Published in final edited form as:

Lang Cogn Neurosci. 2018 ; 33(2): 211–226. doi:10.1080/23273798.2017.1375129.

The role of linguistic experience in the processing of probabilistic information in production

Erin Gustafson and Matthew Goldrick*

Department of Linguistics, Northwestern University, 2016 Sheridan Rd., Evanston IL, 60208 USA

Abstract

Speakers track the probability that a word will occur in a particular context and utilize this information during phonetic processing. For example, content words that have high probability within a discourse tend to be realized with reduced acoustic/articulatory properties. Such probabilistic information may influence L1 and L2 speech processing in distinct ways (reflecting differences in linguistic experience across groups and the overall difficulty of L2 speech processing). To examine this issue, L1 and L2 speakers performed a referential communication task, describing sequences of simple actions. The two groups of speakers showed similar effects of discourse-dependent probabilistic information on production, suggesting that L2 speakers can successfully track discourse-dependent probabilities and use such information to modulate phonetic processing.

Keywords

language production; L2 speech; phonetics; reduction

Inter- as well as intra-speaker factors shape variation in the acoustic-phonetic properties of words. Striking differences in speech production behaviors, including the acoustic-phonetic characteristics of productions, can be observed when comparing native, first language (L1) speakers of a language to non-native, second language (L2) speakers. For example, L2 speakers produce overall longer word durations than L1 speakers (Munro & Derwing, 1995). Variation in the duration of speech can also be observed within an individual. Content words that have high probability within a particular discourse are typically planned more quickly (e.g., Kahn & Arnold, 2012) and are reduced in duration (e.g., Fowler & Housum, 1987) relative to low probability words.

In the current work, we examine how the processing of such probabilistic information across word classes is influenced by linguistic experience. Because L2 speakers have less experience with a language than L1 speakers, L2 speech processing is more demanding, and L2 speakers may lack knowledge of key aspects of the L1 (e.g., they may have not yet mastered the prosodic structure of the L2). Such factors may prevent L2 speakers from utilizing probabilistic information during planning and articulation for both content and function words. We begin by discussing how probabilistic information influences processing

*Corresponding author; +1-847-491-8053.

and how it may impact content and function word processing differently. Next we turn to existing work on the influence of probabilistic information on processing during content (specifically, noun) and function (specifically, determiner) word production by both L1 and L2 speakers. Finally, we consider how differences among L1 and L2 speakers may shed light on the mechanisms underlying the processing of probabilistic information.

The Processing of Probabilistic Information in Speech Production

The probability of some linguistic unit can have a demonstrable influence on word-form processing in production. The current study focuses on the inverse relationship between the probabilities of individual words and their phonetic prominence. Acoustic-phonetic reduction of some kind, whether it be in the duration of the stressed vowel of the word, the spectral qualities of the vowel, or the duration of the word itself, has been associated with high vs. low probability words (e.g., Aylett & Turk, 2004; Baker & Bradlow, 2009; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Fowler & Housum, 1987; Gahl, 2008; Gahl & Garnsey, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001; Kahn & Arnold, 2015; Lam & Watson, 2010; Liberman, 1963; Munson & Solomon, 2004; Pate & Goldwater, 2015; Scarborough, 2010). Reduction in the time required to plan a word or utterance (indexed by response time, or RT) is also impacted by word-specific probability (e.g., Kahn & Arnold, 2012, 2015). In the current study, we examine the association of both measures – planning time and acoustic-phonetic reduction – to discourse-dependent probability (categorically manipulated in terms of discourse-given vs. discourse-new status).

Processing of Content vs. Function Words

A fundamental contrast between lexical items that has long been noted in linguistic theories is between semantically rich, but syntactically weak, content words and syntactically rich, but semantically impoverished, function words (although this dichotomy is likely an oversimplification; c.f. Altmann, Pierrehumbert, & Motter, 2009). Theories of word production typically consider these word classes separately (Levelt, Roelofs, & Meyer, 1999; but see Arnon & Snider, 2010, for results and a theoretical framework that allows for holistic representations of multi-word strings). Some theories go as far as to say that determiners are not independent lexical entities. According to such theories, function words are retrieved as part of some syntactic structure, or frame, and thus do not undergo lexical selection processes (as content words do; Garrett, 1975). In contrast, more recent theories argue that function words have lexical representations and undergo similar selection processes as for content words (Bürki, Laganaro, & Alario, 2014; Janssen, Schiller, & Alario, 2014; Jescheniak, Schriefers, & Lemhöfer, 2014). Under either type of theory, the selection of a determiner depends upon (or, occurs after) selection of the noun phrase's head noun (Bock & Levelt, 1994), as the specific determiner to be selected often depends upon the form of the following noun (e.g., *a* vs. *an* in English, or *le* vs. *la* vs. *l'* in French; Caramazza, Miozzo, Costa, Schiller, & Alario, 2001).

While psycholinguistic theories of production typically assume that function and content words share common phonological encoding and phonetic implementation processes following selection (Lapointe & Dell, 1989), evidence from L1 English speech suggests that

function and content words differ in their phonological form and prosodic implementation. Unlike content words, which are typically stressed, (monosyllabic) function words tend to be unstressed unless produced in isolation (Selkirk, 1996) or when following a disfluency (Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003). Due to their “weak” phonological form, function words cannot form a foot (a prosodic structure made up of at least one strong, accented syllable) or a prosodic word (a prosodic structure consisting of at least one foot). Therefore, without a prosodic word of their own, function words cliticize to an adjacent content word to form a single prosodic word (Selkirk, 1996).

While L1 and L2 speech production are assumed to rely on the same types of selection processes for all word classes, L2-specific factors influence the dynamics of these processes. For content word production, L2 speech is often found to have overall longer word durations than L1 speech (Baker, Baese-Berk, Bonnasse-Gahot, Kim, Van Engen, & Bradlow, 2011; Guion, Flege, Liu, & Yeni-Komshian, 2000; Munro & Derwing, 1995). This across-the-board slowing can be attributed to two non-mutually exclusive sources: cross-language interference or impoverished linguistic knowledge.

According to the cross-language interference account, both L1 and L2 lexical representations are automatically activated in parallel during L2 speech production (e.g., Colomé, 2001; Costa, Caramazza, & Sebastian-Galles, 2000; Colomé & Miozzo, 2010; Hermans, Bongaerts, De Bot, & Schreuder, 1998). Delays in both speech planning (e.g., Hermans et al., 1998) and articulation (Sadat, Martin, Alario, & Costa, 2012) are consequences of this processing difficulty attributed to cross-language interference.

Additional processing difficulty could arise because L2 speakers necessarily have less experience with the language compared to L1 speakers. One proposal, the frequency lag hypothesis (Gollan, Slattery, Goldenberg, Van Assche, Duyck, & Rayner, 2011), argues that L2 speakers have accrued lower frequency counts for all words than L1 speakers due to their lower level of exposure to the language. Slower speech articulations for L1 compared to L2 speakers could, therefore, be due to the L2 frequency lag, given that lower frequency words have longer word durations (e.g., Bell et al., 2009).

Duration differences between L1 and L2 speech also manifest in function word production. A number of studies have found that L2 speakers fail to reduce function words to the same degree as L1 speakers (Aoyama & Guion, 2007; Baker et al., 2011). Deficits are found even for speakers who would reduce function words in their own L1 (e.g., Mandarin; Shi, Morgan, & Allopenna, 1998), although in such cases lack of mastery of the L2 prosodic system might contribute to difficulties with reduction (Baker et al., 2011). However, it is worth noting that the function word differences may not be uniform over this entire class of words. For example, in Baker et al. (2011) the L2 speakers reduced *the* (the third highest frequency (function) word in English) to the same degree as L1 speakers. This suggests that L2 speakers accrue enough experience with high frequency words to reduce them to an L1-like degree (consistent with the frequency lag hypothesis). With enough experience with a particular word, L2 speakers may then be able to overcome general issues with English prosody.

Altogether, across L1 and L2 speech, we can see marked differences in the functional properties (i.e., semantically vs. syntactically focused) and prosodic implementations of content and function words. As discussed in more detail below, studies also observe distinct influences of probabilistic information across these word classes. In the following sections, we discuss these diverging results and consider possible differences that might be present across L1 and L2 speakers stemming from general differences in speech production behavior across groups.

Effects of Variation in Content Word Probabilities

L1 speech

Influences of probabilistic information on content word production have been well-established in the literature for L1 speech production. Bell et al. (2009) analyzed a corpus of spontaneous, conversational speech and found that various types of probabilistic information influenced word duration. Words with higher lexical frequency and higher preceding/following conditional probability (i.e., the probability of a word given the previous or following word) were reduced relative to lower frequency/probability words. Critically, controlling for these factors, content words with high discourse-dependent probability (i.e., discourse-given words) were significantly reduced compared to words with low discourse-dependent probability (i.e., discourse-new words). Robust effects of discourse-dependent reduction have also been observed in experimental tasks that simulate spontaneous conversation and/or create a discourse within the context of the experiment, such as interactive map tasks (Aylett & Turk, 2004; Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon, & Newlands, 2000; Meagher & Fowler, 2014), referential communication tasks (Jacobs, Yiu, Watson, & Dell, 2015; Kahn & Arnold, 2012, 2015; Lam & Watson, 2010, 2014), and paragraph reading (Baker & Bradlow, 2009; Baker et al., 2011).

Note that many of these studies have suggested that reduction should not be attributed to explicit computation of discourse-dependent probabilities, but rather to facilitation during planning and execution that occurs due to priming from the recently repeated nouns (Arnold & Watson, 2012; Fraundorf, Watson, & Benjamin, 2014; Kahn & Arnold, 2012; Lam & Watson, 2010, 2014; c.f., Fowler, 1988). We return to this issue below.

L2 speech

Cross-language interference and frequency lag, the two possible sources of L2 slowing during content word production, could also lead to deficits in how probabilistic information influences L2 processing. In fact, the exposure-driven frequency lag hypothesis is motivated by empirical findings that L2 speakers exhibit probabilistic effects differing in magnitude than effects from L1 speakers. For example, variation in lexical frequency (a type of word-specific probability) has a larger effect on RTs for L2 compared to L1 speakers, both in speech production and perception (e.g., Diependaele, Lemhöfer, & Brysbaert, 2013; Gollan, Montoya, Cera, & Sandoval, 2008; Gollan et al., 2011; Hernández, Costa, & Arnon, 2016), which indicates differences in how probabilistic processes influence L1 vs. L2 speech.

However, such differences have not always been found, particularly in phonetic measures. Differences in frequency's influence on noun durations have not been reliably observed (Baker et al., 2011; Sadat et al., 2012). For discourse-dependent reduction in word durations, Baker et al. (2011) also did not find reliable differences between groups; L2 speakers reduced discourse-given content words to the same degree as L1 speakers in read speech. This result suggests that although cross-language interference or low levels of experience may impact overall word durations and frequency effects on planning, the influence of discourse-dependent probabilities on processing remains unimpaired during L2 speech production. Therefore, it seems to be the case that L2 speakers also track the discourse status of words and reduce appropriately (for further discussion of reduction of nouns in L2 speech, see Lam & Marian, 2015). Such a null result is consistent with the view that reduction reflects priming (e.g., Lam & Watson, 2014). Since this mechanism is shared across L1 and L2 speech, both types of speakers should exhibit comparably levels of reduction.

However, a key issue unaddressed in this work is that L2 effects on word durations appear to be mitigated in less demanding production tasks, such as single word repetition compared to picture naming (Gustafson, Engstler, & Goldrick, 2013). Therefore, the seemingly unimpaired processing of probabilistic information by L2 speakers in Baker et al. (2011) could be attributed to the paragraph reading task utilized in that study. This makes the clear prediction that deficits in processing probabilistic information will reliably be observed when L2 speakers engage in a more demanding task that requires them to generate full sentences without a written prompt (e.g., an event description task; Lam & Watson, 2010). The increased processing demands of such a task could impede the ability of L2 speakers to track the probability of words within a discourse, which could lead to differences between L1 and L2 speakers in the magnitude of discourse-dependent probabilistic reduction produced. We test this hypothesis in the current study.

Effects of Variation in Function Word Probabilities

L1 speech

In addition to content word reduction, Bell et al. (2009) also considered the influence of different sorts of word-specific probabilities on function word production. Notably, unlike for content words, they found that function words did not undergo discourse-dependent reduction. However, function word reduction is strongly predicted by multi-word probability (i.e., probability conditioned on either the previous or following word; Bell et al., 2003; Bell et al., 2009; Jurafsky et al., 2001). For example, Bell et al. (2003) found a strong relationship between the duration of determiners and the conditional probability of the determiner given the following word. Therefore, it could be the case that discourse-dependent probabilistic reduction only occurs on function words when they modify nouns with high discourse-dependent probability. We refer to this as the *probability inheritance hypothesis*. This is supported by previous results showing significant reduction of function words in discourse-given vs. discourse-new trials in an event description task (Kahn & Arnold, 2012, 2015).

As with reduction of content words, such effects could be attributed specifically to processing of probabilistic structure. If production takes into account the probability of the

determiner conditioned on the following noun, the determiner could ‘inherit’ the high discourse-dependent probability of the noun. Alternatively, as with the reduction of content words, the reduction of determiner durations could be attributed to priming from the recently repeated nouns (e.g., Kahn and Arnold, 2012).

L2 speech

Little research has considered how probabilistic information affects L2 function word production. Schertz and Ernestus (2014) measured the duration of determiner vowels produced by Norwegian and Czech speakers of English during spontaneous speech. The results showed that the Czech, but not Norwegian, speakers produced shorter determiner vowels when followed by more vs. less frequent nouns. While this provides some (mixed) evidence that L2 English speakers reduce function words based on the probability of the content words that they modify, it is not clear if L2 and L1 speakers differ in the degree of reduction that they produce. Within the task outlined above, we will be able to directly compare the degree of reduction across these two speaker groups.

The Current Study: Summary

The current study investigates the presence or absence of L2 deficits in how discourse-dependent probabilistic information influences processing during a demanding speech production task. We address this using an event description task, in which speakers describe animations of a series of pictures (e.g., *The candle rotates*). We compare indices of planning and execution (RTs and word durations, respectively) across discourse-new (the target word is produced only once within a trial) and discourse-given (the target word is produced twice) trials. A decrease in the magnitude of discourse-dependent reduction in RTs, content word, and function word durations speaks to the presence of difficulty in the processing of probabilistic information during L2 speech.

Method

Participants

A total of 64 speakers participated in this study. All work was conducted with the formal approval of the Institutional Review Board of Northwestern University. Participants were divided into two groups: native speakers of American English (henceforth, the L1 group) and native speakers of Mandarin who learned English as a second language (henceforth the L2 group). Speakers comprising the L1 group ($N = 25$) were recruited from the linguistics participant pool at Northwestern University and received partial course credit for their participation. All participants in this group were L1 speakers of English with no history of speech impairments or color blindness. One participant was excluded due to poor equipment performance in a companion study not discussed here, leaving 24 native speakers of American English (17 female; mean age: 19, range: 18-21) for the L1 group.

Many speakers included in the L2 group were recruited from the International Summer Institute at Northwestern University, a month-long program for incoming international students that offers intensive English instruction and one-on-one tutoring prior to the start of their first academic quarter as graduate students. Other participants were recruited from the

Northwestern community via flyers and a database of current and former students in Northwestern's English Language Learners Program. Each of these participants was compensated \$10/hour. Two participants were recruited via the linguistics department participant pool, and thus received partial course credit for their participation.

Thirty-nine individuals whose L1 was Mandarin participated as part of the L2 group. Eleven participants were excluded who were unable to produce at least 70% of the target items name during the experiment. This criterion was set to ensure 1) sufficient statistical power and 2) sufficient levels of English proficiency. An additional three participants were excluded due to poor equipment performance in a companion study, and one English-dominant speaker was excluded to ensure comparable linguistic experience across the participants. The remaining 24 participants (18 female; mean age: 23.2, range: 18-31) were Mandarin-dominant, L2 speakers of English who did not learn English at home (i.e., English exposure began at school).

Each participant completed a detailed language background questionnaire. In this questionnaire, participants were asked to provide information about their exposure and experience with all languages they spoke. Table 1 reports information provided by participants that summarizes variation in linguistic experience across groups.¹ The LexTale vocabulary test (an unspeeded lexical decision task; Lemhöfer & Broersma, 2012) was used as an objective measure of English proficiency.

Materials and Design

The stimuli were a set of 48 pictures taken from the Bank of Standardized Stimuli (BOSS; Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and other sources as needed. The BOSS database includes a large set of full color photographs that have been normed along a number of dimensions, including name agreement, category, familiarity, and visual complexity. The names for the target pictures have high name agreement by L1-English speakers (mean = 94.6%, min = 63.2%, sd = 9.2%), as established in a separate norming study with 19 participants. The names also have above at least 40% name agreement by L2-English speakers from the same population as the participants for the current study, also established in a separate norming study with 10 participants (mean = 77.9%, sd = 20.3%).

The target items comprise 24 pairs in which the items overlap by least two phonemes (e.g., /kæn-/ in *candy* and *candle*) and have the same number of syllables. (n.b. This pairing is utilized in ongoing perception studies in our lab; no experimental conditions in this production study were sensitive to this relationship among target items.) Each of these pairs was assigned three other items that served as non-target distractors. The non-targets were not semantically or phonologically related to either target in the pair. In addition to the 48 target and 72 non-target pictures, 144 filler pictures were selected from BOSS. Care was taken to avoid overlap with the target and non-target pictures, and to select images that L2 speakers

¹A series of analyses considering the influence of these proficiency variables on performance revealed no reliable effects. Given the small sample size of the groups and the substantial variation in the dependent measure we draw no strong conclusions from these null effects.

were likely to be able to name successfully (e.g., *bird* but not *saxophone*). See Appendix A for the full set of stimuli, with name agreement and frequency data for each target item.

The participants' task was to describe events presented on a computer screen. At the start of a trial, an array of eight pictures appeared on the screen (see Figure 1). On target trials, the array included the target picture and one, two, or all three of the assigned non-target distractors. The array of pictures did not include the other target item in the pair to guard against influences of phonological overlap from phonologically related competitors on target productions (Meyer & Damian, 2007; Morsella & Miozzo, 2002). The remaining four to six pictures in each array were randomly selected from the entire set of target and filler pictures.

The experiment included 144 trials, including 96 target trials and 48 filler trials. Each target trial included three or four events (48 trials of each length). Variable trial lengths were implemented to discourage participants from adopting a prosodic strategy of utterance-final lengthening for descriptions of the third (critical) event, which could counteract the effect of discourse-dependent reduction. In each event within a trial, the pictures underwent an action: *expand*, *rotate*, *shrink*, or *fade*. Within a trial, each event involved a different, randomly selected action. For discourse-given target trials, the first and third events occurred with the same target picture. The second event occurred with a non-target picture; this was inserted so that participants did not place contrastive focus on the action verb, which would provide another source for reduction of the preceding noun. For discourse-new trials, all events occurred with different pictures (the target and two or three of the assigned non-targets). The target picture in discourse-new trials always appeared in the third event, for comparison with the third event in discourse-given trials. In a discourse-given trial as illustrated in Figure 1, the target item (*tie*) would occur in the first event (e.g., *The tie rotates*). Next, a non-target item (*kite*) would occur in the second event (e.g., *The kite shrinks*). Finally, the target item would occur in the third event with a distinct action (e.g., *The tie fades*).

So that all comparisons were within item and within participant, each participant produced each target word twice over the course of the experiment (once in a discourse-given trial and once in a discourse-new trial). These conditions were blocked such that for any target word, the discourse-given and discourse-new trials appeared in different blocks (separated by breaks in the experiment). Items were also assigned to sub-blocks such that the two targets in a pair (e.g., *tie* and *tire*) occurred in separate sub-blocks. Therefore, blocks included trials of both conditions, although never for the same target item or target pair. Finally, each block contained 12 filler trials, which ranged in length from two to four events. All filler trials were discourse-new, and the pictures were composed into sets by choosing at random from the set of candidate filler pictures. Each filler picture appeared in only one filler trial.

Eight lists were created to counterbalance the sub-block assignment of each target picture (1A, 1B, 2A, or 2B), discourse condition order (discourse-given in block 1 vs. discourse-new in block1), and trial length (3 or 4). Trials of all lengths were evenly distributed throughout the experiment, such that each sub-block (of 36 trials) contained four length-2 filler trials, four length-3 filler trials, four length-4 filler trials, 12 length-3 target trials, and 12 length-4 target trials. For each event in a trial (both target and filler), an action was chosen at random with no repetition of actions within a trial. Three versions of each list were

generated with different within-block random orders, creating 24 experiment lists.² Each participant was randomly assigned to one of these lists, with no more than one participant from each group assigned to each list.

Procedure

Participants were seated in a sound-attenuated booth in front of a computer screen. They were told that they would see an array of objects and were instructed to describe the actions that occurred with those objects as soon as they recognized the action. They were familiarized with the desired names for these actions (i.e., *rotate*, *shrink*, *expand*, and *fade*), and were told that each trial would include two, three, or four events. Participants were instructed to describe the event using a complete sentence, such as *The dog rotates*, and were asked to use single word names for the pictures (e.g., *kettle* but not *tea kettle*, or *ball* but not *blue ball*). The experimenter demonstrated the structure of the experiment and the desired type of description with two trials. Then, the participants completed four practice trials under the supervision of the experimenter, with any mistakes corrected. Participants were permitted to complete the practice trials an additional time upon request.

Following the practice trials, participants began the experiment, which was presented using Max/MSP software (Puckette et al., 2011). Participants controlled the initiation of events by clicking a *Go* button on the screen. Upon clicking *Go*, a grid of pictures appeared. After a two second delay, the first event occurred. When the participant finished speaking and was ready for the next event, they clicked *Go* again. The second event occurred after a 500 ms delay. Clicking *Go* the third time either elicited a third event (after a 500 ms delay), or a new grid of pictures appeared and the sequence restarted. On some trials, a fourth event occurred after clicking *Go* the fourth time (again, after a 500 ms delay). There were three breaks during the experiment, each occurring after 36 trials. Participants determined the length of the breaks. Recordings were made with a boom-mounted Shure SM81 Condenser Handheld Microphone sampling at 44,100 Hz.

Measurement

Speech onset latencies and determiner vowel durations and target noun durations in critical productions were manually annotated and measured using Praat (Boersma & Weenink, 2016). Prior to measurement, accuracy of the target productions was assessed and it was verified that speakers produced the target correctly in both the first and third events of discourse-given trials. During acoustic annotation, the annotator was blind to the experimental condition of each trial.

Speech onset latencies (RTs) were calculated from the onset of the action animation (marked by the experimental software in a second acoustic channel synced to the speech) to the onset of the vowel of the determiner. Vowel onset for the determiner (rather than fricative onset) was chosen due to the difficulty in reliably identifying the onset of frication for /ð/. To

²An error in the counterbalancing of one list was discovered during testing. In length 4 trials, all fourth events used the same action (*rotate*), leading to some repetition of that action within a trial. However, because the fourth event of trials were never analyzed, the general structure of the list was still acceptable for the purposes of the experiment. Therefore, data from this participant was included. The error was rectified for the other two randomized versions of this list.

obtain determiner vowel durations, vowel onsets were marked at a rising zero crossing when clear formant structure had emerged for the vowel and vowel offsets were marked at a rising zero crossing on/after a sharp drop in amplitude upon closure for an upcoming consonant. To obtain target noun durations, word onsets and offsets were marked at zero crossings. Phoneme- and class-specific criteria for noun onset and offset boundary marking are listed in Appendix B. Verb productions were measured but not analyzed, as they were not balanced across trials and were not designed to match across conditions for each target.

The reliability of acoustic measurements was assessed by having an additional phonetically-trained annotator measure 10% of the target trials ($n = 192$). These trials were randomly sampled from the entire set of trials selected for the noun duration analysis, with an equal number of trials sampled from each group, for each target word, for each experimental condition, and (as much as possible) from each participant. All measurements from each annotator were highly correlated (RTs: Pearson's $r = 0.899$; nouns: Pearson's $r = 0.954$; determiner vowels: Pearson's $r = 0.906$). The mean absolute difference was 27.1 ms between RT measurements, 19.7 ms between noun duration measurements, and 5.5 ms between determiner vowel duration measurements.

Analysis

Duration measures—Three duration measures were considered for analysis: RT, determiner vowel duration, and target noun duration. The maximum number of observations available for analysis was 4,608 (48 participants produced 48 target words in 2 conditions). Trials were excluded from duration analysis if they were audibly disfluent (repetitions, false starts, elongations, pauses longer than 250 ms) at any point during the trial ($N = 505$, 11% of data) or if the speaker did not produce the correct target word ($N = 688$, 14.9% of data).

Separate linear mixed effects regressions were built for each of these duration measures using the lme4 package, version 1.1-7 (Bates, Maechler, Bolker, & Walker, 2015), in R 3.2.4 (R Core Team, 2016). Following previous work with this paradigm (Kahn & Arnold, 2015), baseline models with a series of control factors were first built. After fitting the initial baseline model, control factors were included in the final baseline model only if they contributed significantly to model fit (assessed by testing the significance of each factor via model comparison; see Appendix C for final baseline model structure for each measure). Candidate variables for control factors were RTs, determiner durations, noun durations, lexical frequency (all continuous factors were log-transformed and centered), and block (contrast-coded; block 1 vs. block 2). These models included the maximal random effects structure supported by the data, determined by building the maximal possible random effects structure (all possible random slopes for the by-item and by-participant intercepts) and simplifying until convergence was achieved. Baseline models included random intercepts for participants and items (target nouns) and random slopes for all significant control factors.

After fitting the baseline model, the fixed effects of interest were included. The dependent measure for each regressions was the log-transformed duration measure of interest. Fixed effects for these analyses included contrast-coded effects for group (L1 vs. L2) and discourse condition (discourse-given vs. discourse-new), and their interaction. These factors were also included in the random effects structure. The maximum random effects structure

allowed by the data was included in all models (Barr, Levy, Scheepers, & Tily, 2013). Significance of main effects and interactions was assessed via nested model comparison.

Disfluencies—The rate of disfluencies produced by participants was considered in a separate analysis using a series of logistic mixed effects regressions. Trials excluded from duration analyses for both disfluency and naming errors were included in this analysis. The dependent variable for these analyses was a binary measure (disfluent vs. fluent). Fixed effects for these models included contrast-coded effects for group (L1 vs. L2) and discourse condition (discourse-given vs. discourse-new), as well as their interaction. Models included the maximum random effects structure supported by the data (Barr et al., 2013), including decorrelated random slopes for group, discourse condition, and block by item and decorrelated random slopes for discourse condition and block by participant. Nested model comparison was used to perform significance tests.

Results

All de-identified data and analysis scripts can be accessed via the Open Science Foundation (<https://osf.io/jdwbp/>). The raw acoustic data (for the subset of participants that consented to release of the data) are available in the Online Speech/Corpora Archive and Analysis Resource (OSCAAR; <https://oscaar.ci.northwestern.edu/>; collection ‘Probabilistic Reduction Production’).

The results for response times and durations are summarized in Table 2.

Response Times

Observations lying more than 3 standard deviations from each participant's condition mean were excluded from analysis ($N = 25$, 0.5% of data). The model for response times (RTs) included decorrelated random slopes for block, determiner duration, noun duration, group, discourse condition, and the interaction between group and discourse condition by item. For by participant random effects, decorrelated random slopes for block, determiner duration, noun duration, and condition were included. Models were refit after removing observations with greater than 2.5 standardized residuals in the original model ($N = 76$, 1.6% of data).

Overall, L2 speakers were slower to initiate speech than L1 speakers (L1: mean = 1833.60 ms, SE = 82.30 ms; L2: mean = 2117.23 ms, SE = 81.83 ms; $\beta = -0.13$, SE = 0.06, $\chi^2(1) = 4.91$, $p < 0.05$). Similarly, response times were longer in discourse-new compared to discourse-given conditions (discourse-given: mean = 1918.08 ms, SE = 61.88 ms; discourse-new: mean = 2037.05 ms, SE = 69.07 ms; $\beta = -0.04$, SE = 0.01, $\chi^2(1) = 24.43$, $p < 0.001$), indicating words with high discourse-dependent probability were easier to plan and initiate. However, the difference in response times across discourse conditions was similar across groups ($\beta = 0.02$, SE = 0.01, $\chi^2(1) = 2.18$, $p > 0.05$). These results are shown in Figure 2.

Together, the results of this analysis replicate a series of previous findings. As in Kahn and Arnold (2015), L1 speakers exhibited significantly longer planning times (indexed by RTs) in discourse-new compared to discourse-given conditions. Furthermore, these results indicate delays in speech planning for L2 vs. L1 speakers. This bilingualism-related slowing

has been well-attested in the literature for bare picture naming tasks (e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005) and as well as picture naming in sentence generation tasks similar to the current task (Runnqvist, Gollan, Costa, & Ferreira, 2013). However, despite this overall processing slow-down, and a numerical decrease in discourse-dependent effects, L2 speakers exhibited no significant deficits in how probabilistic information influenced processing; similar facilitatory effects were found for L1 and L2 speakers in sentences containing a target noun with high discourse-dependent probability. This result stands in contrast to existing findings of differences between L1 and L2 speakers in the influence of other types of probabilistic information on processing (namely, influence of lexical frequency; Gollan et al., 2011).

Noun Durations

Observations lying more than 3 standard deviations from each participant's condition mean were excluded from analysis ($N = 7$, 0.2% of data). Furthermore, it is important for analysis of noun durations to be within participant and within item, as word durations vary greatly across individual speakers and words. Therefore, observations in the noun duration dataset were matched within participant such that an observation was only included if the target item was produced (following disfluency and outlier trimming) for both discourse conditions ($N = 598$ excluded; 13% of data). Observations were further matched across groups by randomly excluding L1 observations until each target item was represented equally across groups ($N = 574$ excluded; 12.5% of data). This matching across groups was important to ensure that any differences across groups could not be due to imbalances in the individual words included in the analysis (i.e., many observations of certain words in the L1 group vs. few for the L2 group). After these exclusions, a model for noun durations was built that included decorrelated random slopes for block, RT, group, discourse condition, and the interaction between group and discourse condition by item. For by-participant random effects, decorrelated random slopes for block, RT, frequency, and condition were included. Model-based outlier trimming was not performed for this analysis, as it would require re-matching the observations across items and groups.

Overall, L2 speakers produced longer word durations than L1 speakers (L1: mean = 335.27 ms, SE = 8.36 ms; L2: mean = 409.23 ms, SE = 18.59 ms; $\beta = -0.18$, SE = 0.05, $\chi^2(1) = 12.63$, $p < 0.001$). Furthermore, speakers produced significantly shorter word durations in discourse-given compared to discourse-new conditions (discourse-given: mean = 357.55 ms, SE = 11.64 ms; discourse-new: mean = 386.97 ms, SE = 11.46 ms; $\beta = -0.07$, SE = 0.01, $\chi^2(1) = 43.61$, $p < 0.001$). However, the magnitude of the effect of discourse condition on word durations did not differ across groups, indicated by a non-significant group by discourse condition interaction ($\beta = 0.008$, SE = 0.02, $\chi^2(1) = 0.28$, $p > 0.05$).

These results reveal that L1 and L2 speakers do not differ in the magnitude of discourse-dependent probabilistic reduction produced on target noun durations (see Figure 3). However, substantial differences in variability in word durations across groups raise the possibility that the lack of interaction could be driven by differences across groups in the set of words included in the analysis. To investigate this possibility, a Monte Carlo simulation was performed. The random sampling undertaken to match observations across groups was

repeated 1000 times. In other words, we repeated the procedure by which L1 speaker data was matched to the smaller set of L2 data, insuring equal representation of lexical items across datasets (see above). The model built for analyzing word durations was re-run on each of these 1000 samples (the models failed to converge an additional 42 times). The result of interest for this simulation was the distribution of p-values across samples; for each effect in the model, we considered the proportion of p-values that passed the threshold of significance (i.e., $p < 0.05$).

The results of this simulation revealed that the majority of effects held across the random samples. In particular, among the control factors, the effects of log frequency and RT replicated (frequency: 99.3% significant samples; RT: 100% significant samples). However, the effect of block on word durations was not reliable across samples (22.6% significant samples). Among the critical effects of interest, all results replicated; the effects of group and discourse condition on word durations were reliable (100% significant samples for each), while the group by discourse condition interaction never achieved significance across samples (0% significant samples). These results reveal that the lack of group by discourse condition interaction could not be attributed to differences in the words represented in the L1 vs. L2 dataset.

As with RTs, the noun duration results are consistent with existing findings in the literature. A number of other studies have also found that nouns are significantly reduced in duration when discourse-given compared to discourse-new (e.g., Kahn & Arnold, 2015). Existing studies of L2 speech production have also found that L2 speakers produce overall longer word durations than L1 speakers (Baker et al., 2011; Guion et al., 2000). The current results also extend previous findings that L2 speakers show comparable levels of reduction of nouns when discourse-given vs. discourse-new in reading tasks (Baker et al., 2011) by demonstrating this effects hold in a more challenging picture naming task, which requires semantic processing.

Determiner Vowel Durations

Observations lying more than 3 standard deviations from each participant's condition mean were excluded from analysis ($N = 11$, 0.2% of data). Furthermore, observations were excluded when the vowel of the determiner was devoiced, as the duration of the vowel was difficult to measure reliably ($N = 191$, 4.1% of data). In contrast to the noun analysis, there was no concern about imbalances in word representation across groups (as this analysis includes only one word). Therefore, random sampling to match observations across groups was not done for determiners. The model for determiner vowel durations included decorrelated random slopes for block, RT, noun duration, group, discourse condition, and the interaction between group and discourse condition by item. For by participant random effects, decorrelated random slopes for block, RT, noun duration, and condition were included. Models were refit after removing observations with greater than 2.5 standardized residuals in the original model ($N = 64$, 1.4% of data). Control factors and fixed effects of interest included in the model are summarized in Table 2.

The duration of determiner vowels did not differ significantly across groups (L1: mean = 43.26 ms, SE = 2.55 ms; L2: mean = 48.75 ms, SE = 2.89 ms; $\beta = -0.04$, SE = 0.08, $\chi^2(1) =$

0.30, $p > 0.05$). There was a main effect of discourse condition on determiner vowel durations (discourse-given: mean = 44.78 ms, SE = 1.99 ms; discourse-new: mean = 47.27 ms, SE = 1.95 ms; $\beta = -0.03$, SE = 0.01, $\chi^2(1) = 6.82$, $p < 0.01$), indicating that determiners modifying discourse-given nouns were significantly shorter than those modifying discourse-new nouns. However, a marginal interaction with group ($\beta = -0.05$, SE = 0.02, $\chi^2(1) = 3.80$, $p = 0.051$) suggested that reduction effects differed across groups. Follow-up regressions revealed a significant main effect of condition for L1 speakers (discourse-given: mean = 41.70, SE = 2.42; discourse-new: mean = 44.87, SE = 2.75; $\beta = -0.05$, SE = 0.02, $\chi^2(1) = 9.16$, $p < 0.01$) but not L2 speakers (discourse-given: mean = 47.87, SE = 3.07, discourse-new: mean = 49.68, SE = 2.74; $\beta = -0.003$, SE = 0.02, $\chi^2(1) = 0.03$, $p > 0.05$). These results suggest that L1, but not L2, speakers reduce the duration of determiners modifying discourse-given nouns compared to those modifying discourse-new nouns. L2 speakers' vowel durations in both conditions are roughly equivalent to L1 speakers' durations of determiners preceding discourse-new nouns (see Figure 4).

These results are consistent with previous findings. As in Kahn and Arnold (2015), L1 speakers produce significantly shorter determiners in sentences with a discourse-given vs. discourse-new target noun. Furthermore, while Baker et al. (2011) observed that L2 speakers produced significantly longer function words than L1 speakers, this result did not seem to hold for the most common function words, such as the definite determiner *the*. Interestingly, despite showing no general deficit in the production of determiners, the marginal interaction of group and discourse condition provides some support for a lack of reduction of determiners by L2 speakers.

Comparison of Effects on Nouns and Determiners: Multivariate Bayesian Analysis

To explore the potential determiner-specific deficit in the use of probabilistic information, we conducted a second analysis that allowed a more direct comparison of these two measures. We constructed a mixed-effects multivariate regression, simultaneously predicting *both* determiner and noun durations on each trial. We used the R package MCMCglmm (Hadfield, 2010) to perform a Bayesian analysis, sampling coefficients from the posterior probability distribution of the mixed-effects model conditioned on the data and the model's prior. These samples were then used to estimate the 95% central interval for each coefficient. This allowed us to assess which coefficients are likely to make a non-null contribution to the model (assessed by whether the central interval for a coefficient overlaps with zero) and which effects are different across the two dependent measures (assessed by whether the central interval for the difference in coefficients across noun vs. determiner durations overlaps with zero).

We utilized the subset of the data (matched across L1 and L2 speakers) constructed for the noun duration analyses above. In order to compare effect sizes across measures, log noun and determiner vowel duration were normalized (using standard scores) so that each measure was on the same scale. Following our univariate analyses, we first constructed a baseline model which predicted (normalized) log noun and determiner vowel durations from log reaction time, log frequency, and block (all centered), utilizing the maximal random effects structure. A weak prior³ assumed that effects were independent. The posterior

distribution was estimated by 10,000 samples⁴. Log reaction time (but neither block nor log frequency) made a non-null contribution to both noun and determiner durations (see Appendix D for control factor results).

Based on this, we built a model including log reaction time along with the fixed effects of interest (group, condition, and their interaction). The maximal random effects structure was included and posterior estimation followed the procedures outlined above. As shown in Table 3, the results fail to support different effects for nouns vs. determiner vowel durations. Similar to the univariate analyses, this analysis suggests that group and condition (but not their interaction) contribute to variation in noun durations: L1 speakers have shorter noun durations than L2 speakers, and discourse-given nouns are shorter than discourse-new nouns. In contrast to the previous analysis, none of these factors contributed to variation in determiner vowel durations. Critically, for each factor, the central intervals for effects on determiner vowel vs. noun durations exhibit a very high degree of overlap. The central intervals for the difference in determiner vowel vs. noun coefficients for the main effects included zero (group effect central interval: $[-0.87, 0.17]$; discourse condition interval: $[-0.43, 0.16]$) This suggests there is no strong evidence that these factors differentially modulate determiner vs. noun durations.

To confirm that these results are not specific to the particular random subset of matched nouns, a second Monte Carlo analysis was performed. Similar to the analysis above, the multivariate analysis was repeated over 1,000 different random subsets of the data. Qualitatively similar results were found across each of these analyses.

Taken together, the univariate and multivariate analyses replicate previous studies showing that L2 speakers produce longer word durations than L1 speakers and discourse-given nouns are shorter than discourse-new nouns. Critically, both analysis methods suggest that the effect of discourse status is equivalent across two groups. While the results are less clear with respect to determiners, there is no strong evidence that the two groups show different effects of discourse status; the interaction of group and discourse status is marginal in the univariate and not significant in the multivariate analysis. This suggests that L2 speakers exhibit similar sensitivity to discourse status in their productions.

Disfluencies

The results of the logistic mixed effects regression revealed that there was a main effect of group on disfluencies, where L2 speakers produced a significantly higher proportion of disfluencies compared to L1 speakers (L1: mean = 37.07, SE = 2.02; L2: mean = 14.71, SE = 1.49; $\beta = 1.53$, SE = 0.22, $\chi^2(1) = 38.64$, $p < 0.001$). Speakers also produced a significantly higher proportion of disfluencies in discourse-new vs. discourse-given conditions, indicated by a main effect of discourse-condition (discourse-given: mean = 23.52, SE = 1.93; discourse-new: mean = 28.26, SE = 2.30; $\beta = 0.34$, SE = 0.08, $\chi^2(1) = 15.25$, $p < 0.001$). As the experiment progressed, speakers produced lower proportions of

³MCMCglmm's default fixed effects prior was utilized. The priors for the random effect covariance matrices and residual covariance matrix were each given by an inverse-Wishart distribution, with an identity scale matrix and the lowest possible degrees of freedom.

⁴These were taken from 10 independent chains of 10,000 iterations, thinned so that every 10th sample was used, with 3,000 iterations as burn-in. A multivariate potential scale reduction factor (Brooks & Gelman, 1997) of 1 confirmed mixing of these chains.

disfluencies, shown by a main effect of block (block 1: mean = 28.17, SE = 2.34; block 2: mean = 23.61, SE = 1.88; $\beta = -0.29$, SE = 0.08, $\chi^2(1) = 11.59$, $p < 0.001$). None of the two-way interactions were significant (all $\chi^2(1) < 2$, $ps > 0.05$). There was a trend towards a three-way interaction ($\beta = -0.52$, SE = 0.32, $\chi^2(1) = 2.56$, $p < 0.11$), driven by a block by discourse condition interaction for L2 vs. L1 speakers (for the former group, the effect of discourse condition was stronger in the first block).

Analysis of disfluencies largely parallels the results in fluent speech. In particular, L2 speakers were both more disfluent, were slower to initiate speech, and produced longer fluent word durations on average. Similarly, speech was more likely to be disfluent, be more difficult to initiate, and have longer durations on average in discourse-new compared to discourse-given conditions. Finally, parallel to the majority of duration analyses, L2 speakers did not produce significantly more disfluencies in discourse-new conditions compared to L1 speakers. These results suggest that the processing of probabilistic information proceeds largely similarly across groups, both during successful speech processing and when it breaks down, regardless of differences in linguistic experience.

General Discussion

The current study investigated differences in how probabilistic information – specifically, discourse-dependent probability – affects planning and articulation during L1 and L2 speech production in a demanding referential communication task. For RTs and production of content and function words, we asked whether L2 speakers show deficits in how probabilistic information influences processing; that is, do they produce different levels of probabilistic reduction when compared to L1 productions? The results showed no such deficits. Both L1 and L2 speakers produced significantly shorter RTs and word durations in discourse-given vs. discourse-new conditions, and, critically, there was no significant difference in the size of the reduction effect across groups.

L2 Processing of Probabilistic Information

Consistent with previous research, L2 speakers were significantly slower to initiate speech (e.g., Gollan et al., 2008) and produced significantly longer word durations in comparison to L1 speakers (Baker et al., 2011; Guion et al., 2000). These results were also consistent with previous findings with read speech, where L2 speakers produced discourse-dependent probabilistic reduction of the same degree as L1 speakers (Baker et al., 2011).

While the duration findings are clearly consistent with previous work, the RT results – specifically, finding no differences in the size of discourse probability effects across L1 and L2 speakers – differ for some previous findings. Several studies have demonstrated deficits for probabilities conditioned on a word's distribution within a language (e.g., lexical frequency; Gollan et al., 2008; Gollan et al., 2011; although, see Sadat et al., 2012 for conflicting results). We argue that these differences reflect the different types of linguistic experience that underlie each of these effects. Lexical frequency effects have been claimed to arise from speakers' usage of particular word forms. For example, the frequency lag hypothesis (Gollan et al., 2008) claims that L2 speakers (by virtue of speaking more than one language) have less experience with word forms in each language than a comparable

monolingual speaker. In contrast, discourse-dependent effects are not specific to particular lexical items, but emerge from experience producing and perceiving referring expressions (a discourse role that can be filled by many different lexical items). Our results indicate that L2 speech processing is not so demanding as to prohibit L2 speakers from tracking the discourse-dependent probability of a word. This allows them to gain experience in both languages, no differences in the magnitude of discourse-dependent probabilistic reduction on RTs and word durations.

Probabilistic Information and Determiner Production

Our analyses did not reveal clear patterns of performance for function word production. The univariate analysis provided some support for reduction of determiners preceding nouns (following Kahn & Arnold, 2015). While the multivariate analysis did not find such effects, it also failed to find significant differences between effects of discourse status on nouns vs. determiners. The univariate analysis also provided weak evidence in favor of differences in reduction across L1 and L2 speakers; this too was not replicated in our multivariate analysis. Clearly, more evidence is needed before we can draw strong conclusions about determiner processing.

Probabilistic Information for the Speaker vs. for the Listener

Researchers have proposed a variety of mechanisms to account for how probabilistic information influences planning and articulation, including both speaker-based and listener-based processes. According to speaker-based theories, the probabilities associated with a word directly influence the mechanisms underlying speech production, either in terms of how probabilities are stored, how they influence lexical access, or both. Storage-based accounts argue that the probabilities associated with a word are encoded in (long-term) linguistic representations (via resting activation (Dell, 1990); via phonetically-specified representations (Pierrehumbert, 2001, 2002; Seyfarth 2014)). These representations, and the probabilities they encode, influence processing, and lead to reduction during planning and articulation. For example, retrieval of high probability lexical items may be easier than for low probability items due to high resting activation in long-term memory representations. Words that are easy to retrieve will be selected more quickly and will be easy to articulate, leading to hypoarticulation. Ease of retrieval within a discourse has also been attributed to temporary boosts in resting activation due to priming from recent retrieval (Arnold & Watson, 2015; Fraundorf, Watson, & Benjamin, 2014; Kahn & Arnold, 2012). While these storage-based and processing-based accounts are difficult to dissociate, they both reflect purely speaker-based mechanisms.

In contrast, listener-based theories argue that the probabilities associated with a word reflect some calculation undertaken by the speaker for the benefit of the listener. The main assumption underlying such theories is that there is some communicative intent of the articulatory reduction resulting from the influence of probabilistic information on processing. Lindblom's (1990) Hyperspeech-Hypospeech theory proposed that speakers hyperarticulate words that they think will be difficult for their listener to understand (e.g., low probability words), and reduce words that they think will be easy (e.g., high probability words), thus balancing the need for effective communication and the desire to minimize

articulatory effort. Information theoretic accounts, such as the Smooth Signal Redundancy Hypothesis (Aylett & Turk, 2004) and the Uniform Information Density Hypothesis (Levy & Jaeger, 2007), take a similar approach. These theories contend that effective communication requires a trade-off between a desire to produce brief (i.e., reduced) linguistic signals and the need for communication error to be low (Aylett & Turk, 2004; Levy & Jaeger, 2007; Pate & Goldwater, 2015). However, highly reduced signals could impede successful communication if “noise” enters the system (Shannon, 1948) on the speaker side (e.g., speech errors), in the environment (e.g., a noisy room), or on the listener side (e.g., distraction). Therefore, brief signals are optimal only when the message has high probability, because such messages will have intrinsically low probability of communication error. By contrast, communication of messages with low probability should be encoded with relatively longer signals to avoid errors in communication.

Either class of theory can account for the results of the current study. Following speaker-based theories, high discourse-dependent probability could cause a temporary boost in the resting activation for the representation, which facilitates lexical retrieval and leads to reduction. Under listener-based theories, speakers know they can reduce word with high discourse-given probability without risking perception error on the part of their (potential) listener. It is likely that both speaker-driven and listener-driven forces underlie these effects (e.g., Arnold, Kahn, & Pancani, 2012; Rosa, Finch, Bergeson, & Arnold, 2015). Our results with L2 speakers indicate that whichever of these mechanisms drives the influence of discourse-dependent probability on reduction for L1 speakers also does so for L2 speakers.

An underlying assumption to the listener-based theories discussed above is that listeners are (implicitly) aware of the relationship between probability and reduction (Mitterer & Russell, 2013); listeners should know that reduced words are likely to have high probability, and reduced words have low probability. That is, when listeners hear a reduced syllable, they predict that the syllable belongs to word with high vs. low probability according to some factor, such as discourse status (Arnold, 2008; Dahan, Tanenhaus, & Chambers, 2002). Listeners also use acoustic information earlier in the sentence (e.g., at the determiner) to make predictions about upcoming words (e.g., Arnold, Tanenhaus, Altmann, & Fagnano, 2004). While L1 listeners undoubtedly use discourse-dependent reduction to make predictions during speech production, it is unclear if L2 listeners can. Under certain circumstances, L2 listeners cannot make predictions at all (e.g., Lew-Williams & Fernald, 2010). However, if L2 processing of probabilistic information leads to L1-like reduction, as in the current study, it is possible L2 listeners will also be able to use this reduction as a predictive cue during speech perception (Pickering & Garrod, 2007). Ongoing work in our lab is investigating this issue.

Conclusions

In conclusion, despite overall differences in processing across groups, discourse-dependent probabilistic information has comparable influence on L1 vs. L2 speech planning and execution, unlike lexical frequency. These results provide important evidence for understanding how different types of probabilistic information – and the degree to which

they depend on general experience with a language – influence L2 speech planning and execution.

Acknowledgments

This research was supported by the National Institutes of Health under Grant 1R21HD077140.

Appendix A: Stimuli

Table A1

Pairs of target items.

Target item 1	L1 name agreement	L2 name agreement	SUBTLEX frequency	Target item 2	L1 name agreement	L2 name agreement	SUBTLEX frequency
arch	78.9%	70%	3.7	arm *	94.7%	100%	65.4
baby	100%	80%	509.4	bacon	100%	80%	11.9
bag	100%	70%	94.0	bat *	100%	60%	20.6
beach	100%	90%	56.6	bee *	94.7%	100%	10.4
bed *	100%	100%	187.1	bench *	78.9%	60%	9.7
brick *	94.7%	70%	10.2	bridge *	100%	100%	45.7
butter	94.7%	60%	20.4	button *	100%	40%	28.3
cage	63.2%	60%	20.3	cake	100%	80%	45.1
camel *	100%	90%	5.0	camera *	100%	100%	57.0
candy *	89.5%	100%	35.8	candle *	100%	100%	8.0
chair *	100%	100%	49.2	chain *	100%	70%	21.2
dollar	89.5%	50%	27.7	dolphin *	94.7%	100%	2.8
fish	84.2%	100%	83.5	fist	100%	40%	7.4
ghost	100%	80%	36.6	goat	100%	70%	10.5
mouse *	78.9%	70%	19.1	mouth	100%	90%	104.4
net	100%	60%	15.5	nest *	100%	100%	59.5
peanuts	100%	100%	12.4	pizza	100%	100%	33.5
pickle *	100%	60%	4.6	picture	78.9%	60%	138.5
pig *	94.7%	100%	39.1	pill *	89.5%	50%	11.8
sandwich	100%	90%	21.9	sandal *	89.5%	40%	0.2
scar	63.2%	60%	8.5	star	100%	90%	81.4
tire	100%	80%	12.4	tie *	94.7%	90%	44.4
turtle	100%	90%	17.0	turkey	100%	40%	22.6
witch	100%	50%	27.6	wing	94.7%	100%	20.2

* denotes items from BOSS. SUBTLEX frequencies are occurrences per million words.

Table A2

Non-target items with associated target pairs.

Target pair	Non-target item 1	Non-target item 2	Non-target item 3
arch/arm	pepper *	bottle	clock *
baby/bacon	leaf *	heart	shirt *
bag/bat	table *	onion *	daisy *

Target pair	Non-target item 1	Non-target item 2	Non-target item 3
beach/bee	eye *	watch *	eggs *
bed/bench	headphones *	sock *	wine
brick/bridge	hand *	airplane *	phone
butter/button	monkey	hand *	dog
cage/cake	lemon *	moon *	belt *
camel/camera	lock *	apple	bird *
candy/candle	ear *	bell	toothbrush *
chair/chain	box	window	coins *
dollar/dolphin	grapes	nose *	fork *
fish/fist	key *	road *	guitar *
ghost/goat	cat	sun	bra *
mouse/mouth	basket *	car	spoon *
net/neck	tree *	snake *	cookie *
peanuts/pizza	ruler *	glasses	comb *
pickle/picture	horse *	shoe *	ring
pig/pill	orange *	cart	skirt
sandwich/sandal	book *	coat	chicken *
scar/star	door	pen *	tank *
tire/tie	kite *	knife *	spider
turtle/turkey	hat *	bowl	scarf *
witch/wing	brain	pencil *	toilet *

* denotes items from BOSS.

Appendix B: Measurement Criteria

Table A3

Measurement criteria for marking noun onset boundaries according to initial phoneme/class of the noun.

Phoneme/Class	Criteria
Vowels	F2 lowering
	Pause between determiner and noun
[w]	Amplitude drop off in spectrogram
	Before lowering of F2 and raising of F3
Stops/affricates	Before release burst
Fricatives	Onset of high energy noise in spectrogram
	Noise evident in waveform
Nasals	Overall drop in amplitude in spectrogram
	Shape change in waveform

Table A4

Measurement criteria for marking noun offset boundaries for [f]-initial verb according to final phoneme/class of the noun.

Phoneme/Class	Criteria
Sonorants	Onset of high frequency energy in spectrogram and/or waveform
Stops	End of aspiration Lower overall amplitude in energy bands at lowest frequencies
[θ]	Drop in overall amplitude

Table A5

Measurement criteria for marking noun offset boundaries for [ɛ]-initial verb according to final phoneme/class of the noun.

Phoneme/Class	Criteria
[ɹ]	After raising of F2 and F3 Onset of stable F3
[n]	Sudden amplitude increase Increased clarity of F2 formant structure
[l]	F2 rising
[aɪ]	Stability of formants after diphthong movement
[ɪ]	Before lowering of F2 and F3
Fricatives/affricates	Decreasing amplitude of noise in spectrogram Onset of periodicity
Stops	End of aspiration Onset of periodicity

Table A6

Measurement criteria for marking noun offset boundaries for [ʃ]-initial verb according to final phoneme/class of the noun.

Phoneme/Class	Criteria
Sonorants	Onset high frequency energy in spectrogram and waveform
Stops	End of aspiration Onset of periodicity

Table A7

Measurement criteria for marking noun offset boundaries for [ɹ]-initial verb according to final phoneme/class of the noun.

Phoneme/Class	Criteria
[n]	After lowering of F3

Phoneme/Class	Criteria
Other sonorants	Onset of decrease in overall amplitude
	After lowering of F3 to F2

Appendix C: Control Factor Results, Univariate Analysis

Table A8

Results for control factors in model of response times.

	Beta	SE	Chi-squared	<i>p</i>
Block	0.10	0.01	32.20	< 0.001
Log determiner duration	0.07	0.01	23.26	< 0.001
Log noun duration	0.07	0.01	21.23	< 0.001

Table A9

Results for control factors in model of noun durations.

	Beta	SE	Chi-squared	<i>p</i>
Block	-0.02	0.01	2.08	> 0.05
Log frequency	-0.05	0.02	5.35	< 0.05
Log RT	0.13	0.03	15.28	< 0.001

Table A10

Results for control factors in model of determiner durations.

	Beta	SE	Chi-squared	<i>p</i>
Block	0.03	0.02	3.00	< 0.09
Log RT	0.23	0.04	25.93	< 0.001
Log noun duration	0.32	0.04	47.39	< 0.001

Appendix D: Control Factor Results, Multivariate Analysis

Table A11

Central intervals from control factors in baseline multivariate analysis.

		Central Interval
Log RT	Determiner vowel	0.31, 0.99
	Noun	0.21, 0.81
Block	Determiner vowel	-0.12, 0.32
	Noun	-0.25, 0.15
Log Word Frequency	Determiner vowel	-0.19, 0.18
	Noun	-0.35, 0.04

Table A12

Central intervals from control factors in multivariate analysis.

Central Interval		
Log RT	Determiner vowel	0.38, 1.07
	Noun	0.08, 0.68

References

- Altmann EG, Pierrehumbert JB, Motter AE. Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLoS ONE*. 2009; 4(11)doi: 10.1371/journal.pone.0007678
- Aoyama, K., Guion, SG. Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In: Bohn, OS., Munro, M., editors. *Language experience in second language speech learning*. Amsterdam: John Benjamins Publishing Co; 2007. p. 281-297.
- Arnold JE. THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. *Cognition*. 2008; 108:69–99. DOI: 10.1016/j.cognition.2008.01.001 [PubMed: 18358460]
- Arnold JE, Kahn JM, Pancani GC. Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*. 2012; 19(3):505–12. DOI: 10.3758/s13423-012-0233-y [PubMed: 22419403]
- Arnold JE, Tanenhaus MK, Altmann RJ, Fagnano M. The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science*. 2004; 15(9):578–582. DOI: 10.1111/j.0956-7976.2004.00723.x [PubMed: 15327627]
- Arnold JE, Watson DG. Synthesising meaning and processing approaches to prosody: performance matters. *Language and Cognitive Processes*. 2015; 30:88–102. DOI: 10.1080/01690965.2013.840733
- Arnon I, Snider N. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*. 2010; 62:67–82.
- Aylett M, Turk A. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*. 2004; 47(1):31–56. DOI: 10.1177/00238309040470010201 [PubMed: 15298329]
- 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. DOI: 10.1177/0023830909336575 [PubMed: 20121039]
- Baker RE, Baese-Berk M, Bonnasse-Gahot L, Kim M, Van Engen KJ, Bradlow AR. Word durations in non-native English. *Journal of Phonetics*. 2011; 39(1):1–17. DOI: 10.1016/j.wocn.2010.10.006 [PubMed: 21516172]
- 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.
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68(3):255–278. DOI: 10.1016/j.jml.2012.11.001
- Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. 2015; 67(1):1–48. DOI: 10.18637/jss.v067.i01
- 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(1):92–111. DOI: 10.1016/j.jml.2008.06.003
- 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. *The Journal*

- of the Acoustical Society of America. 2003; 113(2):1001–1024. DOI: 10.1121/1.1534836 [PubMed: 12597194]
- Bock, K., Levelt, WJM. Language production: Grammatical encoding. In: Gernsbacher, MA., editor. Handbook of psycholinguistics. San Diego: CA: Academic Press; 1994. p. 954-984.
- Boersma, P., Weenink, D. Praat: doing phonetics by computer [Computer program]. 2016. Version 6.0.16, retrieved 6 April 2016 from <http://www.praat.org/>
- Brodeur MB, Dionne-Dostie E, Montreuil T, Lepage M. The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. PLoS ONE. 2010; 5(5):1–13.
- Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1997; 7:434–455.
- Colomé À. Lexical activation in bilinguals' speech production: Language-specific or language-independent? Journal of Memory and Language. 2001; 45(4):721–736. DOI: 10.1006/jmla.2001.2793
- Colomé A, Miozzo M. Which words are activated during bilingual word production? Journal of Experimental Psychology Learning, Memory, and Cognition. 2010; 36(1):96–109. DOI: 10.1037/a0017677
- Costa A, Caramazza A, Sebastian-Galles N. The cognate facilitation effect: implications for models of lexical access. Journal of Experimental Psychology Learning, Memory, and Cognition. 2000; 26(5):1283–96.
- Dahan D, Tanenhaus MK, Chambers CG. Accent and reference resolution in spoken-language comprehension. Journal of Memory and Language. 2002; 47:292–314. DOI: 10.1016/S0749-596X(02)00001-3
- Dell GS. Effects of frequency and vocabulary type on phonological speech errors. Language and Cognitive Processes. 1990; 5(4):313–349.
- Diependaele K, Lemhöfer K, Brysbaert M. The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. The Quarterly Journal of Experimental Psychology. 2013; 66(5):843–863. DOI: 10.1080/17470218.2012.720994 [PubMed: 23025801]
- Ferreira F, Swets B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. Journal of Memory and Language. 2002; 46(1):57–84.
- Fowler CA. Differential shortening of repeated content words produced in various communicative contexts. Language and Speech. 1988; 31(4):307–319. DOI: 10.1177/002383098803100401 [PubMed: 3271893]
- 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. DOI: 10.1016/0749-596X(87)90136-7
- Fraundorf SH, Watson DG, Benjamin AS. Reduction in prosodic prominence predicts speakers' recall: implications for theories of prosody. Language, Cognition and Neuroscience. 2014; 30:606–619. DOI: 10.1080/23273798.2014.966122
- 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(4):748–775.
- Garrett, MF. The analysis of sentence production. In: Bower, GH., editor. The psychology of learning and motivation. San Diego: Academic Press; 1975. p. 133-175.
- Gollan TH, Montoya RI, Fennema-Notestine C, Morris SK. Bilingualism affects picture naming but not picture classification. Memory & Cognition. 2005; 33(7):1220–1234. DOI: 10.3758/BF03193224 [PubMed: 16532855]
- Gollan TH, Montoya RI, Cera C, Sandoval TC. More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. Journal of Memory and Language. 2008; 58(3):787–814. DOI: 10.1016/j.jml.2007.07.001 [PubMed: 19343088]

- Gollan TH, Slattery TJ, Goldenberg D, Van Assche E, Duyck W, Rayner K. Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*. 2011; 140(2):186–209. DOI: 10.1037/a0022256 [PubMed: 21219080]
- Guion SG, Flege JE, Liu SH, Yeni-Komshian GH. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*. 2000; 21(2):205–228. DOI: 10.1017/S0142716400002034
- Gustafson E, Engstler C, Goldrick M. Phonetic processing of non-native speech in semantic vs non-semantic tasks. *The Journal of the Acoustical Society of America*. 2013; 134(6):EL506.doi: 10.1121/1.4826914 [PubMed: 25669296]
- Hadfield J. MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *Journal of Statistical Software*. 2010; 33(2):1–22. [PubMed: 20808728]
- Hermans D, Bongaerts T, De Bot K, Schreuder R. Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*. 1998; 1(3):213–229. DOI: 10.1017/S1366728998000364
- Hernández M, Costa A, Arnon I. More than words: multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*. 2016; 3798(April):1–16. DOI: 10.1080/23273798.2016.1152389
- Jacobs CL, Yiu LK, Watson DG, Dell GS. Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language*. 2015; 84:37–48. DOI: 10.1016/j.jml.2015.05.004 [PubMed: 26089592]
- 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*. Amsterdam: Benjamin; 2001. p. 229–254.
- Kahn JM, Arnold JE. A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*. 2012; 67(3):311–325. DOI: 10.1016/j.jml.2012.07.002
- Kahn JM, Arnold JE. Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *Language, Cognition and Neuroscience*. 2015; 30(1-2):103–119. DOI: 10.1080/01690965.2013.848989
- Lam TQ, Marian V. Repetition reduction during word and concept overlap in bilinguals. *Journal of Memory and Language*. 2015; 84:88–107. DOI: 10.1016/j.jml.2015.05.005 [PubMed: 26166943]
- Lam TQ, Watson DG. Repetition is easy: why repeated referents have reduced prominence. *Memory & Cognition*. 2010; 38(8):1137–1146. DOI: 10.3758/MC.38.8.1137 [PubMed: 21156876]
- Lam TQ, Watson DG. Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2014; 40(3):829–843.
- Lapointe, SG., Dell, GS. A synthesis of some recent work in sentence production. In: Carlson, GN., Tanenhaus, MK., editors. *Linguistic structure in language processing*. Springer Netherlands: 1989. p. 107–156.
- Lemhöfer K, Broersma M. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*. 2012; 44(2):325–343. DOI: 10.3758/s13428-011-0146-0 [PubMed: 21898159]
- Levelt WJ, Roelofs A, Meyer aS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 1999; 22(1):1–75. [PubMed: 11301520]
- Levy, R., Jaeger, TF. Speakers optimize information density through syntactic reduction. In: Schlökopf, B. Platt, J., Hoffman, T., editors. *Advances in neural information processing systems (NIPS)*. Vol. 19. Cambridge, MA: MIT Press; 2007. p. 849–856.
- Lew-Williams C, Fernald A. Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*. 2010; 63(4):447–464. DOI: 10.1016/j.jml.2010.07.003 [PubMed: 21076648]
- Lieberman P. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*. 1963; 6(3):172–187.
- Lindblom, B. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, WJ., Marchal, A., editors. *Speech production and speech, modelling*. Dordrecht: Kluwer; 1990. p. 403–439.

- Meagher BR, Fowler CA. Embedded articulation: shifts in location influence speech production. *Language, Cognition and Neuroscience*. 2014; 29(5):1–7.
- Meyer AS, Damian MF. Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*. 2007; 35(3):494–503. [PubMed: 17691148]
- Mitterer H, Russell K. How phonological reductions sometimes help the listener. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2013; 39(3):977–84. DOI: 10.1037/a0029196
- Morsella E, Miozzo M. Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28(3):555–563. DOI: 10.1037//0278-7393.28.3.555
- Munro MJ, Derwing TM. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*. 1995; 45(1):73–97.
- Munson B, Solomon NP. The Effect of Phonological Neighborhood Density on Vowel Articulation. *Journal of Speech, Language, and Hearing Research*. 2004; 47:1048–1058.
- Pate JK, Goldwater S. Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*. 2015; 78:1–17. DOI: 10.1016/j.jml.2014.10.003
- Pickering MJ, Garrod S. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*. 2007; 11(3):105–110. DOI: 10.1016/j.tics.2006.12.002 [PubMed: 17254833]
- Pierrehumbert, JB. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, JL., Hopper, PJ., editors. *Frequency effects and emergent grammar*. Amsterdam: John Benjamins Publishing; Co: 2001. p. 137-157.
- Pierrehumbert, JB. Word-specific phonetics. In: Gussenhoven, C., Warner, N., editors. *Laboratory Phonology*. Vol. 7. Berlin/New York: Mouton; 2002. p. 101-139.
- Pluymaekers M, Ernestus M, Baayen RH. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*. 2005; 62(2-4):149–159.
- Puckette, M., Zicarelli, D., Sussman, R., Clayton, JK., Bernstein, J., Nevile, B., Place, T., Grosse, D., Dudas, R., Jourdan, E., Lee, M., Schabtach, A. MAX/MSP [Computer program]. 2011. Version 5.1.9 retrieved from <https://cycling74.com/downloads/older/#.V7UrUWUqafQ>
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer program]; 2016. Version 3.2.4 retrieved from <https://www.R-project.org/>
- Rosa EC, Finch KH, Bergeson M, Arnold JE. The effects of addressee attention on prosodic prominence. *Language, Cognition and Neuroscience*. 2015; 30(1-2):48–56. DOI: 10.1080/01690965.2013.772213
- Runnqvist E, Gollan TH, Costa A, Ferreira VS. A disadvantage in bilingual sentence production modulated by syntactic frequency and similarity across languages. *Cognition*. 2013; 129(2):256–63. DOI: 10.1016/j.cognition.2013.07.008 [PubMed: 23948209]
- Sadat J, Martin CD, Alario FX, Costa A. Characterizing the bilingual disadvantage in noun phrase production. *Journal of Psycholinguistic Research*. 2012; 41(3):159–79. DOI: 10.1007/s10936-011-9183-1 [PubMed: 21997516]
- Scarborough, RA. Lexical and contextual predictability: Confluent effects on the production of vowels. In: Fougerson, C. Kuhnert, M., D'Imperio, M., editors. *Laboratory Phonology*. Vol. 10. Berlin: De Gruyter; 2010. p. 557-586.
- Schertz J, Ernestus M. Variability in the pronunciation of non-native English the: Effects of frequency and disfluencies. *Corpus Linguistics and Linguistic Theory*. 2014; 10(2):329–345. DOI: 10.1515/cllt-2014-0024
- 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*. Mahwah, NJ: Lawrence Erlbaum; 1996. p. 187-213.
- Seyfarth S. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*. 2014; 133(1):140–155. DOI: 10.1016/j.cognition.2014.06.013 [PubMed: 25019178]
- Shannon CE. A mathematical theory of communication. *The Bell System Technical Journal*. 1948; 27(3):379–423.

- 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(1):169–201. DOI: 10.1017/S0305000997003395 [PubMed: 9604573]
- Steedman M. Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry*. 2000; 31(4):649–689.
- Zipf, GK. Human behavior and the principle of least effort. Cambridge, MA: Addison-Wesley; 1949.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Example array from experiment interface.⁵

⁵Images are from the Bank of Standardized Stimuli (BOSS; Brodeur et al., 2010) and are authorized for redistribution according to the Creative Commons Attribution-Share Alike 3.0 license.

Figure 2. Mean response time (ms) across groups and discourse conditions. Error bars show standard error.

Figure 3. Mean duration of nouns (ms) across groups and discourse conditions. Error bars show standard error.

Figure 4. Mean duration of determiner vowels (ms) across groups and discourse conditions. Error bars show standard error.

Table 1

Language background information. Mean (standard deviation).

	Percent Correct, LexTale vocabulary test	Age of first exposure to English [years]	Length studied English [years]	Percent time English used	Self-rated speaking ability [Perfect:10]	Self-rated listening ability [Perfect:10]
L1 group (N = 24)	95.7 (5.2)	0 (0)	18.9 (1.2)	95.7 (5.8)	9.7 (0.6)	9.8 (0.5)
L2 group (N = 24)	80 (13.4)	7.9 (5.8)	15.6 (6.8)	48.6 (27.4)	5.8 (2.4)	6.6 (2.3)

Table 2

Summary of regression results (direction of effects) for temporal measures in the main analyses. Direction of effects only shown for significant and marginal effects. Italicized text indicates a marginal effect.

	Response time (RT)	Determiner vowel	Target noun
Group (L2 vs. L1)	Negative	n.s.	Negative
Discourse condition (new vs. given)	Negative	Negative	Negative
Group $\times$ discourse condition	n.s.	<i>Negative</i> L1: negative L2: n.s.	n.s.

Table 3

Central intervals from Bayesian mixed-effects multivariate regression analysis of normalized determiner vowel and noun durations.

		Central Interval
Group (L2 vs. L1)	Determiner vowel	−0.69, 0.18
	Noun	−1.02, −0.21
Discourse condition (new vs. given)	Determiner vowel	−0.33, 0.11
	Noun	−0.45, −0.04
Group × discourse condition	Determiner vowel	−0.42, 0.27
	Noun	−0.30, 0.35