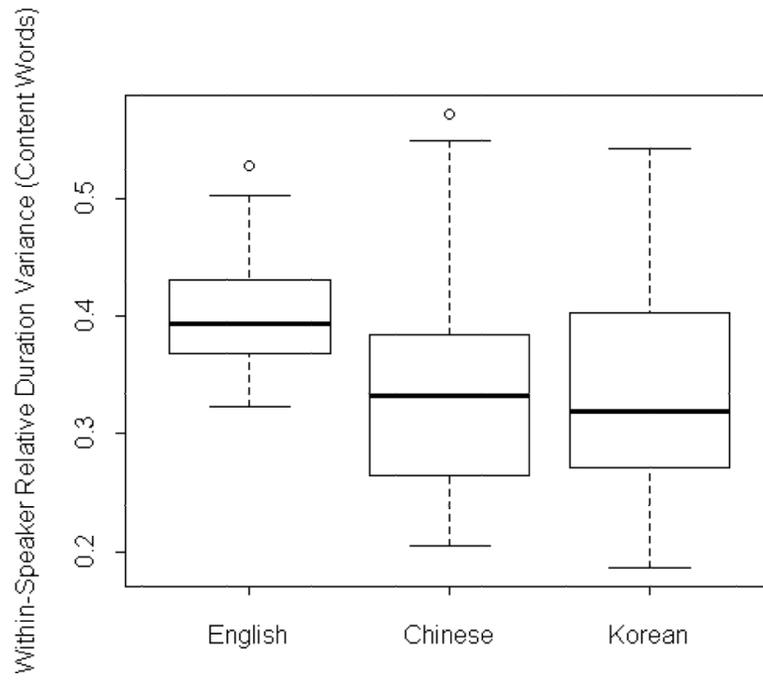


Figure 2. Comparison of within-speaker relative word duration variance for content words in the English, Chinese, and Korean language groups

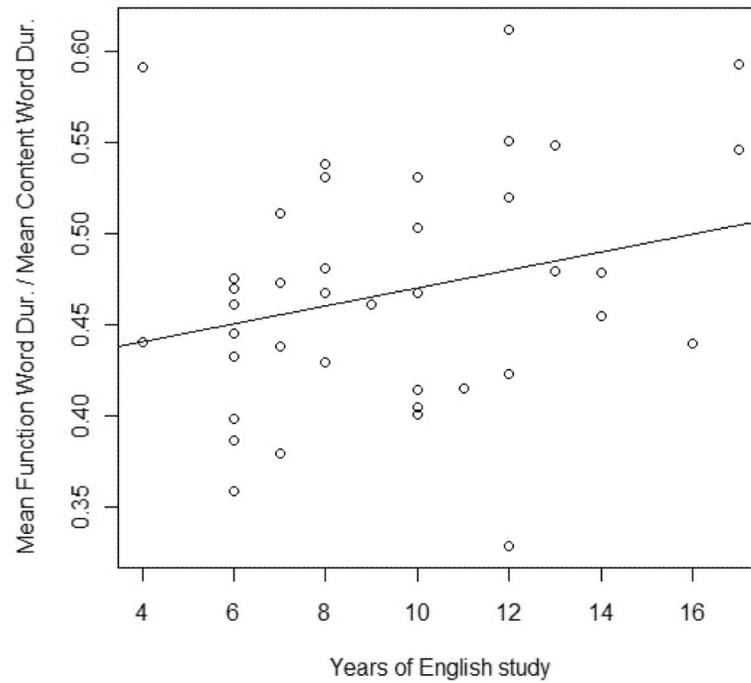


Figure 3. Scatterplot of the relationship between years of English study and mean function word duration divided by mean content word duration for non-native participants

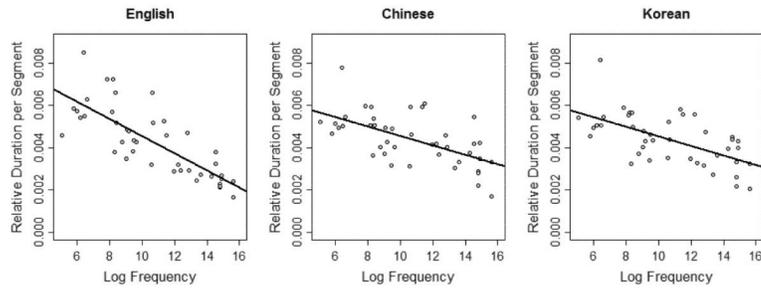


Figure 4.

Scatterplots showing the relationship between relative duration per segment and log word frequency in the English, Chinese, and Korean language groups. Relative durations per segment were calculated by dividing the mean relative duration for each word by the number of phonemes in the word (as listed in the MRC psycholinguistic database).

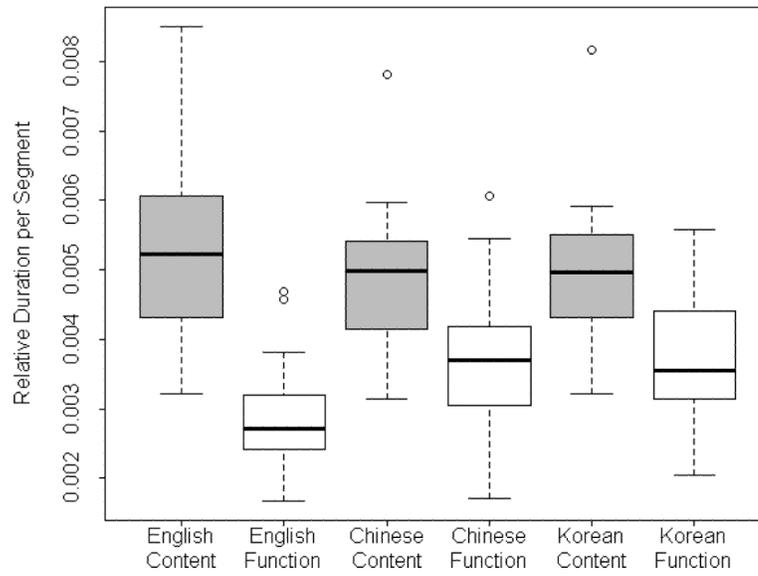


Figure 5. Comparison of relative durations per segment in content and function words for the English, Chinese, and Korean speaker groups. Relative durations per segment were calculated by dividing the mean relative duration for each word by the number of phonemes in the word (as listed in the MRC psycholinguistic database).

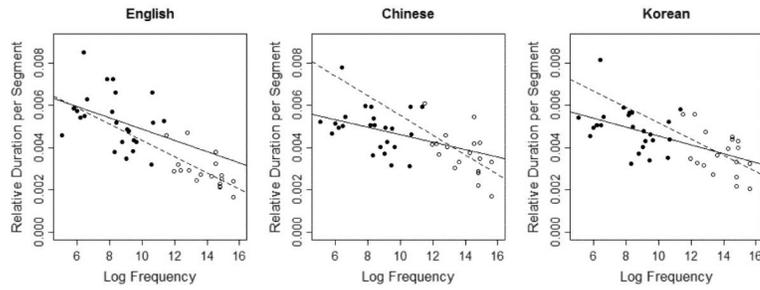
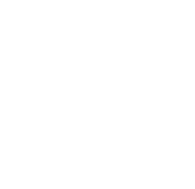


Figure 6.

Scatterplots showing the relationship between relative duration per segment and log word frequency for content and function words in the English, Chinese, and Korean language groups. Relative durations per segment were calculated by dividing the mean relative duration for each word by the number of phonemes in the word (as listed in the MRC psycholinguistic database). In these plots, the filled circles and solid lines represent content words, while the empty circles and dashed lines represent function words.



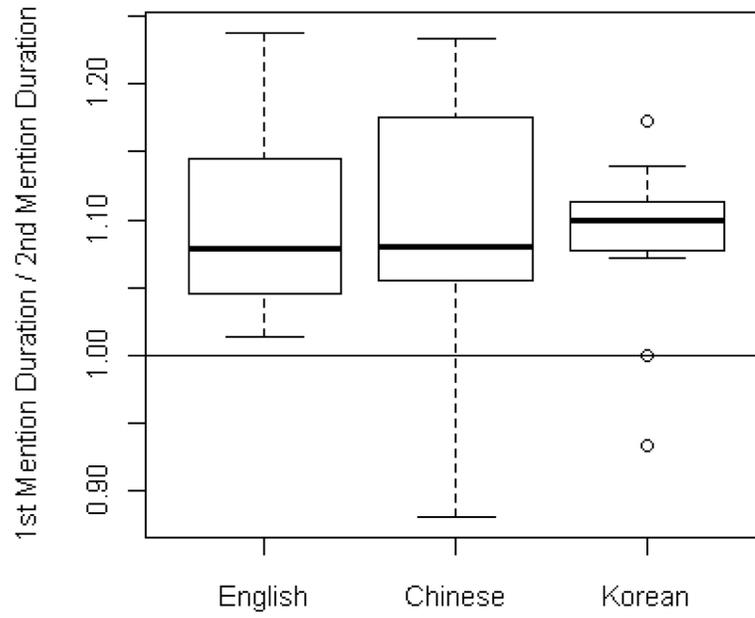


Figure 7. Ratios of first mention mean durations divided by second mention mean durations for the English, Chinese, and Korean language groups. Everything above the line at 1 represents reduction.

Table 1

Range, mean, and standard deviation (in parentheses) of 1) the age at which participants began their formal study of English, 2) the number of years they had formally studied English, and 3) the number of months they had spent in English speaking countries, grouped by native language.

	Age when formal English study began	Years of formal English study	Months in English-speaking country
Korean	10–14, mean=12.5 (1.2)	4–17, mean=8.3 (3.3)	1–60, mean=19 (17.6)
Chinese	9–16, mean=11.5 (1.6)	4–17, mean=10.7 (3.1)	1–132, mean=12.9 (30.6)

Table 2

Parameter estimates and associated confidence intervals and *p*-values for fixed, noncontrol, effects in the native main effect regression on word duration. The confidence intervals and *p*-values were obtained by MCMC sampling (Baayen, Davidson, & Bates, 2008).

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	-0.0995	-0.1766	-0.0350	<0.01
Log Frequency	-0.0276	-0.0337	-0.0171	<0.0005
Word Type	0.0145	-0.0056	0.0431	0.1248

Table 3

Parameter estimates and associated confidence intervals and *p*-values for fixed, noncontrol, effects in the non-native main effect Stella regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	-0.1722	-0.2466	-0.1027	<0.0005
Log Frequency	-0.0326	-0.0433	-0.0209	<0.0005
Word Type	-0.0078	-0.0399	0.0259	0.7074
L1	0.0007	-0.0041	0.0053	0.7832
Age English study began	0.0012	-0.0018	0.0043	0.4684
Years of English study	0.0000	-0.0013	0.0014	0.9508
Months in English speaking country	0.0000	-0.0002	0.0002	0.8070
Studying English at time of experiment	-0.0007	-0.0049	0.0038	0.7524

Table 4

Parameter estimates and associated confidence intervals and *p*-values for fixed, noncontrol, effects (with interactions between two lexical and five L1 and L2 experience factors) in the non-native two-way interaction Stella regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	-0.1746	-0.2447	-0.1023	<0.0005
Log Frequency	-0.0295	-0.0415	-0.0158	<0.0005
Word Type	0.0098	-0.0268	0.0490	0.5718
L1	0.0005	-0.0043	0.0056	0.8278
Age English study began	0.0018	-0.0014	0.0050	0.2678
Years of English study	0.0004	-0.0010	0.0018	0.5884
Months in English speaking country	0.0000	-0.0002	0.0002	0.9016
Studying English at time of experiment	-0.0008	-0.0052	0.0037	0.7222
Log Frequency : L1	0.0010	-0.0019	0.0041	0.5328
Log Frequency : Age English study began	-0.0007	-0.0026	0.0013	0.4992
Log Frequency : Years of English study	-0.0004	-0.0013	0.0005	0.3276
Log Frequency : Months in English speaking country	-0.0001	-0.0002	0.0000	0.2140
Log Frequency : Studying English at time of experiment	0.0000	-0.0027	0.0028	0.9812
Word Type : L1	0.0010	-0.0087	0.0103	0.8260
Word Type : Age English study began	-0.0041	-0.0102	0.0021	0.1970
Word Type : Years of English study	-0.0030	-0.0060	-0.0004	<0.05
Word Type : Months in English speaking country	0.0001	-0.0003	0.0004	0.6682
Word Type : Studying English at time of experiment	0.0007	-0.0085	0.0095	0.8770

Table 5

Parameter estimates and associated confidence intervals and *p*-values for fixed, noncontrol, effects in the native/non-native two-way interaction Stella regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	-0.1453	-0.2229	-0.0711	<0.0005
Log Frequency	-0.0313	-0.0421	-0.0176	<0.0005
Word Type	-0.0030	-0.0355	0.0374	0.9616
Nativeness	-0.0014	-0.0075	0.0043	0.6392
Log Frequency : Word Type	0.0038	-0.0091	0.0144	0.6964
Log Frequency : Nativeness	0.0048	0.0022	0.0074	<0.001
Word Type: Nativeness	0.0131	0.0050	0.0216	<0.005

Table 6

Parameter estimates and associated confidence intervals and *p*-values for fixed, noncontrol, effects in the native/non-native three-way interaction Stella regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i>-value
Intercept	-0.1501	-0.2304	-0.0810	<0.0005
Log Frequency	-0.0306	-0.0415	-0.0169	<0.0005
Word Type	-0.0005	-0.0355	0.0394	0.8482
Nativeness	-0.0103	-0.0207	-0.0004	<0.05
Log Frequency : Word Type	0.0021	-0.0109	0.0125	0.9144
Log Frequency : Nativeness	0.0061	0.0033	0.0089	<0.0005
Word Type: Nativeness	0.0177	0.0081	0.0268	<0.0005
Log Frequency : Word Type : Nativeness	-0.0030	-0.0060	-0.0003	<0.05

Table 7

Parameter estimates and associated confidence intervals and *p*-values for fixed, noncontrol, effects in the native/non-native content word Stella regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	-0.2219	-0.3144	-0.1347	<0.0005
Log Frequency	-0.0294	-0.0438	-0.0159	<0.0005
Nativeness	0.0183	0.0072	0.0292	<0.005
Log Frequency : Nativeness	0.0030	-0.0001	0.0064	0.0636

Table 8

Parameter estimates and associated confidence intervals and *p*-values for fixed, non-control, effects in the native/non-native function word Stella regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	-0.0490	-0.1884	0.0941	0.4524
Log Frequency	-0.0297	-0.0477	-0.0093	<0.005
Nativeness	-0.0425	-0.0583	-0.0251	<0.0005
Log Frequency : Nativeness	0.0091	0.0047	0.0133	<0.0005

Table 9

Parameter estimates and associated confidence intervals and p -values for fixed, non-control, effects in the native main effect Gina's Pizza regression on word duration.

	Estimate	95% CI lower	95% CI upper	p -value
Intercept	0.0536	-0.0392	0.1489	0.2344
Mention	-0.0332	-0.0457	-0.045	<0.0005

Table 10

Parameter estimates and associated confidence intervals and *p*-values for fixed, non-control, effects in the non-native main effect Gina's Pizza regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	0.0773	0.0002	0.1495	<0.05
Mention	-0.0416	-0.0543	-0.0278	<0.0005
L1	-0.0004	-0.0078	-0.0078	0.9322
Age English study began	0.0004	-0.0045	0.0053	0.8952
Years of English study	-0.0007	-0.0030	0.0015	0.5508
Months in English speaking country	0.0001	-0.0002	0.0004	0.5292
Studying English at time of experiment	0.0000	-0.0069	0.0074	0.9866

Table 11

Parameter estimates and associated confidence intervals and *p*-values for fixed, non-control, effects in the non-native two-way interaction Gina's Pizza regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	0.0796	0.0093	0.1583	<0.05
Mention	-0.0431	-0.0568	-0.0295	<0.0005
L1	-0.0117	-0.0341	0.0121	0.3278
Age English study began	0.0073	-0.0084	0.0220	0.3516
Years of English study	-0.0062	-0.0136	0.0004	0.0784
Months in English speaking country	0.0005	-0.0004	0.0014	0.3002
Studying English at time of experiment	-0.0113	-0.0341	0.0100	0.3184
Mention : L1	0.0076	-0.0071	0.0221	0.3146
Mention : Age English study began	-0.0046	-0.0138	0.0056	0.3570
Mention : Years of English study	0.0037	-0.0007	0.0082	0.1012
Mention : Months in English speaking country	-0.0002	-0.0008	0.0003	0.3782
Mention : Studying English at time of experiment	0.0075	-0.0066	0.0213	0.2922

Table 12

Parameter estimates and associated confidence intervals and *p*-values for fixed, non-control, effects in the native/non-native two-way interaction Gina's Pizza regression on word duration.

	Estimate	95% CI lower	95% CI upper	<i>p</i> -value
Intercept	0.0679	0.0068	0.1290	<0.05
Mention	-0.0374	-0.0487	-0.0255	<0.0005
Nativeness	-0.0061	-0.0252	0.0136	0.5460
Mention: Nativeness	0.0042	-0.0075	0.0154	0.4856