The influence of lexical selection disruptions on articulation

Matthew Goldrick[1], Rhonda McClain[2], Emily Cibelli[1],

Yossi Adi[3] , Erin Gustafson[1], Cornelia Moers[4], and Joseph Keshet[3,]

Northwestern University[1]

The Pennsylvania State University[2]

Bar Ilan University[3]

Max Planck Institute for Psycholinguistics[4]

Author Note

Matthew Goldrick, Department of Linguistics, Northwestern University

This work was supported by National Institute of Health (NIH NICHD) Grant No. 1R21HD077140 to M.G. and by a scholarship for CM from the Max Planck International Research Network on Aging (MaxNetAging). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NIH or MaxNetAging.

Correspondence concerning this article should be addressed to Matthew Goldrick, Department of Linguistics, Northwestern University, 2016 Sheridan Rd., Evanston, IL 60208. Email: matt-goldrick@northwestern.edu.

Abstract

Interactive models of language production predict that it should be possible to observe long-distance interactions; effects that arise at one level of processing influence multiple subsequent stages of representation and processing. We examine the hypothesis that disruptions arising in non-form based levels of planning – specifically, lexical selection – should modulate articulatory processing. A novel automatic phonetic analysis method was used to examine productions in a paradigm yielding both general disruptions to formulation processes and, more specifically, overt errors during lexical selection. This analysis method allowed us to examine articulatory disruptions at multiple levels of analysis, from whole words to individual segments. Baseline performance by young adults was contrasted with young speakers' performance under time pressure (which previous work has argued increases interaction between planning and articulation) and performance by older adults (who may have difficulties inhibiting non-target representations, leading to heightened interactive effects). The results revealed the presence of interactive effects. Our new analysis techniques revealed these effects were strongest in initial portions of responses, suggesting that speech is initiated as soon as the first segment has been planned. Interactive effects did not increase under response pressure, suggesting interaction between planning and articulation is relatively fixed. Unexpectedly, lexical selection disruptions appeared to yield some degree of facilitation in articulatory processing (possibly reflecting semantic facilitation of target retrieval) and older adults showed weaker, not stronger interactive effects (possibly reflecting weakened connections between lexical and form-level representations).

Keywords: speech production; interaction; articulation; automatic acoustic analysis

The influence of lexical selection disruptions on articulation

To produce a single word, a speaker must map an intended message to a lexical representation and select detailed representations regarding the word's sound structure (e.g., Garrett, 1975; Schriefers, Meyer, & Levelt, 1990; van Tourennout, Hagoort, & Brown, 1997; see Levelt, Roelofs, & Meyer, 1999, for an overview). This typically assumed to rely on several distinct processing stages collectively referred to as *formulation*. Speech production begins with the selection of a concept to verbalize the message, and then the speaker activates the concept's relevant semantic features. During lexical selection, these meaning-based representations are used to select an appropriate lexical representation. Phonological encoding associates this lexical representation with a form-based planning representation. Phonetic encoding or articulatory processing then implements this plan as a set of movements of the articulators.

Most contemporary perspectives on speech planning agree that speaking involves interaction among stages of formulation. At each stage of processing, multiple representations are co-activated, and subsequently influence the following stage of processing. For example, many studies have shown that the process of lexical selection results in the co-activation of multiple semantically related words (e.g., Peterson & Savoy, 1998). These semantic cohort members influence subsequent phonological encoding. Other work has shown that disruptions originating in phonological planning extend to phonetic processing, altering the phonetic properties of speech (e.g., Goldrick & Blumstein, 2006; McMillan, Corley, Lickley, 2009; Pouplier, 2007). However, evidence for long-distance interactions—effects of disruptions to conceptual processes and lexical selection that influence articulatory processing—have been inconsistent. The current work provides new evidence on such interactive effects, examining the influence of semantic competitors on articulation during picture naming.

Evidence for interactions between adjacent levels of formulation

Abundant evidence supports the idea that lexical selection processes interact with phonological planning. Specifically, semantically related competitors activated during lexical selection activate their corresponding phonological representations (for reviews, see Goldrick, 2006; Melinger, Branigan, & Pickering, 2014). For example, in the picture-word interference paradigm (Schriefers et al., 1990), picture naming is disrupted by the presentation of an auditory or visual distractor word. Distractors that are phonological relatives of a semantic competitor show evidence of priming, suggesting their phonological representations have been activated during target processing (e.g., during processing of target *couch*, the semantic competitor *sofa* primes *soda*; Cutting & Ferreira, 1999; Peterson & Savoy, 1998; Taylor & Burke, 2002). Further support for the semantically-driven activation of phonological representations comes from studies of speech errors showing that mixed errors (sharing both semantic and phonological structure with the target) occur at a higher rate than predicted by the rate of pure semantic or phonological errors (Dell & Reich, 1981; Rapp & Goldrick, 2000).

Interactive effects are also found between phonological planning and articulatory processing. Speech errors reflect a blend of articulatory/acoustic properties of the target and error outcomes (e.g., when producing target *big* as "pig," the production of /p/ reflects a blend of the intended /b/ and error outcome /p/; Frisch & Wright, 2002; Goldrick, Baker, Murphy, & Baese-Berk, 2011; Goldrick & Blumstein, 2006; Goldrick, Keshet, Gustafson, Heller, & Needle, 2016; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; McMillan & Corley, 2010; McMillan, et al., 2009; Pouplier, 2007, 2008). Such effects can be attributed, in part, to the partial activation of the target representation during phonological planning (see Goldrick et al., 2016, for review and discussion). Similar effects are found when phonological competitors are primed not within production processes but by comprehension processes. Yuen, Davis, and Brysbaert (2010) examined articulatory processing during reading aloud of a target while participants listened to a matching (i.e., identical) syllable or a phonologically related (rhyming) competitor. Articulatory processing of the target sound was distorted when a competitor was presented, such that articulation reflected a blend of the target and the initial sound of the spoken competitor.

A key study with respect to the work reported here is Drake and Corley (2015) who examined picture naming when phonologically related competitors were primed by sentence preambles. Participants heard sentences like *Jimmy used a washer to fix the drip from the old leaky…* (priming a mismatching word, *tap)* and *On his head he wore the school...* (priming the target word, *cap*). In both cases, participants then named a picture of a cap. Articulations in these two conditions were compared to a baseline: picture naming with no sentence preamble. Productions following unrelated primes showed greater difference from baseline than those following target primes, suggesting that the primed competitor disrupted articulatory processing of the target.

Evidence for long-distance interactions: The influence of disruptions to lexical selection on phonetic processing

Given the evidence that lexical selection interacts with phonological processing, which in turn interacts with articulatory processing, one would expect long-distance interactions between lexical selection and articulation. In this work, we focus on how disruptions to lexical selection influence phonetic processing. We review evidence from paradigms using conditions that slow reaction times and/or increase errors relative to baselines, deferring discussion of facilitatory effects until the following section.

Kello, Plaut, and MacWhinney (2000) adapted the Stroop task to examine how lexical selection disruptions modulate phonetic processing. In the typical color-word version of the Stroop task, written words specifying color concepts are presented. The word is printed in color font and only the color of the font is to be named aloud. In some conditions, the color is congruent with the conceptual representation of the written word (e.g., say *green* to the word "GREEN" presented in a green colored font) or neutral (e.g., say *green* to XXXX presented in a green colored font). The Stroop interference effect refers to the fact that relative to the two conditions above responses are initiated more slowly when the color of the font is incongruent with the meaning of the written word (e.g., say *green* to the word "RED" presented in a green colored font). Kello et al. used this paradigm to examine long distance interactions from lexical selection to articulation. Stroop interference lengthened reaction times as well as spoken word durations – but only when speakers were pressured to respond quickly. To account for their findings, Kello et al. proposed a dynamic interaction hypothesis. This claims that interactive effects extending from lexical selection to phonetic processing will be strongest under processing circumstances that allow insufficient time for speakers to resolve disruptions during lexical selection. Time pressure increases temporal overlap between processes, increasing interaction.

However, other work has found no evidence supporting such effects. Damian (2003) failed to replicate Kello and colleagues' (2000) Stroop task results; even under time pressure, there was no increase in duration of words subject to Stroop interference. Furthermore, he failed to show articulatory effects in two additional tasks. Using the picture-word interference task, participants named pictures aloud in the presence of semantically-related auditory distractors (Schriefers et al., 1990). Although the presence of distractors increased reaction times, Damian found no effect of distractors on spoken durations, even when speakers were pressured to respond quickly. Above and beyond the empirical uncertainty regarding effects in durations, there is debate over whether these two paradigms actually tap lexical selection processes (Roelofs, 2014) or arise in processes external to lexical selection (e.g., Dhooge, De Baene, & Hartsuiker, 2013; Finkbeiner & Caramazza, 2006a, 2006b). That is, it is unclear whether these findings speak to the question of interaction between lexical selection and articulation or rather to articulatory interactions with other aspects of formulation and more general cognitive processing.

The semantic blocking paradigm (Belke, Meyer, & Damian, 2005; see also Damian, Vigliocco, & Levelt, 2001) induces inhibitory effects that are more widely assumed to arise within lexical selection (Oppenheim, Dell, & Schwartz, 2010). (Note the paradigm can also give rise to facilitatory effects; Belke, 2017.) Pictures are presented in homogeneous blocks (including only pictures from the same semantic category) or heterogeneous blocks (including pictures from a mixture of categories). The so-called blocking effect reflects increased response latencies in homogeneous blocks relative to heterogeneous blocks. Damian (2003) found a very robust blocking effect observed in measures of response latency, but no significant effects in response duration, even under time pressure. However, using a much larger sample of participants (n = 96 compared to 24 in each condition of Damian's 2003 study), Fink, Oppenheim, and Goldrick (2018) found effects of semantic blocking on word durations (with longer durations in homogeneous vs. heterogeneous blocks). Interestingly, this effect was detected only when individual differences in susceptibility to the blocking effect were taken into account: individuals that showed large blocking effects in their reaction times exhibited semantic effects in their word durations; those participants with small blocking effects in reaction times showed no duration effects. Fink et al. (2018) also found evidence in favor of long-distance interactions between lexical selection and articulation using another paradigm, continuous picture naming, which also induces semantic interference effects during lexical selection (Howard, Nickels, Coltheart, & Cole-Virtue, 2006). In this paradigm, participants name pictures from intermixed sets of semantic categories. Within each category, response latencies increase with each successive member of the category. Consistent with long-distance interactions, Fink et al. found that increases were also found in word durations.

Mixed results have also been found in the manual articulatory domain in studies of typing. Logan and Zbrodoff (1998) found that Stroop interference effects impacted typing latencies, but not durations. Parallel to Kello et al. (2000) and Damian (2003), Damian and Freeman (2008) examined Stroop effects under response pressure; similar to Damian (2003), Damian and Freeman found no effects on typewritten response durations either with or without time pressure. However, in a regression-based analysis using a large sample of participants (n = 86) and a diverse array of pictures (n = 260), Scaltritti, Arfé, Torrance, and Peressotti (2016) found that variables influencing lexical selection (word frequency and name agreement) influence both response latencies and typing durations in written picture naming.

Note that all of the studies reviewed above (in both spoken and manual modalities) have focused on duration of the entire response. However, this might obscure interactive effects if they are present only in certain portions of the word. For example, suppose response initiation occurs as soon as the first element (e.g., initial segment, letter) is planned, but planning continues while it is being articulated (i.e., response planning for different components of a word occurs in parallel). During articulation of the initial element, continued planning of subsequent elements in the word might allow such element to overcome the effects of any delays or disruptions. Effects may therefore be limited to initial portions of articulation and dissipate at later positions

(Kawamoto, Kello, Higareda, & Vu, 1999). There is some evidence from typing studies consistent with this possibility. Scaltritti, Pinet, Longcamp, and Alario (2017) found that while semantic priming did not significantly influence the whole-word duration of typewritten responses, there was an influence on duration of the initial inter-keystroke-intervals. However, this effect was not reliable in some of the subset analyses they performed, and the study of Stroop effects by Damian and Freeman (2008) found no effects in initial position or whole word durations. Furthermore, using electroencephalographic measures, Scaltritti et al. failed to find evidence that semantic priming interacted with motor response preparation.

Another important limitation of the work reviewed above is that focuses on young, monolingual participants, whose formulation abilities are likely operating at peak efficiency. In particular, young adults may possess strong selection processes, which serve to enhance the activation of a single lexical representation relative to its co-activated competitors (via boosting target activation and/or inhibiting non-target activation). The reduced strength of competitors relative to targets will significantly reduce the strength of interactive effects (see Dell & O'Seaghdha, 1992; Rapp & Goldrick, 2000, for discussion and simulation data). Consistent with this claim, Jescheniak, Hahne, Hoffmann, and Wagner (2006) found that children (~7 years of age) were more susceptible to interactive effects than young adults (~24 years old). Interactive processing accounts predict that cascading activation from the semantic representation of a target (e.g., *cat*) will activation category coordinates (*dog*) which will in turn activate phonologically related words (*doll*). Critically, the strength of such effects will depend on the relative activation of target vs. non-target representations. Jescheniak et al. found that children showed significant effects, whereas young adults did not; this is consistent with the claim that children have weaker selection processes than young adults. The focus of the literature on long-distance interactions on young adults may therefore have reduced our ability to detect effects.

Facilitation of formulation and its impact on phonetic processing

While the above discussion has focused on disruptions, other work has focused on how phonetic processing may be facilitated by ease of formulation. A large number of studies have shown that words that are more predictable with respect to the linguistic and/or discourse context[1] in which they appear are retrieved more quickly and accurately (suggesting facilitated processing) and produced with reduced phonetic forms (e.g., shorter acoustic duration, de-accented variants, centralization of vowels, etc.; see Arnold, 2016, for a recent review). Some accounts have claimed that such effects are due, in part, to long-distance interactive mechanisms; context-driven facilitation of the retrieval of representations at the lexical or conceptual level modulates the activation of word form and phonetic representations, producing reduction (e.g., Balota, Boland, & Shields, 1989). However, other accounts have emphasized more local interactions, presenting evidence that facilitation of phonological retrieval, over and above facilitation to lexical or conceptual processing, is the key mechanism within the production system that results in phonetic reduction. Jacobs, Yiu, Watson, and Dell (2015) examined reductions in time to initiate production of a target word as well its duration when the target was repeated. They contrasted conditions where the first utterance of a word was fully articulated vs. read silently. Critically, while target word initiation was facilitated to the same degree in both conditions, reduction in target duration only occurred when the word was read aloud. Jacobs et

[1] In contrast to the context-specific effects of disruption examined here, other work has examined how acoustic properties are related to context-independent features of words (e.g., word frequency: Gahl, 2008; phonological neighborhood density: Gahl & Strand, 2016; informativity: Seyfarth, 2014). Note that these effects may due to mechanisms that overlap with those driving context-specific predictability effects (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009).

al. concluded that repetition of any kind was sufficient to facilitate lexical and conceptual processing (reducing the time required to initiate production). In contrast, facilitation of phonetic processing (i.e., reduction) requires facilitation specifically of word form processing.

These findings suggest that independent of the presence or absence of effects of disruption, there are widespread situations where phonetic processing is facilitated by the context in which a word is produced. However, in situations where processing at multiple levels of representation are enhanced, it is unclear whether phonetic effects arise due to long-distance or more local interactive mechanisms.

The current study

In the current study, we utilized a sentence completion paradigm (Ferreira & Griffin, 2003) to examine long-distance interactive effects. Ferreira and Griffin used visually presented sentence preambles to prime competitors during picture naming. Participants in their study read sentences like *The woman went to the convent to become a….* (priming *nun*) and then attempted to name a picture of a priest. These primes disrupted processing, resulting in the overt production of semantic errors. Semantically related primed words were produced significantly more often than control trials where the sentence primes a semantically unrelated word (e.g., *He lit the candle with just one…* priming *match*; here, participants had less difficulty naming *priest*). Interestingly, errors were also produced at a rate greater than control trials when the sentence primes a homophone of a semantic competitor (*I thought that there would be some cookies left, but there were…* priming *none*), but there was no increase in errors when the sentence primed a purely phonologically related word (*present*; Ferreira & Griffin, 2003; Li & Slevc, 2017; Severens, Ratinckx, Ferreira, & Hartsuiker, 2008). The fact that homophones also induced substitutions suggests that the processing disruption that leads to semantic errors in this task arose specifically at a post-semantic, lexical level of processing (where homophones share representations) but prior to phonological processes manipulating sub-lexical units of form (accounting for the absence of phonological errors). Thus, in contrast to paradigms such as Stroop and picture-word interference, there is clear evidence that this paradigm can induce disruption specifically within lexical selection processes.

As discussed above, using a similar sentence prime paradigm, Drake and Corley (2015) found articulatory interference during picture naming after priming by a sentence stem predicting a semantically unrelated word (n.b. interference was relative to an unprimed baseline). This finding suggests that sentence primes serve to activate representations inconsistent with the target at semantic, lexical, and form-based levels; the activation of these competing representations produces articulatory disruptions. Note that while matching primes could also have facilitated processing, Drake and Corley (2015) found no difference between the articulation of matching primes and an unprimed baseline. This suggests that in this type of paradigm, interference effects dominate processing.

Here, we examine the impact of semantically related primes. While such primes activate semantic representations that overlap with the target, the pattern of errors reviewed above suggest that they produce *enhanced* conflict (relative to unrelated primes) specifically at the lexical level. Comparing articulation of targets following semantically related vs. unrelated primes will therefore provide an index of long-distance interactive effects: how enhanced disruption of lexical selection impacts articulatory processing.

Our design also took into account three other factors that may have contributed to the mixed results observed in previous work. In Experiment 1, we examine healthy, young adult monolinguals' patterns of response time and articulation in this paradigm. In this (and subsequent) experiments, we extended previous work by examining articulatory properties of the whole word as well as properties specific to the initial segments of the word (where, as noted above, interactive effects might be strongest). In Experiment 2, we examined whether interactive effects were modulated by pressure to respond (which, as noted above, has been suggested to increase temporal overlap and interaction between processes).

Finally, in Experiment 3, we examined these effects in healthy, monolingual older adults (parallel to Experiment 1, under no explicit time pressure). As noted above, there is evidence that across the lifespan there are changes in the strength of lexical selection, such that young adults show weaker interactive effects than children (Jescheniak et al., 2006). There is some evidence that formulation processes undergo declines as a consequence of normal aging, as indexed by a higher rate of tip-of-the-tongue (TOTs) retrieval failures (see Gollan & Brown, 2006, for a review) and higher rates of speech errors (Gollan & Goldrick, submitted; but see Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). One account of effects such as these is the *Inhibitory Deficit Hypothesis*, which claims that aging leads to difficulty in suppressing inappropriate responses (e.g., Zacks & Hasher, 1994). If these domain-general inhibitory mechanisms are utilized during lexical selection, their age-related decline would allow for greater activation of non-target lexical representations in older vs. younger adults. Cascade from these representations would be predicted to strengthen interactive effects in the older adults. However, it should be noted that there are other accounts of aging deficits that predict *decreased* interactive effects. Specifically, the *Transmission Deficit Hypothesis* (e.g., Burke, MacKay, Worthley, & Wade, 1991) proposes that language production difficulties in older adults arise due to reduced flow of activation between lexical and phonological representations. This account therefore predicts that there should be less activation of non-target representations at the phonological level, and therefore less disruption of articulatory processing.

This design did not eliminate all potential issues. To facilitate group comparisons, we recruited the same number of participants across groups. We based the sample size for all groups ($N = 18$) on that used in previous studies with this paradigm (Severens et al., 2008). This sample size was achievable given practical limitations on our recruitment of older adults. It's possible this does not provide sufficient power for detecting long-distance interactive effects; replicating our findings with larger groups is an important area for future work.

To summarize, our study includes 3 experiments examining how disruptions to lexical selection modulate articulatory processing. Experiment 1 examines effects in younger adults, examining effects on whole word durations as well as specific properties of initial segments. Experiment 2 aimed to increase interactive effects by increasing response pressure. Experiment 3 aimed to increase effects by testing older adults; difficulties these individuals may have in inhibiting non-target representations would increase interactive effects.

Acoustic Analysis Methods

Key to our study is the measurement of phonetic properties of productions. Using a combination of algorithms (all available at https://github.com/MLSpeech), we limited manual processing in the analysis pipeline. First, participants' speech was recorded on one channel of a stereo recording and the second channel simply recorded when pictures were presented for naming. These clicks were used to automatically segment the original audio stream into separate files containing the signal from each individual trial. After segmentation of each trial's data, several algorithms were combined to extract the phonetic variables of interest. We first used two algorithms to estimate several key time points in the signal:

- Word onset and offset: We developed a novel algorithm (DeepWDM, short for Deep Word Duration Measurement, described below) that, given a signal consisting of speech preceded and followed by 'silence' (minimally noisy non-speech signals), would automatically determine the onset and offset of the word.
- Vowel onset and offset: Given speech consisting of a vowel with one or more flanking consonants on each side, the AutoVowelDuration algorithm (Adi, Keshet, Cibelli, Gustafson, Clopper, & Goldrick, 2016) outputs the onset and offset of the vowel. In monosyllabic words, this can operate without any additional processing. In disyllabic words, the AutoAligner forced aligner (Keshet, Shalev-Shwartz, Singer, & Chazan, 2007; McAllester, Hazan, & Keshet, 2010) was used to determine the location of the initial syllable (always the location of the stressed syllable in this dataset), and then AutoVowelDuration was used to determine the precise location of the vowel onset and offset.

Once these time points had been determined, several duration measures could be extracted: reaction time (the duration between trial onset and word onset); word duration (time between word onset and offset); duration of initial consonant or consonants (time between word onset and vowel onset, for consonant-initial words only); and vowel duration (time between vowel onset and offset). Examination of whole word, initial consonant, and vowel durations allows us to examine both overall effects of articulatory disruption as well as effects that may specifically target the initial segments of the word. Finally, the DeepFormants algorithm (Dissen & Keshet, 2016) was used to estimate first (F1) and second (F2) formant values within the window identified by AutoVowelDuration. Measuring these spectral qualities gives us another index (beyond duration) of vowel articulation. Disruption of processing was indexed by vowel dispersion (calculated as the Euclidean distance from the overall F1/F2 midpoint of the vowel space, within-subject; Lofqvist, Sahlen, & Ibertsson, 2010). Based on previous work, we predict that disruptions to formulation will lead to greater vowel dispersion (i.e., lower distance from the overall midpoint; see Munson, 2007, for discussion).

In the remainder of this section, we describe in detail the structure of the novel DeepWDM algorithm; detailed characterization of the other speech processing algorithms can be found in the publications cited above.

Problem setting

The input to our algorithm is an acoustic signal containing one dominant speech portion (i.e., the uttered word) which can be surrounded by noisy non-speech signals. (Such non-speech noise is a persistent challenge to voice key systems that simply rely on signal intensity to determine speech onset.) The output is the onset and offset times of the speech portion. The acoustic signal can be of an arbitrary length, and its beginning does not need to be synchronized with speech onset.

Let $\boldsymbol{x} = (x_1, \ldots, x_T)$ denote the input acoustic signal, represented as a sequence of feature vectors, where each $x_T \in \mathbb{R}^D (1 \leq t \leq T)$ is a *D*-dimensional vector. The length of the speech portion, T, is not a fixed value since the acoustic signals and target words can have different durations.

Each acoustic signal is associated with a timing pair, denoted by $\boldsymbol{t} = (t_b, t_e)$, where t_b and t_e are the onset and offset of the speech portion respectively (see Figure 1). Our goal is to predict the onset and offset times of the speech portion as accurately as possible.

Model

One approach to determining the duration of a phonetic property is to predict at each time frame whether the property is present or absent; the predicted duration is then the smoothed, continuous set of frames where the property is likely to be present (Adi, Keshet, Dmitrieva, & Goldrick, 2016; Adi, Keshet, & Goldrick, 2015). In this work, we follow this method with generating predictions using a Recurrent Neural Network (RNN).

Learning Model. In order to predict the voice activity's onset and offset (i.e., speech onset and offset) we trained a RNN (Elman, 1991), as a speech detection system. The input to the network is a sequence of T tuples, where each tuple is composed of the feature vector x_t and a corresponding label y_t, from the set of $\{1,-1\}$, for $1 \leq t \leq T$ as follows:

$$y_t = \begin{cases} -1 & 1 \leq t \leq t_b \\ 1 & t_b \leq t \leq t_e \\ -1 & t_e \leq t \leq T \end{cases}$$

We label every frame that is placed inside the boundaries of the speech portion of the acoustic signal as positive and every frame that is outside of those boundaries as negative.

Our RNN model is composed of two stacked layers of bi-directional Long-Short Term Memory (LSTM) Units (Hochreiter & Schmidhuber, 1997), which have shown remarkable results in modeling speech sequences (Graves & Jaitly, 2014; Graves, Mohamed, & Hinton, 2013). The inputs to the network were 39 Mel-Frequency Cepstrum Coefficients (MFCCs), including delta and delta–delta, extracted every 10 milliseconds. In order to avoid overfitting, a dropout layer (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012) is used after each recurrent layer, with rate of 0.5.

Training. The training data consisted of 2,369 hand-annotated productions of single words (drawn from Fink, 2016). Participants named a set of 90 pictures ("carrot", "violin", etc.) in sequence. Ten percent of the data was used as a validation set for tuning model parameters (including hyper-parameters). We optimized the negative log-likelihood loss function using

Adagrad (Duchi, Hazan, & Singer, 2011) with a learning rate of 0.01. Training was stopped after 5 epochs with no loss improvement on the validation set.

Inference. The RNN outputs a probability for each class (speech vs. non-speech) every frame, which can be used to characterize a probability distribution over all possible sequences. In order to extract the onset and offset times from the RNN outputs we perform three steps. First, we predict the class with the largest probability in every frame. Second, we remove noisy predictions by smoothing the predictions using a window of 10 frames. Finally, since we know that in every sequence there is one major voice activity which we are interested in, we output the timing pair with the longest duration.

Validation: To assess performance of the DeepWDM algorithm, novel data (not used in model training) from Fink (2016) was used to compare manual and algorithmic measurements of word duration on 6641 tokens. The correlation between manual and algorithmic measures was 0.72; the mean absolute deviation was 56 msec (*SD* = 73 msec). This level of performance is well within that of the vowel duration algorithm, which our previous work has shown faithfully reproduces results from behavioral data (Adi et al., 2016).

Overview of the Experiments

As discussed above, in this paradigm each trial consists of a visually presented prime sentences followed by picture; participants are asked to orally produce the picture name. Table 1 summarizes our design, which limits non-target primes to semantically related and semantically/phonologically unrelated items. The key findings are summarized in Table 2.

Experiment 1:

Young Adults Naming

Methods

This experiment along with the following two were approved by the Northwestern University Institutional Review Board.

Participants

We recruited 18 participants at Northwestern University using the Linguistics Department subject pool. Participants received course credit. They reported learning no language other than English before age 5 and no history of color blindness or language impairment. Age ranged from 18-24 (*M*=19.6, *SD*=1.5).

Materials

Details of norming procedures for picture and sentence stimuli can be found in Appendix A. One hundred eighty colored photographs were selected from a larger pool of photographs retrieved from the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and Google Images. Each picture had a name agreement of at least 75%. The selected pictures had an average word frequency of 32.2 words per million (from the SUBTLEX-US corpus, Brysbaert & New, 2009, extracted from the CLEARPOND database; Marian, Bartolotti, Chabal, & Shook, 2012). All pictures were 300 x 300 pixels. In addition, 180 unique sentence fragments were created for the experiment. For each picture, we constructed a cloze probability sentence. Sentences were normed with both younger adults and older adults (see Appendix B for details); only sentences with at least 45% cloze agreement were selected. For younger adults, the average cloze probability for all sentences was 90.7% (*SD* = 11.9%). Sentences ranged in length from 5 to 15 words (*M*=8.7, *SD*=2.0). Sentences and target picture names can be found in Appendix B.

A small percentage of sentence contexts (1.7%) had an indefinite article that mismatched with the onset of the picture completion. Visual inspection of response times for these trials indicated that they were not different from trials where no mismatch was present, and excluding these trials from analysis did not appear to qualitatively impact the results.

Design and Procedure

Participants were presented high probability cloze sentences, one word at a time. Each sentence was followed by a picture that was to be named aloud. As shown in Table 1, pictures were paired with one of three sentences: match (where the sentence primed the picture name), or one of two mismatching sentences – competitor (where the sentence primed a semantically related word) and unrelated (where the sentence primed a phonologically and semantically unrelated word).

Each participant named 60 pictures three times, for a total of 180 trials. Across the blocks of picture naming, each appearance of a given picture was paired with a different prime sentence (reflecting the three conditions). Note that sentences were not repeated across blocks so as to minimize experiment-specific expectancy effects. Within a given block, the number of trials was evenly divided between the three conditions. The order of conditions for each picture was counterbalanced across lists.

Participants were tested individually in a sound-proof room. They first provided informed consent and completed a background questionnaire. Speech during the experiment was recorded using a head-mounted microphone. After the experimental task was completed, participants completed a measure of receptive vocabulary (the Shipley-2 Institute of Living Vocabulary Subscale; Shipley, Gruber, Martin, & Kline, 2009) and a separate measure of productive vocabulary (the Multilingual Naming Test; Gollan, Weissberger, Runnqvist, Montoya & Cera, 2012). These helped us control for any effects of differences in vocabulary knowledge on lexical processing (e.g., Mainz, Shao, Brysbaert, & Meyer, 2017).

Sentence Prime with Picture Naming Task. Each trial began with a fixation cross (+) presented in the center of the screen for 500 ms. The fixation was followed by the first word of the sentence fragment. Subsequently, the remaining words of the sentence were presented one at a time at the center of the screen in standard rapid serial visual presentation fashion. Each word remained on the screen for 275 ms. After the final word of the sentence fragment was presented, a picture appeared and remained on the screen for 600 ms. Participants were instructed to read the words within the sentence silently for comprehension and to name the picture aloud before it disappeared. If the participant did not respond within 600 ms, their response could be registered for an additional 300 ms, during which a blank screen was displayed. An inter-stimulus interval of 1500 ms occurred between trials.

Multilingual Naming Test (MINT). Immediately following the picture naming in context task, participants completed the Multilingual Naming Test to measure individual differences in native language vocabulary knowledge. Participants were shown a set of 68 black and white line drawing images and instructed to name each image aloud as quickly as possible. Participants were given two different kinds of prompts if they gave an incorrect response. A semantic cue was provided, in the form of a brief definition of the object. If participants did not retrieve the correct word after receiving the semantic cue, they were also provided a phonological cue, the first letter of the response. If participants still could not respond with the correct name, the response was marked as incorrect and they were instructed to move on. Pictures were presented in an order of ascending difficulty. Score on the test is number of pictures correct.

Shipley-2 Institute of Living Vocabulary Subscale. The Shipley-2 Test comprises 40 stimulus words, presented in generally ascending order of difficulty. Participants selected the word that was the closest synonym to the stimulus word from among four presented options. Score on this test is a standardized score (with 100 indicating average performance on demographically matched sample).

Data on these vocabulary measures is shown in Table 3. Two sample heteroscedastic t-tests (using the Welch-Satterthwaite correction) were used to compare scores from Experiments 2 and 3 to the Experiment 1 baseline. No differences were found in MINT scores (ts < 1.5, ps $>$ .15). Young adults in Experiment 2 had higher Shipley-2 scores than participants in Experiment 1 (t (31.013 = 3.291, $p < .005$) and older adults in Experiment 3 showed a similar trend relative to the Experiment 1 baseline (t (29.370) = 1.786, $p < .09$). We therefore included Shipley-2 scores as a covariate in our analyses.

Results

Errors

Responses were categorized as correct or one of four types of errors: a) name agreement errors (production of names that differed from those designated by the experimenter); b) verbal disfluencies (stuttering, utterance repairs, and production of nonverbal sounds); c) omissions; and d) completion errors (where the sentence prime completion was produced instead of the picture name). We assessed inter-rater reliability on error classification by taking data from a random selection of 4 of the 54 participants across the 3 experiments. Two raters agreed on response classification for 98.5% (n = 720).

Participants were quite accurate, with a mean of 8.6% of trials (s.e. 0.9%) eliciting errors. Given the fairly high rate of agreement errors (mean 5.7% of trials, s.e. 0.7%) we excluded from analysis pictures that elicited 60% or more name agreement errors across Experiments 1 and 2 (*beaver, canoe, cheetah, dropper, jeep, leg, peeler, raccoon, sheep*). With these items removed, the average cloze probability of the target sentences (as assessed by younger adults) increased from 90.7% to 91.1%.

After removal of items with low name agreement, 17 completion errors remained in the data set (0.5% of trials); of these, 15 were in the semantic competitor condition, and 2 in the unrelated condition. This was not sufficient to fit a regression model; however, this pattern indicates that participants were more likely to make a completion error when a semantic competitor was primed, as compared to an unrelated prime.

Acoustic properties: Data analysis methods

All items excluded in accuracy analyses were also excluded from the articulation models. Six additional items were also excluded (*cookie, mop, leg, olive, tire,* and *foot*) because participants frequently named a competitor for these items. Once these data were excluded, outlier removal on each dependent variable was conducted, by removing trials with measurements 3 standard deviations above and below each participant's mean. We first removed response time and word durations outliers from the entire data set, and fit models to these dependent variables using this set of data (98.2-98.5% of data retained within each experiment). At this point, the data set contained 7462 tokens (2168 in experiment 1, 2045 in experiment 2, and 2136 in experiment 3). Then, in conducting more exploratory analyses of sub-word components, we separately removed outliers for these three dependent variables: first consonant duration (98.5-98.9% of remaining data retained), initial consonant duration (99.1-99.6%

retained), and vowel distance (99.1-99.4% retained)[2]. For the initial consonant duration model, tokens which were word-initial were also removed; after this, 85.1% of the data was retained (6349 data points across all three experiments).

Models were fit in R using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014). Fixed effects of each models are described in the sections for each dependent variable that follow. Selection of random effects followed Bates, Kliegl, Vasishth, and Baayen (2015); models were initially fit with the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013), and PCA was used to identify components of the structure that did not contribute variance to the model fit. After model selection, each model was refit excluding data points with extreme residuals (> 2.5 s.d., following Baayen, 2008). Likelihood ratio tests were used to assess the significance of fixed effects, as they are less anti-conservative than t-as-Z tests (Barr et al., 2013).

Results: Acoustic properties

Below, we discuss key predictions of the theoretical accounts discussed above (repeated below; see Table 2). The full results of the linear mixed effects regressions can be found in Appendix C. In assessing these effects, we included a series of control variables in our regression model:

---

[2] We have also conducted models using more restrictive criteria, in which we restrict the data pool to only items where a response is available for all three conditions (match, unrelated, competitor) for a single participant. While this more conservative approach better controls the distribution of data across conditions, it resulted in the removal of nearly 30% of the data. We found the effects in this more conservative analysis to be broadly consistent with the analyses reported here, so we have chosen to present this analysis, which is a more complete data set.

- *block, along with interactions of block with match status and semantic relatedness*: controls for any effects due to the repetition of picture targets across blocks (e.g., repetition reduction; Baker & Bradlow, 2009).
- *Shipley-2 score*: controls for effects of overall differences in vocabulary on lexical processing (as noted above).
- *trial-level response time*; *overall response speed of participant*: if articulation is ballistic, such that all effects of formulation are fixed at the moment of response initiation, planning measures (i.e., reaction time) should be positively correlated with acoustic measures (e.g., word duration), and there should be no remaining independent effects of formulation variables on the acoustic measures (see Buz & Jaeger, 2016; Fink et al., 2018, for discussion; see Strijkers & Costa, 2016, for additional discussion). We control for this in two ways: response time for each individual trial and a measure of the overall response speed of an individual. The latter is estimated by the best linear unbiased predictors (BLUPS; Baayen, 2008) of by-participant intercepts in a mixed effects model of RTs (detailed in Appendix C).

With these factors under statistical control, we examine the key predictions tested in our study. Table 4 provides descriptive statistics for our principal dependent measure, word duration, in each condition.

Interaction of formulation and articulation. As predicted, mismatch trials had longer word durations than match trials ($\beta = 18.783$, χ^2 (1) = 30.09, $p < 0.001$), suggesting that formulation disruptions lead to articulatory disruptions.

Long-distance interaction. There was no significant difference between word durations following unrelated vs. semantically related primes (β = –4.268, χ^2 (1) = 2.33, p = 0.127), suggesting that disruptions to lexical selection do not yield enhanced articulatory disruptions.

Dynamic interaction. The match vs. mismatching primes effect did not significantly interact with reaction time (β = –19.718, χ^2 (1) = 1.56, p = 0.212), suggesting that overlap between formulation and articulation did not increase when responses were more speeded. There was a significant positive interaction of semantic relatedness and reaction time (β = 50.667, χ2 (1) = 9.84, p = 0.002). This was driven by a greater relationship between reaction time and word duration for semantically related (r = 0.104) vs. unrelated primes (r = 0.0597). Critically, unrelated trials were particularly longer than semantically related trials at fast RTs (below the median RT, difference in word durations = 3.3 ms) as compared to slow RTs (above the median RT, mean difference = 2.1 ms; note that in the model RT was entered as a continuous factor). This suggests that across the range of RTs semantically related primes always have less of an effect on articulation than unrelated primes – contra the predictions of long-distance interaction.

Initial segment speech initiation. A series of regression models (including the control variables above) examined whether these condition effects could be found in the duration of initial consonants and following vowel, as well as in spectral properties of the vowel. Full model specifications and condition means for each measure can be found in Appendices C and D respectively.

The effect of match status was significant for initial consonants (β = 10.852, χ^2 (1) = 16.90, p < 0.001) but not for vowel durations (χ^2 (1) < 1, *p* > .80). Semantic relatedness did not significantly affect either duration measure (χ^2s (1) < 2, *p*s > 0.15). Figure 2 provides a visualization of the relative effect sizes on the three duration measures. This suggests that, as predicted by an account where speech is initiated prior to completing planning, the effects of formulation disruptions on articulation are largest for initial consonants.

There was mixed evidence as to whether these duration effects increased in faster responses. For initial consonants, there was a significant negative interaction of match status and reaction time ($\beta = –31.232$, χ^2 (1) = 9.87, $p = 0.002$), such that disruptions caused by mismatch effects increased at shorter reaction times (as predicted by dynamic interaction accounts). For vowel durations, there was no significant interaction of match status and reaction time (χ^2s (1) < 1, *p*s > 0.40). However, there was a significant interaction of response time and semantic relatedness (as is found in the whole word analysis; $\beta = 20.607$, χ^2 (1) = 6.99, $p = 0.008$).

Finally, with respect to spectral properties of the vowel, there was no significant effect of match status or semantic relatedness (χ^2s (1) < 1, *p*s > 0.10).

Discussion

Experiment 1 provided evidence consistent with interactions between formulation and articulation, but no evidence for long-distance interaction effects. As predicted by accounts where speech is initiated as soon as the initial segment is planned, these formulation-articulation interactions appeared strongest for initial segments.

There was no evidence that these interactive effects were dynamic. This may be due to our reliance on natural, planned variation in response time. In Experiment 2, we followed the design of previous work argued to support dynamic interaction (e.g., Kello et al., 2000) and tested young adults naming pictures with explicit time pressure. This stronger manipulation should increase the chances that the degree of overlap between planning and articulation will increase.

Experiment 2:

Younger Adults Naming Under Explicit Time Pressure

Methods

Participants

We recruited 18 younger adult participants (9 male, 9 female) at Northwestern University (NU) using the Linguistics Department subject pool and ads recruiting participants on campus for monetary payment. Participants received either course credit or were paid $10/hour. They reported learning no language other than English before age 5. Age ranged from 18-22 years (*M*= 19.72, *SD*=0.89).

Materials

The sentence completion materials were identical to Experiment 1.

Design and Procedure

The design and procedure was nearly identical to Experiment 1, with the following exceptions. Following Severns et al. (2008, experiment 3), a deadline was imposed for initiating the naming response. Pictures appeared for 600 ms. Participants were instructed to name the picture before it disappeared. After the picture disappeared, a blank screen appeared and responses could still be registered for an additional 900 ms. When a software voice-key (with amplitude threshold adjusted for each participant) detected a response or when the 900 ms period ended, there was a blank screen for 1,500 ms. If participants did not initiate their response within 600 ms (as measured by the software voice key), a written message appeared on the right portion of the screen at the end of the trial to indicate that the response was too slow.

Results

Errors

Analysis indicated that participants were much less accurate in Experiment 2, with 20.8% of trials eliciting errors (compared to 8.6% in Experiment 1). Most of the errors were categorized as name agreement errors, with an error rate of 12.1%. However, participants made many more completion errors with time pressure. 4.6% of responses were completion errors. We identified 2.9% of the responses as verbal disfluencies and only 1.2% as omissions. After removing low name agreement pictures, 97.8% of the dataset was retained (2636 trials).

We analyzed the rate of completion errors (relative to correct trials) for semantic vs. unrelated primes. The average rate of completion errors for competitor trials was 9.716% (s.e.: 2.025%) versus 4.039% (s.e.: 1.303%) for unrelated trials. We fit a logistic mixed effects regression model to this data, with semantic relatedness and block as predictors (random effects included correlated by-subject slopes for semantic relatedness, block, and their interactions; random intercept for items) . There was a significant main effect of semantic relatedness (β = 5.770, χ^2 (1) = 9.25, p = 0.002), indicating that there were more completion errors from semantic vs. unrelated primes. There was a marginal main effect of block (β = -1.308, χ^2 (1) = 3.83, p = 0.051), indicating that errors decreased over the course of the experiment. The interaction of block and semantic relatedness was also significant (β = -1.143, χ^2 (1) = 6.25, p = 0.0124). The rate of errors following unrelated primes was relatively constant across blocks (Block 1: 4.18%; 2: 3.79%; 3: 3.61%), whereas the rate of errors on semantically related primes decreased over repeated presentations (Block 1: 14.4%; 2: 7.84%; 3: 7.26%). Because the error rate is relatively low, we confirmed this pattern by running a logistic regression of rare events model (Choirat et al., 2018; King & Zeng, 2001). This approach confirmed the critical significant effect of semantic relatedness ($\beta = 0.937$, $z = 4.404$, $p < 0.001$) and the marginal effect of block ($\beta = -0.242$, $z = -1.950$, p = 0.064); the interaction did not reach significance ($\beta = -0.330$, $z = -1.262$, $p = 0.207$).

Results: Acoustic properties

Analysis methods followed Experiment 1. We first note that our experimental manipulation not only increased error rates relative to Experiment 1, but also successfully produced decreased reaction times relative to Experiment 1 (see Appendix D). Table 5 provides descriptive statistics for our principal dependent measure, word duration, in each condition. (Comparison with Table 4 will reveal that word durations are overall shorter in Experiment 2 vs. 1.)

Interaction of formulation and articulation. Mismatch trials had longer word durations than match trials ($\beta = 12.717$, χ^2 (1) = 10.12, $p = 0.002$), suggesting that the conflicting representations activated by mismatched primes disrupted target articulation.

Long-distance interaction. There was a non-significant (marginal) *decrease* in word durations following semantically related vs. unrelated primes ($\beta = -5.298$, χ^2 (1) = 3.66, $p = 0.056$). As can be seen in Table 5, semantically related primes had essentially the same duration as match trials. Note this occurred in spite of an increase in the production of completion errors under speeded responding. This pattern is opposite that predicted by long-distance interaction; the disruption that leads to overt speech errors should lead to distortions in articulation.

Dynamic interaction. There was a non-significant (marginal) positive interaction of match vs. mismatching primes with reaction time ($\beta = 27.527$, χ^2 (1) = 3.45, $p = 0.063$); if anything, effects of mismatching primes increase with longer reaction times. In contrast, dynamic interaction accounts predict that interactions should increase with *shorter* reaction times (where there is greater overlap in processing). There was no significant interaction of semantic relatedness and reaction time ($\beta = 12.328$, χ^2 (1) = 0.76, $p = 0.382$).

Experiment 2 also allows a stronger test of the dynamic interaction hypothesis; a cross-experiment comparison between Experiment 1 (without explicit time pressure, yielding longer reaction times) and the current experiment (with explicit time pressure, yielding overall shorter reaction times). We examined this via a separate regression over data from both experiments. This was structured similarly to the overall model of word durations, with the addition of fixed effects for experiment and interactions of experiment and all other effects (full results can be found in Appendix C). Critically, the interaction of experiment with the effect of matching primes was not significant, nor was the interaction with semantic relatedness of the prime (χ^2s (1) < 0.10, ps > 0.80)[3]. This suggests that the main influence of explicit time pressure was to simply speed responses, not increase interactive effects.

Initial segment speech initiation. The effect of match status was significant for initial consonants ($\beta = 8.621$, χ^2 (1) = 15.36, $p < 0.001$) but not significant (marginal) for vowel durations ($\beta = 3.395$, χ^2 (1) = 3.85, $p = 0.05$). Figure 3 provides a visualization of the relative effect sizes on the three duration measures. Similar to Experiment 1, this suggests that the relative strength of condition effects is largest on initial consonants (although note that the range of effect sizes exhibits considerable overlap across measures).

This conclusion is tempered by a significant effect of match status on spectral properties of vowels ($\beta = 12.291$, χ^2 (1) = 6.01, $p = 0.014$); mismatch trials had greater vowel distances than

[3] There was a significant three-way interaction of experiment, match status, and RT; this reflected the non-significant negative interaction of match status and RT in Experiment 1 vs. the non-significant positive interaction in the current experiment. This does not provide evidence in favor of increased effects with decreased RTs. (Note a similar interaction was found for initial consonant duration as well.)

match trials, consistent with disruption to vowel articulation. Thus, while duration effects are larger on initial consonants vs. vowels, effects of formulation disruption persist into the vowel. There was no evidence that any of these by-position effects interacted with reaction time (see Appendix C for full results).

With respect to long-distance interactions in the context of initial segment speech initiation, the effect of semantic relatedness was significant for initial consonants (such that semantic competitors resulted in less interference than unrelated primes; $\beta = -3.759$, χ^2 (1) = 6.44, p = 0.011) but not vowel durations ($\beta = -0.284$, χ^2 (1) = 0.04, p = 0.841). Again, the direction of this effect is opposite that predicted by long-distance interactions. With respect to vowel distance, there was no significant effect of semantic relatedness ($\beta = 1.172$, χ^2 (1) = 0.07, $p = 0.790$). Additionally, there was no evidence that any of these by-position effects interacted with production speed (see Appendix C for full results).

Finally, to examine whether dynamic interaction effects would appear when comparing experiments with vs. without explicit time pressure, models compared each of these measures of vowel and consonant articulation across Experiment 1 vs. Experiment 2. There were no significant interactions of interactive effects with reaction time and experiment (see footnote 3 and Appendix C for full results).

Discussion

Parallel to previous work (Severens et al., 2008), imposing a deadline for responding produced greater error rates. The phonetic effects were largely parallel to Experiment 1. We saw clear evidence for interactions between formulation and articulation. Speech appeared to be initiated by planning of the initial segment, as disruption was particularly strong for initial consonants (but also extending into vowels). However, there was no clear evidence of long-distance interactions; furthermore, in spite of the inclusion of explicit time pressure, there was no evidence for dynamic interaction effects (parallel to Damian, 2003, but in contrast to the effects observed by Kello et al., 2000).

The finding that increased disruptions to lexical selection (due to time pressure) did not yield enhanced articulatory effects may reflect the overall stability of lexical selection in younger adults. We examine this in Experiment 3 by testing older adults. If, following the Inhibitory Deficit hypothesis, the older adults are less effective at suppressing non-target representations, interactive effects are predicted to be stronger than in younger adults.

Experiment 3:

Older Adults Naming

Methods

Participants

We recruited 18 older adult participants (3 male, 15 female) from communities in Chicago and Evanston, IL. Participants were paid $10/hour. They reported learning no language other than English before age 5 and no history of color blindness or language/cognitive impairment. Age ranged from 60-77 years (M= 68.38, SD=5.88).

Materials and Design

The sentence completion materials used here were identical to Experiment 1. We separately normed cloze probability in older adults; the average probability was 86.7% (s.d. 17.9%)

Procedure

The procedure was identical to Experiment 1.

Results

Errors

Analysis indicated that 14.9% of trials elicited errors. As in Experiments 1 and 2, most of the errors were categorized as name agreement errors (error rate = 11.4%). There were few completion errors (error rate = 0.5%). We identified 2.0% of the responses as verbal disfluencies and only 1.0% as omissions. We excluded 17 pictures from analysis that elicited a large number of agreement errors within this group of participants (*beaver, cheetah, couch, dollar, dolphin, fly, headphones, ipod, jeans, juice, laptop, nachos, speaker, vase, lime, peeler, mop, olive, ring, wheel)*. This allowed for retention of 90.3% of the dataset (2614 trials). Excluding these items the average cloze probability of the target sentences for older adults remained relatively stable, from 86.7% to 86.8%.

After removal of items with low name agreement, 14 completion errors remained in the data set (0.5% of trials); of these, all were in the semantic competitor condition. As such, it was not possible to fit a model to the error data in Experiment 3. The pattern matches those in Experiment 1, and indicates that participants were more likely to make a completion error when a semantic competitor was primed.

Results: Acoustic properties

Analysis methods followed Experiment 1. Four additional items were removed from articulation models because of high rates where participants named the competitor (*mop, olive, ring, wheel*). Table 6 provides descriptive statistics for our principal dependent measure, word duration, in each condition. (Comparison with Table 4 will reveal that word durations are overall longer in Experiment 3 vs. 1, consistent with slower speech in older adults.)

Interaction of formulation and articulation; long-distance interaction. In contrast to the experiments with young adults, collapsing across semantically unrelated and semantically related, mismatch trials did not have longer word durations than match trials ($\beta = 5.105$, χ^2 (1) = 1.31, p = 0.253)[4]. However, there was a significant *decrease* in word durations following semantically related vs. unrelated primes ($\beta = -11.304$, χ^2 (1) = 8.86, $p = 0.003$). As can be seen in Table 6, semantically related primes had essentially the same duration as match trials (with unrelated trials showing longer durations). So while the result with unrelated primes is consistent with interactions between formulation and articulation, the pattern with semantically related primes is opposite that predicted by long-distance interaction; the disruption that leads to overt speech errors should lead to distortions in articulation (certainly relative to the match condition).

Dynamic interaction. Neither match status ($\beta = -20.459$, χ^2 (1) = 1.75, $p = 0.186$) nor semantic relatedness interacted with reaction time ($\beta = 7.360$, χ^2 (1) = 0.19, $p = 0.667$), failing to provide support for the predictions of accounts with dynamic interaction between cognitive processes.

Inhibitory Deficit. The Inhibitory Deficit hypothesis predicts that older adults (with less ability to suppress competitors) should show larger interactive effects than young adults. We assessed this prediction using a regression model; this was structured similarly to the overall model of word durations, with the addition of fixed effects for experiment (Experiment 3 vs. Experiment 1) and interactions of experiment and all other effects (full results can be found in Appendix C). The interaction of experiment with the effect of matching primes was not

[4] There was a significant interaction of block and match status, such that overall reduction of word durations across blocks was strongest for mismatch trials.

significant[5] (β =-2.320, χ^2 (1) = 0.14, p = 0.704). However, the effect of semantic relatedness did interact with experiment (β = –10.309, χ^2 (1) = 4.41, p = 0.038); underscoring the unexpected finding above, there was a *larger* (more negative) effect was found with older adults. This fails to support an increased effect of lexical selection disruptions on processing.

Initial segment speech initiation. In contrast to the previous two experiments, there was no clear effect on initial consonants, nor any effect on vowels. The effect of match status was non-significant for initial consonants (β = 3.381, χ^2 (1) = 2.72, p = 0.099) and non-significant for vowel durations (β = 2.743, χ^2 (1) = 1.90, p = 0.168). Similarly, the effect of semantic relatedness was not significant for initial consonants (β = –2.486, χ^2 (1) = 2.39, p = 0.122) as well as vowel durations (β = –0.729, χ^2 (1) = 0.22, p = 0.636). Figure 4 provides a visualization of the relative effect sizes on the three duration measures. Spectral properties of the vowel also failed to show significant effects (match status: β = 1.295 , χ^2 (1) = 0.04, p = 0.843; semantic relatedness: β = -2.664, χ^2 (1) = 0.36, p = 0.547). There were no significant modulation of these effects by reaction time (see Appendix C for full models.).

There was also little evidence of dynamic long-distance interactions in the context of initial speech initiation effects. Experiment 3 RTs did not interact with semantic relatedness for initial consonants (β = –5.706, χ^2 (1) = 0.67, p = 0.413), vowel durations (β = –0.173, χ^2 (1) = 0, p = 0.975), or vowel spectral properties (β = –19.634, χ^2 (1) = 1.09, p = 0.297).

Finally, the cross-experiment models examining consonant and vowel articulation measures failed to find support for Inhibitory Deficits (see Appendix C for full results). There

[5] There was a significant three-way interaction of experiment, match status, and block; this reflected the non-significant interaction of match status and block in Experiment 1 vs. the significant interaction in the current experiment.

was, however, a significant effect of age; vowel distances were overall larger for older speakers ($\beta = 35.565$, $\chi^2(1) = 4.650$, $p = 0.031$).

Discussion

While older adults produced more errors (14.1% vs. 8.6% in experiment 1) and had overall longer word durations (consistent with slowed speech associated with aging), they did not show heightened sensitivity to differences across experimental conditions. Their rate of completion errors was numerically lower than that of younger speakers, and the majority of the effects on articulatory measures were weaker (e.g., lack of significant condition effects on initial consonants). We discuss the implications of this finding in the following section.

General Discussion

Previous research has yielded mixed results on the extent and scope of interactions between formulation and articulatory processes. This study aimed to provide novel evidence on these issues using a paradigm that disrupted formulation. Participants named a picture following a sentence priming a matching or non-matching picture label (Ferreira & Griffin, 2003). A novel automatic phonetic analysis tool allowed us to examine word durations; in concert with other analysis tools, we used this to automatically measure more fine-grained aspects of word articulation associated with specific segments. Consistent with a number of previous studies, when priming a non-target word disrupts formulation, articulation is disrupted as well. This effect appears to be stronger in the initial position of the words, suggesting that speech is initiated as soon as these segments are planned; subsequent segments benefit from additional planning, suppressing the temporary effects of disruption. Some of the mixed evidence in the literature may therefore reflect the use of coarse-grained measures of articulatory processing. However, two other possible confounds in previous work did not appear to influence the current results. Disruptions to formulation did not exert a stronger influence with explicit time pressure or faster reaction times, suggesting overlap between formulation and articulation is relatively fixed. Older participants showed, if anything, weaker interactive effects than younger adults, suggesting that aging does not decrease the ability to suppress non-target representations during speech planning.

A central goal of this work was to provide a clearer picture of long-distance interactive effects: specifically, the impact of disruptions to lexical selection (as opposed to other aspects of formulation) on articulatory processing. The priming paradigm successfully disrupted lexical selection; semantically related primes yielded overt, whole-word picture naming errors at a higher rate than unrelated primes. However, this enhanced disruption of lexical selection was associated with *weaker* disruptions to articulation than the unrelated condition.

Challenges raised by results from semantic primes

As discussed in the introduction, a large body of work has demonstrated that there are many situations where articulatory processing is facilitated when formulation processes are facilitated. Several theoretical proposals in the language production literature assume that conceptual processes preceding lexical selection are facilitated by semantic relationships (for reviews and discussion see Abdel & Melinger, 2007; Scaltritti, Peresotti, & Navarrete, 2017); under some contexts, this can facilitate subsequent processing. In the context of a task using a sentence prime, previous studies have measured the difficulty of reading a word following a high-cloze probability sentence; in such tasks, there is *less* disruption when the unexpected word is semantically related to the cloze completion than when it is semantically unrelated to the target (Federmeier & Kutas, 1999). Furthermore, results from a wide array of studies have suggested that older adults show larger semantic priming effects (e.g., Laver & Burke, 1993); here, older adults show larger facilitatory effects of semantic primes than younger adults. These results make it plausible that semantic primes could facilitate target retrieval, and as a consequence, articulatory processing. The challenge for such an account is to explain the error data. Whatever facilitation the semantic primes provide the target during formulation, it is clearly insufficient to suppress the occurrence of semantic errors. If articulatory effects arise as a result of cascading activation, why does the heightened activation of semantic competitors fail to disrupt articulatory processing?

One possibility (suggested by an anonymous reviewer) is that these effects represent a kind of speed/accuracy tradeoff; participants capitalize on the facilitatory effects of semantic primes to speed formulation and articulation at the cost of allowing more errors to be produced. Consistent with this, there is a trend towards facilitatory effects of semantic primes in RTs (see Appendices C and D for analyses and statistical models). However, a key prediction of this account is not confirmed by our data. Experiment 2 differs from Experiment 1 in that speed is explicitly emphasized. This results in a clear increase in errors and a decrease in RTs (illustrating a speed/accuracy tradeoff). However, it is not accompanied by a significant increase in semantic effects on articulatory processing. While the effect of semantic primes is significant in Experiment 2 but not in Experiment 1, regression models explicitly comparing the two experiments show no interaction of semantic primes with experiment. This possibility deserves further investigation; it is possible that our study did not have sufficient power to detect such interactions.

Another alternative is to allow for multiple processes or information sources to directly influence articulatory processing. Rather than articulation being modulated solely by the relative activation of target and competitor representations (i.e., the output of formulation processes), it may be that target activation has a privileged effect on articulation, irrespective of the activation of competitors. For example, Baker and Bradlow (2009) presented analyses showing that the facilitatory effects of contextual predictability cannot be reduced to the prosodic structure of an utterance (with predictable items occurring in less prominent positions). They therefore propose that predictability exerts a direct influence on articulation. Such a direct influence could serve to reduce articulatory disruptions for semantically related primes.

A natural question is then whether predictability provides a sufficient account of the data, obviating the need to appeal to interaction to account for the data reported here. Other results in the literature on interactive effects suggest a pure predictability account is insufficient (see Arnold, 2016, for a review). As reviewed in the introduction, Jacobs et al. (2015) shows that facilitation of articulation requires the prior facilitation of word form processing. Findings such as these suggest that predictability and interactive effects co-determine articulatory prominence.

Initiation of speech by initial segment

The enhanced effects on initial positions are consistent with previous findings from reading aloud (Kawamoto et al., 1999) and typing (Scaltritti et al., 2017; but see Damian & Freeman, 2008). Such results are predicted if planning continues following articulation onset. While the initial portions of the word are being articulated, planning may continue for later portions. This extra time may allow the production system to resolve planning conflicts before having to articulate, reducing effects at later portions of the word. Future research should be focused on examining more fine-grained measures of speech articulation to capture such transient effects.

Dynamic interaction

The failure to find effects of naturally occurring variation in response speed is consistent with the absence of such effects in Fink et al.'s (2018) study of semantic interference effects. The current study provides converging evidence from the absence of effects from explicit response pressure, consistent with Damian (2003) and Damian and Freeman (2008). This suggests that the coordination of formulation processes and articulation processes is not as flexible as suggested by Kello et al. (2000). Future work, perhaps including high-powered replications of Kello et al. (2000), might allow us to determine if the original results were spurious. The automated phonetic analysis developed here could facilitate such work.

Lack of inhibitory deficits in aging

While older adults produced longer word durations (suggesting a reduction in articulatory rate), and somewhat lower accuracy overall, aging did not enhance the effect of lexical selection disruptions on articulation. In fact, older adults showed enhanced *facilitatory* effects of semantic primes. This is consistent with previous work suggesting that various predictions of the Inhibitory Deficit hypothesis are not confirmed by empirical studies of cognitive aging (see Burke, 1997, for a review). The Transmission Deficit hypothesis (Burke et al., 1991; MacKay, 1987) might provide a framework for understanding such effects. According to this theory, connections between representational units are weakened with increasing age. Specifically, weak connections between phonological and lexical levels prevent form-level representations from becoming adequately activated, yielding difficulty in retrieval (MacKay, 1987; Taylor & Burke, 2002). Furthermore, because of the degraded links from semantics to phonology, the amount of activation that cascades to subsequent phonetic/articulatory processes should be significantly reduced. This would reduce the influence of weakly activated competitor representations, decreasing interactive effects.

The relationship between formulation and articulation

More generally, this set of findings suggests that interactions between formulation and articulation do not simply cause articulation to mirror all aspects of formulation processing. While disruptions to formulation can yield articulatory disruptions, the extent and nature of these disruptions is not a straightforward consequence of the outcome of formulation processes. Articulatory effects may be sensitive to multiple aspects of processing (e.g., lexical selection difficulties as well as positive effects of contextual predictability). A more nuanced model of planning-articulation interactions may better account for this complex relationship and yield predictions about contexts that facilitate vs. inhibit observing long-distance interactive effects.

References

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes*, *24*, 713-734.

Adi, Y., Keshet, J., Cibelli, E., & Goldrick, M. (2016). Sequence segmentation using joint RNN and structured prediction models. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2017* (pp. 2422-2426). Institute of Electrical and Electronics Engineers.

Adi, Y., Keshet, J., Cibelli, E., Gustafson E., Clopper C., & Goldrick, M. (2016). Automatic measurement of vowel duration via structured prediction. *Journal of the Acoustical Society of America, 140*, 4517-4527.

Adi, Y., Keshet, J. Dmitrieva, O., & Goldrick, M. (2016) Automatic measurement of voice onset time and prevoicing using recurrent neural networks. In *Proceedings of Interspeech 2016* (pp. 3152-3155). doi:10.21437/Interspeech.2016-893

Adi, Y., Keshet, J., & Goldrick, M. (2015, September). Vowel duration measurement using deep neural networks. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing.*

Arnold, J. E. (2016). Explicit and emergent mechanisms of information status. *Topics in Cognitive Science, 8(4)*, 737-760.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baker, R. E., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech, 52*, 391–413.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28*(1), 14–36.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68(3)*, 255-278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint :1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Belke, E. (2017). The role of task-specific response strategies in blocked-cyclic naming. *Frontiers in Psychology, 7*, 1955. doi: 10.3389/fpsyg.2016.01955

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, *58*, 667-692.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. J*ournal of Memory and Language, 60*, 92–111

Biegler, K. A., Crowther, J. E., & Martin, R. C. (2008). Consequences of an inhibition deficit for word production and comprehension: Evidence from the semantic blocking paradigm. *Cognitive Neuropsychology*, *25*, 493-527.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). 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*, *5*, e10773.

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

Burke, D. M. (1997). Language, aging, and inhibitory deficits: Evaluation of a theory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 52*(6), 254-264.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language, 30*, 542-579.

Choirat, C., Gandrud, C., Honaker, J., Imai, K., King, G., Lau, O. (2018). relogit: Rare events logistic regression for dichotomous dependent variables. In C. Choirat, C. Gandrud, J. Honaker, K. Imai, G. King, and O. Lau, *Zelig: Everyone's statistical software* http://zeligproject.org/

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318-344.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*, 416–431.

Damian, M. F., & Freeman, N. H. (2008). Flexible and inflexible response components: a Stroop study with typewritten output. *Acta Psychologica*, *128*, 91-101.

Damian, M. F., Vigliocco, G., & Levelt, W. J. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*, B77-B86.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287–314.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611–629.

Dhooge, E., De Baene, W., & Hartsuiker, R. J. (2013). A late locus of the distractor frequency effect in picture–word interference: Evidence from event-related potentials. *Brain and Language*, *124*, 232-237.

Dissen, Y., & Keshet, J. (2016). Formant estimation and tracking using deep learning. In *The 17th Annual Conference of the International Speech Communication Association*.

Drake, E., & Corley, M. (2015). Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level? *Memory & Cognition*, *43*, 111–120.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research,* 12, 2121–2159.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure, *Machine Learning*, 7, 195–225.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469-495.

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis)selection. *Psychological Science*, *14*, 86–90.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8, e77661*.

Fink, A. (2016). *The role of domain-general executive functions, articulation, and* *conceptualization in spoken word production*. Doctoral dissertation, Northwestern University.

Fink, A., Oppenheim, G.M., and Goldrick, M. (2018). Interactions between lexical access and articulation. *Language, Cognition, and Neuroscience, 33,* 12-24.

Finkbeiner, M., & Caramazza, A. (2006a). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*, 790-796.

Finkbeiner, M., & Caramazza, A. (2006b). Lexical selection is not a competitive process: A reply to La Heij et al. (2006). *Cortex*, *42*, 1032-1036.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*, 139-162.

Gahl, S. (2008). "Thyme" and "time" are not homophones. The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*(3), 474-496.

Gahl, S., & Strand, J. F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language, 89*, 162-178.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789-806

Garrett, M. F. (1975). The analysis of sentence production. *Psychology of Learning and Motivation*, *9*, 133-177.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21*, 817-855.

Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*, 58-72.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*, 649-683.

Goldrick, M., Keshet, J., Gustafson, E., Heller, J., & Needle, J. (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition*, *149*, 31-39.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*, 386-412.

Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology: General, 135*, 462-483.

Gollan, T. H., & Goldrick, M. (submitted). *Aging deficits in speech production revealed through reading aloud*. Manuscript submitted for publication.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism*, *15*, 594-615.

Graves A. & Jaitly, N. (2014). Towards end-to-end speech recognition with recurrent neural networks. In *ICML* (Vol. 14, pp. 1764-1772).

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on* (pp. 6645-6649). IEEE.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*, 1735-1780.

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language, 84*, 37–48.

Jescheniak, J. D., Hahne, A., Hoffmann, S., & Wagner, V. (2006). Phonological activation of category coordinates during speech planning is observable in children but not in adults: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 373-386.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language, 67*, 311–325.

Kahn, J. M., & Arnold, J. E. (2013). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *Language and Cognitive Processes, 30*, 103-119.

Kello, C. T. (2004). Control over the time course of cognition in the tempo-naming task. *Journal of Experimental Psychology. Human Perception and Performance*, *30*, 942–955.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: an empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology. General*, *129*, 340–360.

Keshet, J., Shalev-Shwartz, S., Singer, Y., & Chazan, D. (2007). A large margin algorithm for speech-to-phoneme and music-to-score alignment. *IEEE Transactions on Audio, Speech, and Language Processing*, *15*, 2373-2382.

King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*(2), 137-163.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: why repeated referents have reduced prominence. *Memory & Cognition, 38*, 1137-1146.

Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging, 8*, 34-43.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.

Li, L., & Slevc, L. R. (2017). Of papers and pens: Polysemes and homophones in lexical (mis)Selection. *Cognitive Science*, *41*, 1532–1548.

Logan, G. D., & Zbrodoff, N. J. (1998). Stroop-type interference: Congruity effects in color naming with typewritten responses.. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 978–992. doi:10.1037/0096- 1523.24.3.978

MacKay, D. G. (2012). *The organization of perception and action: A theory for language and other cognitive skills*. Berlin: Springer.

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A.S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology,* 8, 1164. doi: 10.3389/fpsyg.2017.01164

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS one*, *7*, e43230.

McAllester, D., Hazan, T., & Keshet, J. (2010). Direct loss minimization for structured prediction. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.) *Advances in neural information processing systems 23* (pp. 1594-1602).

McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*, 44-66.

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*, 243-260.

Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. *Language, Cognition and Neuroscience, 29*, 663-683.

Mortensen, L., Meyer, A. S., & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes*, *21*, 238-290.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. In J. Cole & J. I. Hualde (Eds.) *Papers in laboratory phonology 9* (pp. 201-228). Berlin: Mouton de Gruyter.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition, 114*, 227-252.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 539–557.

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*, *50*, 311-341.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of*

Phonetics, *36*, 114-140.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6*, 5–42.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460-499.

Roelofs, A. (2014). Integrating psycholinguistic and motor control approaches to speech production: where do they meet?. *Language, Cognition and Neuroscience*, *29*, 35-37.

Scaltritti, M., Arfé, B., Torrance, M., & Peressotti, F. (2016). Typing pictures: Linguistic processing cascades into finger movements. *Cognition, 156*, 16–29. doi:10.1016/j.cognition.2016.07.006

Scaltritti, M., Peressotti, F., & Navarrete, E. (2017). A joint investigation of semantic facilitation and semantic interference in continuous naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 818-823.

Scaltritti, M., Pinet, S., Longcamp, M., & Alario, F. X. (2017). On the functional relationship between language and motor processing in typewriting: An EEG study. *Language, Cognition and Neuroscience, 32*, 1086-1101. DOI: 10.1080/23273798.2017.1283427

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*, 86-102.

Severens, E., Ratinckx, E., Ferreira, V. S., & Hartsuiker, R. J. (2008). Are phonological influences on lexical (mis)selection the result of a monitoring bias? *The Quarterly Journal of Experimental Psychology*, *61*, 1687–1709.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*, 140–155.

Shipley, W.C., Gruber, C.P., Martin, T.A., & Klein, A.M. (2009). *Shipley-2 manual.* Los Angeles, CA: Western Psychological Services.

Strijkers, K., & Costa, A. (2016). On words and brains: linking psycholinguistics with neural dynamics in speech production. *Language, Cognition, and Neuroscience, 31(4)*, 524-535.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: naming in the picture-word interference task. *Psychology and Aging*, *17*, 662-676.

Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 787-806.

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, *107*, 592-597.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241-264). San Diego, CA, US: Academic Press.

Table 1

Example illustrating the design of the study.

Prime Condition	Cloze Sentence	Primed Response	Picture Target
Match	The fairy tale princess lived in a majestic…	"castle"	
Semantically Related (Mismatch)	Every Halloween, they turned their home into a haunted…	"house"	
Unrelated (Mismatch)	The joint connecting the thigh and shin is the…	"knee"	

Table 2

Key predictions and results across the three experiments.

Key Prediction	**Results**
Interaction of Formulation and Articulation Interactive theories of speech production predict that disruptions to formulation processes (lexical selection and/or phonological encoding) disrupt articulation via cascading activation.	**Confirmed** *Articulation is disrupted following mismatching vs. matching primes.*
Long-Distance Interaction Interactive theories that allow for long-distance interactions predict that disruptions to lexical selection will disrupt phonological encoding and, in turn, articulation processes.	**Not confirmed** *Semantically related primes, which yield overt speech errors during lexical selection, show no more disruption than unrelated primes, which lead to fewer errors.*
Dynamic Interaction Theories incorporating dynamic interaction predict that insufficient time for selection process will increase the overlap between formulation and articulation processes; this greater overlap will increase interactive effects relative to conditions which allow for greater processing time.	**Not confirmed** *Articulatory effects do not interact with trial-level reaction time, nor do they increase with response pressure.*
Inhibitory Deficit The Inhibitory Deficit Hypothesis for cognitive aging predicts that older adults will be less able to inhibit non-target representations; in the context of interactive theories of production, older adults are predicted to show stronger interactive effects than younger adults.	**Not confirmed** *Articulatory effects in older adults are of comparable magnitude to younger adults*
Initial Segment Speech Initiation If speech is initiated prior to completion of planning for remainder of the word, interactive theories predict that interactive effects will be stronger in the initial vs. later part of the word (where additional planning time mitigates the effects of disruptions).	**Confirmed** *Proportional effects of disruption are numerically larger on duration of initial consonants vs. duration of following vowels.*

Table 3

*Vocabulary measures for each experiment. * denotes score significantly different from Experiment 1 baseline.*

	Experiment 1 Young Adults No time pressure		*Experiment 2* Young Adults Time pressure		*Experiment 3* Older Adults No time pressure	
	Mean	s.e.	Mean	s.e.	Mean	s.e.
MINT (productive vocabulary)	64.556	0.623	62.222	0.712	64.500	0.526
Shipley-2 (receptive vocabulary)	113.444	1.190	118.278*	0.863	117.333	1.811

Table 4

Mean word durations with standard errors across participants for each condition, Experiment 1.

Condition	Mean word duration	s.e.
Match	370.838	(3.479)
Semantically Related	376.807	(3.762)
Unrelated	380.109	(3.801)

Table 5

Mean word durations with standard errors across participants for each condition, Experiment 2.

Condition	Mean word duration	s.e.
Match	324.491	(3.295)
Semantically Related	323.857	(3.663)
Unrelated	331.083	(3.532)

Table 6

Mean word durations with standard errors across participants for each condition, Experiment 3.

Condition	Mean word duration	s.e.
Match	393.569	(3.815)
Semantically Related	397.651	(4.105)
Unrelated	421.518	(4.501)

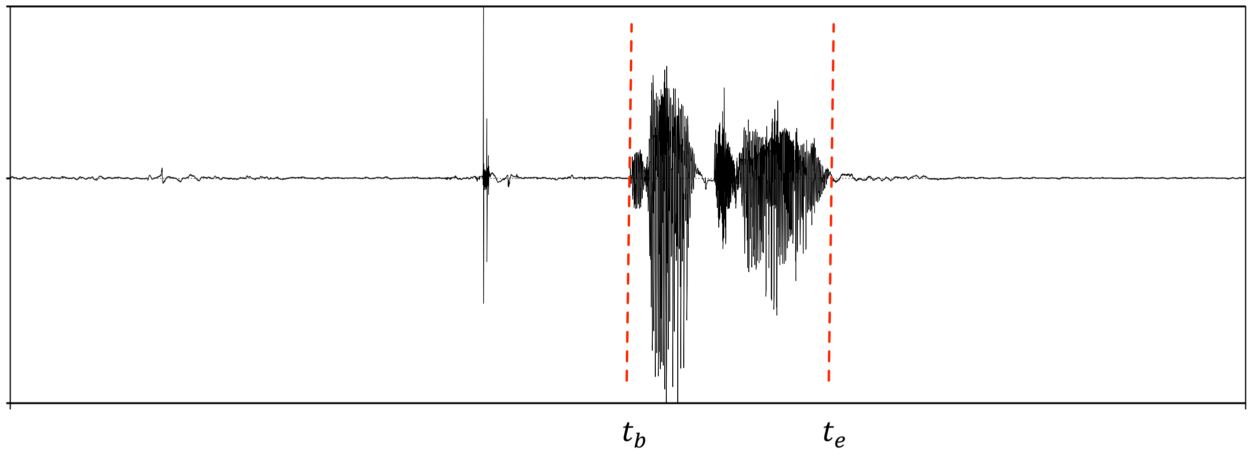

Figure 1. Example acoustic signal with annotations marking onset t_b and offset t_e of speech. Note prior to speech onset there is a high intensity non-speech signal (lip smack) that the DeepWDM algorithm can learn to ignore.

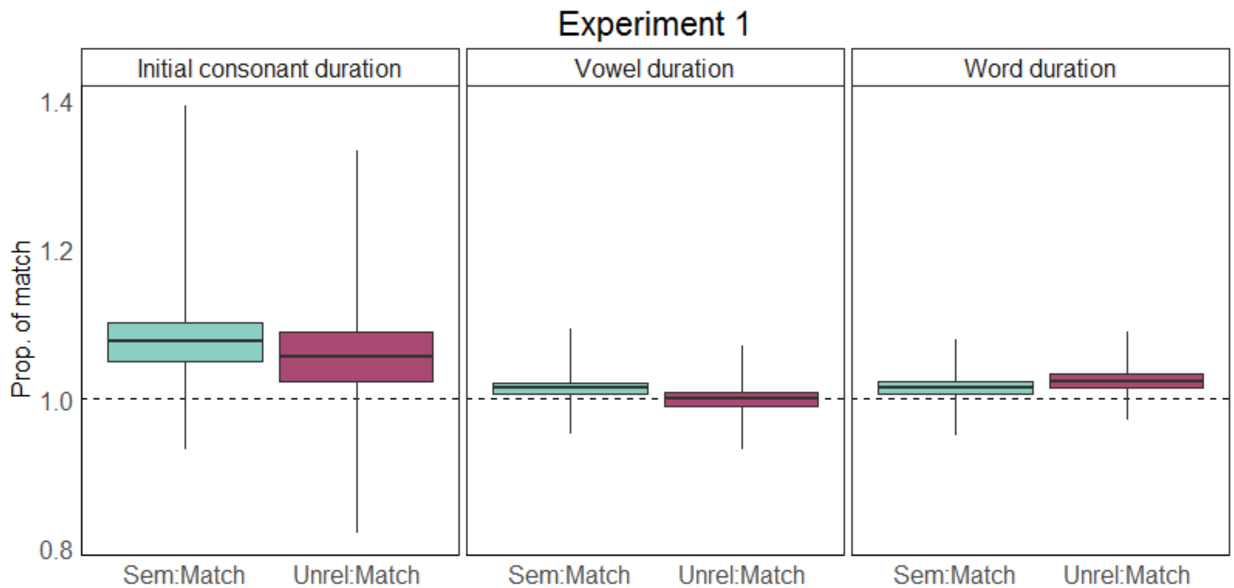

Figure 2. Condition means of each dependent variable for the no-match conditions (sem = semantically related competitor, unrel = unrelated) as a proportion of match condition. An average value within each condition was generated for each participant. The boxplots show the mean (central line) and standard error across participants (width of box). Wings show range. A value of 1 indicates that the match condition and no-match condition had the same average measurement; values above 1 indicate that the no-match condition had a longer or larger measurement than the match condition.

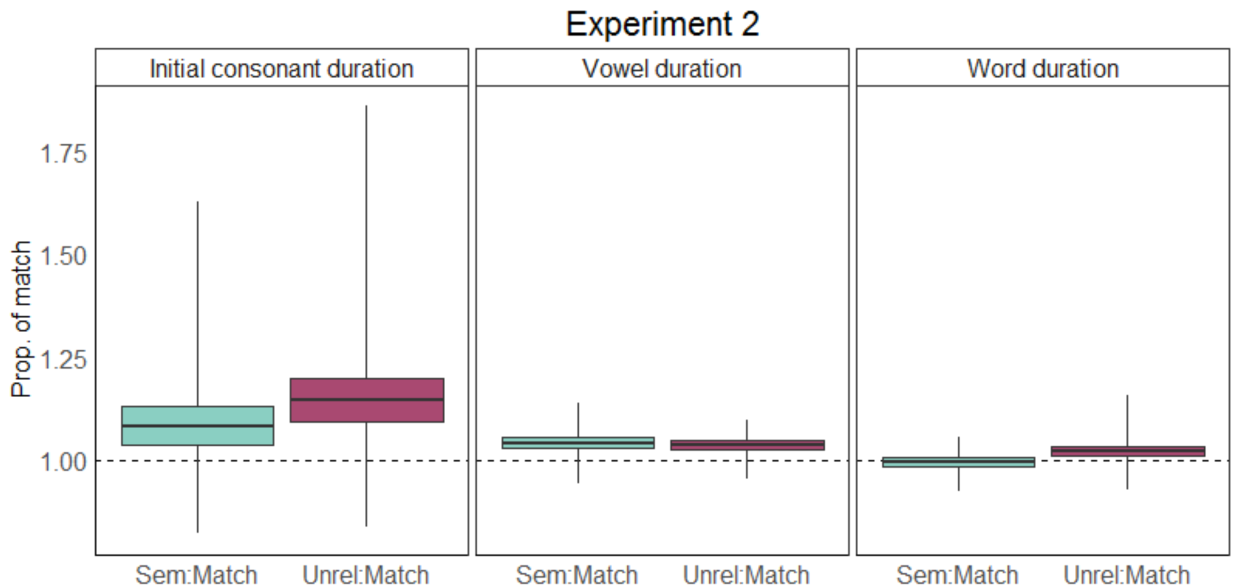

Figure 3. Condition means of each dependent variable for the no-match conditions (sem = semantically related competitor, unrel = unrelated) as a proportion of match condition. An average value within each condition was generated for each participant. The boxplots show the mean (central line) and standard error across participants (width of box). Wings show range. A value of 1 indicates that the match condition and no-match condition had the same average measurement; values above 1 indicate that the no-match condition had a longer or larger measurement than the match condition.

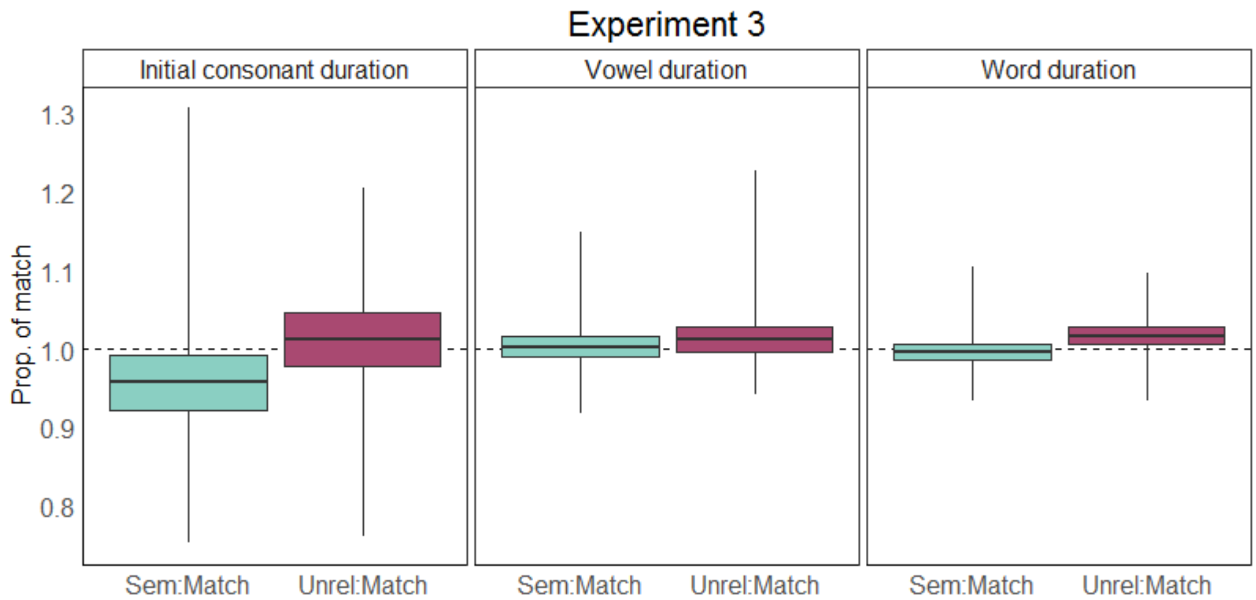

Figure 4. Condition means of each dependent variable for the no-match conditions (sem = semantically related competitor, unrel = unrelated) as a proportion of match condition. An average value within each condition was generated for each participant. The boxplots show the mean (central line) and standard error across participants (width of box). Wings show range. A value of 1 indicates that the match condition and no-match condition had the same average measurement; values above 1 indicate that the no-match condition had a longer or larger measurement than the match condition.

Appendix A: Stimulus norming procedures

Separate groups of participants were recruited to norm the picture and sentence materials. To ensure high name agreement among the selected pictures, a large pool of images were first normed for name agreement using Mechanical Turk (mTurk). mTurk workers (n=20) were native English speakers between the ages of 19 and 60. During norming, participants were shown a picture one at a time and asked to provide a single word that best described the picture. Another group of mTurk workers normed sentences for cloze probability. mTurk workers (n=19) were native English speakers between the age of 19 and 60. During sentence norming, participants were shown pictures and asked to provide a single word that was the best completion for the sentence. Sentence norming data were used in a similar manner to the data trimming procedure described by Li and Slevc (2017). Cloze norms were inspected prior to the analysis of data from the main task to identify items which over 45% of participants responded with an incorrect response. In the current study, we attempted to exclude incorrect responses in which the participant provided the competitor as a completion to the sentence. These responses would introduce a confound. Our inspection revealed that no items met these criteria and no sentences were removed.

Appendix B: Items

The table below presents items used in the experiment. Items consist of a picture and a sentence context. There are 180 unique pictures across the items; each picture is part of a semantically-related pair (e.g. “anchor” and “rope”). These items were split into six lists of 180 items, with a subset of 60 unique pictures in each list. Pictures were evenly distributed across these lists, and sentence contexts were not repeated within a list.

Within a list, each picture was repeated three times, once in each condition. Match and semantic competitor sentence contexts for each picture were designed to prime that specific picture or its semantic pair, respectively. Unrelated contexts within a list were selected from match sentences for pictures not presented in that list. Across lists, a picture could be paired with one of two unrelated sentence contexts; in the table below, only one of the two appears.

Each participant saw one list of 180 items split across three blocks, with each picture occurring once per block. The order of conditions across blocks was counterbalanced. The assignment of lists was distributed such that picture/sentence pairings were evenly split across participants.

Table A1: List of items.

Picture	Condition	Sentence	Completion	Cloze (younger)	Cloze (older)
anchor	match	The sailor stopped the ship and dropped the ...	anchor	0.842	0.667
anchor	semantic	They tied the ship to the dock with an ...	rope	0.842	0.667
anchor	unrelated	Her favorite dish is macaroni and ...	cheese	0.842	0.667
ant	match	The colony worked together to build a hill for the queen ...	ant	0.789	0.917
ant	semantic	Honey is produced by an insect called a ...	bee	0.789	0.917
ant	unrelated	Her favorite dish is macaroni and ...	cheese	0.789	0.917
apple	match	Snow White was poisoned when she bit into an ...	apple	1.000	1.000
apple	semantic	Her favorite dish is macaroni and ...	cheese	1.000	1.000
apple	unrelated	At the ballpark, the boys enjoyed a bag of salty ...	peanuts	1.000	1.000
axe	match	He chopped down the tall pine with an ...	axe	1.000	0.583
axe	semantic	He put the bookshelf together not with a screwdriver, but with a ...	hammer	1.000	0.583
axe	unrelated	The colony worked together to build a hill for the queen ...	ant	1.000	0.583
baby	match	The parents bought a stroller for their newborn ...	baby	0.947	0.833
baby	semantic	The newborn sleeps peacefully in her ...	crib	0.947	0.833
baby	unrelated	He kicked the ball with his left ...	foot	0.947	0.833
backpack	match	The student took her books home in her ...	backpack	0.632	0.750
backpack	semantic	After home room, she dropped some books off in her ...	locker	0.632	0.750
backpack	unrelated	The salesman was as sly as a ...	fox	0.632	0.750
bagel	match	Angela likes cream cheese on a ...	bagel	0.947	0.917
bagel	semantic	John spread butter and grape jelly on his morning ...	toast	0.947	0.917
bagel	unrelated	Sandy watered her garden using a rubber ...	hose	0.947	0.917
balloon	match	The clown blew up a big green ...	balloon	1.000	1.000
balloon	semantic	It was windy enough to fly a ...	kite	1.000	1.000
balloon	unrelated	The man happily sat down in the comfortable ...	chair	1.000	1.000
basket	match	He collected Easter eggs in his ...	basket	1.000	1.000
basket	semantic	The shipment arrived in a large cardboard ...	box	1.000	1.000
basket	unrelated	When I eat in Maine, I always order a big red ...	lobster	1.000	1.000
bat	match	The blind, flying rodent that lives in caves is the ...	bat	0.895	0.833
bat	semantic	Matt grabbed the swatter to kill the ...	fly	0.895	0.833

bat	unrelated	The stagecoach wasn't moving because of the broken wagon ...	wheel	0.895	0.833
beach	match	Mary went to Hawaii to tan on a sandy ...	beach	1.000	1.000
beach	semantic	She likes to swim laps at the ...	pool	1.000	1.000
beach	unrelated	At the petting zoo Suzie's snack was stolen by a pesky billy ...	goat	1.000	1.000
beans	match	Mexican food comes with a side of rice and ...	beans	0.895	1.000
beans	semantic	Caesar salad is made with romaine ...	lettuce	0.895	1.000
beans	unrelated	The message was broadcast to the students over a loud ...	speaker	0.895	1.000
bear	match	The campers were frightened by a large grizzly ...	bear	0.947	0.750
bear	semantic	The nocturnal animal that looks like a masked robber is a ...	raccoon	0.947	0.750
bear	unrelated	The man brushed his hair using a fine-toothed ...	comb	0.947	0.750
beaver	match	The rodent famous for building dams in rivers is called a ...	beaver	1.000	0.917
beaver	semantic	He laid a trap with cheese to catch the ...	mouse	1.000	0.917
beaver	unrelated	Others won't hear your music when you listen to it wearing ...	headphones	1.000	0.917
bed	match	Bob sleeps in a king-sized ...	bed	1.000	0.580
bed	semantic	He lies his head down to sleep on his ...	pillow	1.000	0.580
bed	unrelated	To garnish the margarita, the bartender sliced a wedge from a sour green ...	lime	1.000	0.580
bee	match	Honey is produced by an insect called a ...	bee	1.000	0.250
bee	semantic	The colony worked together to build a hill for the queen ...	ant	1.000	0.250
bee	unrelated	A Vespa is a type of motorized ...	scooter	1.000	0.250
bench	match	The homeless man slept on a park ...	bench	1.000	0.250
bench	semantic	The family sat in their living room on a big comfy ...	couch	1.000	0.250
bench	unrelated	His mouth puckered when he ate the sour yellow ...	lemon	1.000	0.250
boat	match	He grabbed the oars and got in the row ...	boat	0.842	0.417
boat	semantic	They paddled down the river in a wooden ...	canoe	0.842	0.417
boat	unrelated	She needed new laces for just one ...	shoe	0.842	0.417
bomb	match	The explosion came from a homemade pipe ...	bomb	0.895	0.947
bomb	semantic	The soldiers were protected by the armored fighting vehicle called a ...	tank	0.895	0.947
bomb	unrelated	Alice ended up in Wonderland when she followed the ...	rabbit	0.895	0.947
boot	match	Her foot was cold even though she wore a sock and a winter ...	boot	0.684	0.833
boot	semantic	Each year his grandmother knit him an ugly Christmas ...	sweater	0.684	0.833
boot	unrelated	For his birthday, his mom baked him a chocolate ...	cake	0.684	0.833
box	match	The shipment arrived in a large cardboard ...	box	1.000	1.000
box	semantic	He collected Easter eggs in his ...	basket	1.000	1.000
box	unrelated	She felt hot in the office and plugged in the ...	fan	1.000	1.000
bra	match	Under their shirts, most women wear a supportive ...	bra	1.000	0.500
bra	semantic	Between his jacket and shirt, the usher had buttoned his ...	vest	1.000	0.500
bra	unrelated	The chef chopped the vegetables with a ...	knife	1.000	0.500
broccoli	match	You can eat the green stalk and flowering head of ...	broccoli	0.526	1.000
broccoli	semantic	Bugs Bunny chewed on a ...	carrot	0.526	1.000
broccoli	unrelated	Because it was out of fluid, there was no flame but only sparks from the ...	lighter	0.526	1.000
broom	match	The boy swept up his mess with the ...	broom	0.947	0.833
broom	semantic	Mary cleaned the floor using a bucket and ...	mop	0.947	0.833

broom	unrelated	The activists scolded the poacher for clubbing a baby ...	seal	0.947	0.833
cactus	match	In the desert, he pricked his finger on the spine of a ...	cactus	1.000	1.000
cactus	semantic	The bird built a nest high up in an elm ...	tree	1.000	1.000
cactus	unrelated	The joint connecting the thigh and shin is the ...	knee	1.000	1.000
cake	match	For his birthday, his mom baked him a chocolate ...	cake	1.000	0.917
cake	semantic	My favorite treat is a chocolate chip ...	cookie	1.000	0.917
cake	unrelated	She needed new laces for just one ...	shoe	1.000	0.917
candle	match	When the power went out, they lit a ...	candle	0.842	1.000
candle	semantic	She placed the flowers in a glass ...	vase	0.842	1.000
candle	unrelated	The shipment arrived in a large cardboard ...	box	0.842	1.000
candy	match	On Halloween kids trick-or-treat to collect ...	candy	0.947	0.750
candy	semantic	At the ballpark, the boys enjoyed a bag of salty ...	peanuts	0.947	0.750
candy	unrelated	The sailor stopped the ship and dropped the ...	anchor	0.947	0.750
canoe	match	They paddled down the river in a wooden ...	canoe	0.737	1.000
canoe	semantic	He grabbed the oars and got in the row ...	boat	0.737	1.000
canoe	unrelated	The footwear typically worn in the summer is a ...	sandal	0.737	1.000
carrot	match	Bugs Bunny chewed on a ...	carrot	1.000	0.500
carrot	semantic	You can eat the green stalk and flowering head of ...	broccoli	1.000	0.500
carrot	unrelated	The shipment arrived in a large cardboard ...	box	1.000	0.500
castle	match	The fairy tale princess lived in a majestic ...	castle	1.000	0.917
castle	semantic	Every Halloween, they turned their home into a haunted ...	house	1.000	0.917
castle	unrelated	Chocolate glazed with sprinkles is his favorite kind of ...	donut	1.000	0.917
cat	match	Susan hates dogs, but loves Garfield, her fat tabby ...	cat	1.000	1.000
cat	semantic	The animal bacon and ham come from is the ...	pig	1.000	1.000
cat	unrelated	He chopped down the tall pine with an ...	axe	1.000	1.000
chair	match	The man happily sat down in the comfortable ...	chair	0.895	0.833
chair	semantic	I like to sit at the bar on a tall wooden ...	stool	0.895	0.833
chair	unrelated	The clown blew up a big green ...	balloon	0.895	0.833
cheese	match	Her favorite dish is macaroni and ...	cheese	1.000	0.917
cheese	semantic	Snow White was poisoned when she bit into an ...	apple	1.000	0.917
cheese	unrelated	During their bike tour Mike had to fix a flat ...	tire	1.000	0.917
cheetah	match	The wild cat that runs the fastest is the ...	cheetah	0.895	1.000
cheetah	semantic	The king of the jungle is the ...	lion	0.895	1.000
cheetah	unrelated	Levi makes high quality denim blue ...	jeans	0.895	1.000
chips	match	At the party, we had some salsa and tortilla ...	chips	0.947	1.000
chips	semantic	Before the movie started everyone bought some buttery ...	popcorn	0.947	1.000
chips	unrelated	Nick sneezed and blew his ...	nose	0.947	1.000
clock	match	He keeps track of time by looking to the wall at the ...	clock	0.947	0.833
clock	semantic	The ornate shade covered the bulb of the ...	lamp	0.947	0.833
clock	unrelated	The smallest bank note is one ...	dollar	0.947	0.833
comb	match	The man brushed his hair using a fine-toothed ...	comb	0.947	0.417
comb	semantic	Proper dental hygiene includes cleaning each day with a bristled ...	toothbrush	0.947	0.417
comb	unrelated	The campers were frightened by a large grizzly ...	bear	0.947	0.417
compass	match	The navigation device that points north is a ...	compass	0.947	0.917
compass	semantic	The lead broke so she sharpened the ...	pencil	0.947	0.917
compass	unrelated	She placed the flowers in a glass ...	vase	0.947	0.917
cookie	match	My favorite treat is a chocolate chip ...	cookie	0.947	0.917
cookie	semantic	For his birthday, his mom baked him a chocolate ...	cake	0.947	0.917
cookie	unrelated	They paddled down the river in a wooden ...	canoe	0.947	0.917

corn	match	The vegetable that comes on a cob is ...	corn	0.947	0.917
corn	semantic	The vegetables that come in pods are ...	peas	0.947	0.917
corn	unrelated	The bike was protected from theft by an expensive ...	lock	0.947	0.917
couch	match	The family sat in their living room on a big comfy ...	couch	0.579	1.000
couch	semantic	The homeless man slept on a park ...	bench	0.579	1.000
couch	unrelated	To garnish the margarita, the bartender sliced a wedge from a sour green ...	lime	0.579	1.000
cow	match	Every day the farmer goes to milk his only ...	cow	1.000	0.917
cow	semantic	The farmer woke up to the cock-a-doodle-doo of the ...	rooster	1.000	0.917
cow	unrelated	For his interview Mr. Jones needed a new ...	suit	1.000	0.917
crib	match	The newborn sleeps peacefully in her ...	crib	0.895	0.833
crib	semantic	The parents bought a stroller for their newborn ...	baby	0.895	0.833
crib	unrelated	He wished the actor luck by saying break a ...	leg	0.895	0.833
desk	match	During work Danny sits all day long at his ...	desk	0.842	0.947
desk	semantic	For dinner, the family gathers at the dining room ...	table	0.842	0.947
desk	unrelated	She likes to swim laps at the ...	pool	0.842	0.947
dollar	match	The smallest bank note is one ...	dollar	0.895	1.000
dollar	semantic	A one cent coin is called a ...	penny	0.895	1.000
dollar	unrelated	He keeps track of time by looking to the wall at the ...	clock	0.895	1.000
dolphin	match	The fisherman came upon a pod with a baby bottlenose ...	dolphin	0.684	0.750
dolphin	semantic	The activists scolded the poacher for clubbing a baby ...	seal	0.684	0.750
dolphin	unrelated	Mary cleaned the floor using a bucket and ...	mop	0.684	0.750
donut	match	Chocolate glazed with sprinkles is his favorite kind of ...	donut	0.684	0.917
donut	semantic	At brunch, Maggie either eats pancakes or a Belgian ...	waffle	0.684	0.917
donut	unrelated	She likes to swim laps at the ...	pool	0.684	0.917
dropper	match	He applied the medicine to his eye using a ...	dropper	0.632	0.833
dropper	semantic	Andy removed a splinter with some ...	tweezers	0.632	0.833
dropper	unrelated	The Canadian flag features a maple ...	leaf	0.632	0.833
eagle	match	The national bird of the United States is the ...	eagle	0.947	0.917
eagle	semantic	Every ugly duckling eventually becomes a beautiful ...	swan	0.947	0.917
eagle	unrelated	He melted cheese over tortilla chips to make ...	nachos	0.947	0.917
elbow	match	The joint connecting the forearm to the bicep is the ...	elbow	0.947	0.833
elbow	semantic	The joint connecting the thigh and shin is the ...	knee	0.947	0.833
elbow	unrelated	He collected Easter eggs in his ...	basket	0.947	0.833
fan	match	She felt hot in the office and plugged in the ...	fan	0.947	1.000
fan	semantic	To let some cool air in the apartment they opened a ...	window	0.947	1.000
fan	unrelated	He chopped down the tall pine with an ...	axe	0.947	1.000
feather	match	The sense of relief made him feel as light as a ...	feather	1.000	1.000
feather	semantic	She went to the salon to color her ...	hair	1.000	1.000
feather	unrelated	The boy took a pole to the lake to catch a ...	fish	1.000	1.000
fish	match	The boy took a pole to the lake to catch a ...	fish	1.000	0.917
fish	semantic	When I eat in Maine, I always order a big red ...	lobster	1.000	0.917
fish	unrelated	To fix his torn paper he needs some ...	tape	1.000	0.917
floss	match	After brushing his teeth, Mike also uses dental ...	floss	0.947	0.833
floss	semantic	Every morning the man shaves using a ...	razor	0.947	0.833
floss	unrelated	The animal with a long neck and long legs is a ...	giraffe	0.947	0.833
fly	match	Matt grabbed the swatter to kill the ...	fly	1.000	1.000

fly	semantic	The blind, ..ing rodent that lives in caves is the ...	bat	1.000	1.000
fly	unrelated	On Halloween kids trick-or-treat to collect ...	candy	1.000	1.000
foot	match	He kicked the ball with his left ...	foot	0.947	1.000
foot	semantic	He wished the actor luck by saying break a ...	leg	0.947	1.000
foot	unrelated	He found a pot of gold at the end of the ...	rainbow	0.947	1.000
fox	match	The salesman was as sly as a ...	fox	1.000	0.583
fox	semantic	Alice ended up in Wonderland when she followed the ...	rabbit	1.000	0.583
fox	unrelated	The student took her books home in her ...	backpack	1.000	0.583
giraffe	match	The animal with a long neck and long legs is a ...	giraffe	1.000	1.000
giraffe	semantic	The horse-like animal with black and white stripes is the ...	zebra	1.000	1.000
giraffe	unrelated	The newborn sleeps peacefully in her ...	crib	1.000	1.000
glasses	match	She is as blind as a bat without her ...	glasses	0.947	1.000
glasses	semantic	Swimmers protect their eyes by wearing ...	goggles	0.947	1.000
glasses	unrelated	You can eat the green stalk and flowering head of ...	broccoli	0.947	1.000
globe	match	The spherical object showing the entire world is a ...	globe	0.947	1.000
globe	semantic	The directions did not match any roads on the ...	map	0.947	1.000
globe	unrelated	In the desert, he pricked his finger on the spine of a ...	cactus	0.947	1.000
glue	match	Emily fixed the broken mug with some ...	glue	0.947	0.750
glue	semantic	She cut the paper using ...	scissors	0.947	0.750
glue	unrelated	Peter serves the soup out of the pot with a ...	ladle	0.947	0.750
goat	match	at the petting zoo Suzie's snack was stolen by a pesky billy ...	goat	0.895	0.917
goat	semantic	The farmer shaved the wool off of the ...	sheep	0.895	0.917
goat	unrelated	Mary went to Hawaii to tan on a sandy ...	beach	0.895	0.917
goggles	match	Swimmers protect their eyes by wearing ...	goggles	1.000	1.000
goggles	semantic	She is as blind as a bat without her ...	glasses	1.000	1.000
goggles	unrelated	Bugs Bunny chewed on ...	carrot	1.000	1.000
goose	match	The children loved to play Duck-Duck ...	goose	0.947	0.833
goose	semantic	The bird that looks like it's wearing a tuxedo is a ...	penguin	0.947	0.833
goose	unrelated	after brushing his teeth, Mike also uses dental ...	floss	0.947	0.833
hair	match	She went to the salon to color her ...	hair	1.000	0.833
hair	semantic	The sense of relief made him feel as light as a ...	feather	1.000	0.833
hair	unrelated	Sam measured the length of the paper using a ...	ruler	1.000	0.833
hammer	match	He put the bookshelf together not with a screwdriver, but with a ...	hammer	0.684	0.917
hammer	semantic	He chopped down the tall pine with an ...	axe	0.684	0.917
hammer	unrelated	Honey is produced by an insect called a ...	bee	0.684	0.917
hat	match	To protect his head from sunburn, the bald man wore a wide-brimmed ...	hat	1.000	1.000
hat	semantic	A+B224fter doing laundry Derek noticed he was missing just one ...	sock	1.000	1.000
hat	unrelated	Andy removed a splinter with some ...	tweezers	1.000	1.000
headphones	match	Others won’t hear your music when you listen to it wearing ...	headphones	0.947	0.750
headphones	semantic	The message was broadcast to the students over a loud ...	speaker	0.947	0.750
headphones	unrelated	The rodent famous for building dams in rivers is called a ...	beaver	0.947	0.750
hose	match	Sandy watered her garden using a rubber ...	hose	0.895	1.000
hose	semantic	The gardener collects the leaves in a pile using a ...	rake	0.895	1.000
hose	unrelated	Angela likes cream cheese on a ...	bagel	0.895	1.000
house	match	Every Halloween, they turned their home into a haunted ...	house	1.000	0.917

house	semantic	The fairy tale princess lived in a majestic ...	castle	1.000	0.917
house	unrelated	She wrote the list on a piece of ...	paper	1.000	0.917
ipod	match	Apple's mp3 player is called an ...	ipod	0.895	0.833
ipod	semantic	Before the CD existed, music was listened to off of a black vinyl ...	record	0.895	0.833
ipod	unrelated	The old girlfriends chatted over a bottle of red ...	wine	0.895	0.833
jeans	match	Levi makes high quality denim blue ...	jeans	1.000	0.917
jeans	semantic	He wore a suit with a Windsor knot in his ...	tie	1.000	0.917
jeans	unrelated	The wild cat that runs the fastest is the ...	cheetah	1.000	0.917
jeep	match	Many people like the Grand Cherokee, but the Wrangler is my favorite kind of ...	jeep	0.684	0.894
jeep	semantic	We couldn't get a truck, but we managed to pack everything in a moving ...	van	0.684	0.894
jeep	unrelated	Before the movie started everyone bought some buttery ...	popcorn	0.684	0.894
juice	match	With breakfast Julie always drinks some orange ...	juice	1.000	1.000
juice	semantic	The old girlfriends chatted over a bottle of red ...	wine	1.000	1.000
juice	unrelated	During work Danny sits all day long at his ...	desk	1.000	1.000
key	match	To start a car, you need the ...	key	0.947	0.833
key	semantic	The bike was protected from theft by an expensive ...	lock	0.947	0.833
key	unrelated	In the desert, he pricked his finger on the spine of a ...	cactus	0.947	0.833
kite	match	It was windy enough to fly a ...	kite	1.000	0.917
kite	semantic	The clown blew up a big green ...	balloon	1.000	0.917
kite	unrelated	The animal with antlers that is much larger than a deer is a ...	moose	1.000	0.917
knee	match	The joint connecting the thigh and shin is the ...	knee	0.789	0.917
knee	semantic	The joint connecting the forearm to the bicep is the ...	elbow	0.789	0.917
knee	unrelated	The fairy tale princess lived in a majestic ...	castle	0.789	0.917
knife	match	The chef chopped the vegetables with a ...	knife	1.000	0.833
knife	semantic	She heated the stew in a large metal ...	pot	1.000	0.833
knife	unrelated	Under their shirts, most women wear a supportive ...	bra	1.000	0.833
ladle	match	Peter serves the soup out of the pot with a ...	ladle	0.789	0.583
ladle	semantic	Cooks remove skins from vegetables using a ...	peeler	0.789	0.583
ladle	unrelated	He garnished the martini with a green ...	olive	0.789	0.583
lamp	match	The ornate shade covered the bulb of the ...	lamp	0.737	1.000
lamp	semantic	He keeps track of time by looking to the wall at the ...	clock	0.737	1.000
lamp	unrelated	A portable computer is called a ...	laptop	0.737	1.000
laptop	match	A portable computer is called a ...	laptop	1.000	0.750
laptop	semantic	He hooked up his computer and discovered the ink had run dry in the ...	printer	1.000	0.750
laptop	unrelated	The ornate shade covered the bulb of the ...	lamp	1.000	0.750
leaf	match	The Canadian flag features a maple ...	leaf	1.000	1.000
leaf	semantic	This frozen turkey is as hard as a ...	rock	1.000	1.000
leaf	unrelated	He applied the medicine to his eye using a ...	dropper	1.000	1.000
leg	match	He wished the actor luck by saying break a ...	leg	1.000	1.000
leg	semantic	He kicked the ball with his left ...	foot	1.000	1.000
leg	unrelated	The blind, flying rodent that lives in caves is the ...	bat	1.000	1.000
lemon	match	His mouth puckered when he ate the sour yellow ...	lemon	0.737	0.830
lemon	semantic	To garnish the margarita, the bartender sliced a wedge from a sour green ...	lime	0.737	0.830
lemon	unrelated	The homeless man slept on a park ...	bench	0.737	0.830
lettuce	match	Caesar salad is made with romaine ...	lettuce	1.000	0.917
lettuce	semantic	Mexican food comes with a side of rice and ...	beans	1.000	0.917

lettuce	unrelated	The clown blew up a big green ...	balloon	1.000	0.917
lighter	match	Because it was out of fluid, there was no flame but only sparks from the ...	lighter	0.526	0.940
lighter	semantic	She opened the fireplace and lit the kindling with a wooden ...	match	0.526	0.940
lighter	unrelated	You can eat the green stalk and flowering head of ...	broccoli	0.526	0.940
lightning	match	Kids are often frightened by thunder and ...	lightning	0.947	0.917
lightning	semantic	He found a pot of gold at the end of the ...	rainbow	0.947	0.917
lightning	unrelated	The newborn sleeps peacefully in her ...	crib	0.947	0.917
lime	match	To garnish the margarita, the bartender sliced a wedge from a sour green ...	lime	0.947	0.833
lime	semantic	His mouth puckered when he ate the sour yellow ...	lemon	0.947	0.833
lime	unrelated	Bob sleeps in a king-sized ...	bed	0.947	0.833
lion	match	The king of the jungle is the ...	lion	0.737	0.417
lion	semantic	The wild cat that runs the fastest is the ...	cheetah	0.737	0.417
lion	unrelated	He dried his wet hands with a ...	towel	0.737	0.417
lips	match	In the winter, she uses lots of chapstick on her ...	lips	1.000	1.000
lips	semantic	Nick sneezed and blew his ...	nose	1.000	1.000
lips	unrelated	Many people like the Grand Cherokee, but the Wrangler is my favorite kind of ...	jeep	1.000	1.000
lobster	match	When I eat in Maine, I always order a big red ...	lobster	0.789	1.000
lobster	semantic	The boy took a pole to the lake to catch a ...	fish	0.789	1.000
lobster	unrelated	In the Olympic games, she won a gold ...	medal	0.789	1.000
lock	match	The bike was protected from theft by an expensive ...	lock	0.737	0.833
lock	semantic	To start a car, you need the ...	key	0.737	0.833
lock	unrelated	The vegetable that comes on a cob is ...	corn	0.737	0.833
locker	match	After home room, she dropped some books off in her ...	locker	0.947	0.667
locker	semantic	The student took her books home in her ...	backpack	0.947	0.667
locker	unrelated	The explosion came from a homemade pipe ...	bomb	0.947	0.667
map	match	The directions did not match any roads on the ...	map	0.947	0.833
map	semantic	The spherical object showing the entire world is a ...	globe	0.947	0.833
map	unrelated	Every Halloween, they turned their home into a haunted ...	house	0.947	0.833
match	match	She opened the fireplace and lit the kindling with a wooden ...	match	0.947	0.667
match	semantic	Because it was out of fluid, there was no flame but only sparks from the ...	lighter	0.947	0.667
match	unrelated	Bugs Bunny chewed on a ...	carrot	0.947	0.667
medal	match	In the Olympic games, she won a gold ...	medal	1.000	0.667
medal	semantic	The team that won the tournament took home a ...	trophy	1.000	0.667
medal	unrelated	After brushing his teeth, Mike also uses dental ...	floss	1.000	0.667
moon	match	Neil Armstrong was the first man to walk on the ...	moon	1.000	1.000
moon	semantic	The hopeful girl wished upon a ...	star	1.000	1.000
moon	unrelated	Every ugly duckling eventually becomes a beautiful ...	swan	1.000	1.000
moose	match	The animal with antlers that is much larger than a deer is a ...	moose	0.579	1.000
moose	semantic	In the desert, he got bitten by a rattle ...	snake	0.579	1.000
moose	unrelated	Raymond needed a belt to hold up his ...	pants	0.579	1.000
mop	match	Mary cleaned the floor using a bucket and ...	mop	0.947	0.917
mop	semantic	The boy swept up his mess with the ...	broom	0.947	0.917
mop	unrelated	Neil Armstrong was the first man to walk on the ...	moon	0.947	0.917
mouse	match	He laid a trap with cheese to catch the ...	mouse	0.947	1.000
mouse	semantic	The rodent famous for building dams in rivers is called a ...	beaver	0.947	1.000

mouse	unrelated	Mexican food comes with a side of rice and ...	beans	0.947	1.000
nachos	match	He melted cheese over tortilla chips to make ...	nachos	1.000	1.000
nachos	semantic	She snacks on a peanut butter and jelly ...	sandwich	1.000	1.000
nachos	unrelated	The national bird of the United States is the ...	eagle	1.000	1.000
nose	match	Nick sneezed and blew his ...	nose	1.000	1.000
nose	semantic	In the winter, she uses lots of chapstick on her ...	lips	1.000	1.000
nose	unrelated	The animal bacon and ham come from is the ...	pig	1.000	1.000
notebook	match	A binder of ruled pages used by students is a …	notebook	0.842	1.000
notebook	semantic	She wrote the list on a piece of …	paper	0.842	1.000
notebook	unrelated	Chocolate glazed with sprinkles is his favorite kind of …	donut	0.842	1.000
olive	match	He garnished the martini with a green ...	olive	1.000	1.000
olive	semantic	The burger came with a side of chips and a dill ...	pickle	1.000	1.000
olive	unrelated	At the party, we had some salsa and tortilla ...	chips	1.000	1.000
owl	match	The bird that says "hoo" is the ...	owl	0.947	1.000
owl	semantic	The bird whose tail feathers make a colorful fan is a ...	peacock	0.947	1.000
owl	unrelated	My favorite treat is a chocolate chip ...	cookie	0.947	1.000
pants	match	Raymond needed a belt to hold up his ...	pants	1.000	0.750
pants	semantic	For his interview Mr. Jones needed a new ...	suit	1.000	0.750
pants	unrelated	Susan hates dogs, but loves Garfield, her fat tabby ...	cat	1.000	0.750
paper	match	She wrote the list on a piece of ...	paper	1.000	1.000
paper	semantic	A binder of ruled pages used by students is a ...	notebook	1.000	1.000
paper	unrelated	Every Halloween, they turned their home into a haunted ...	house	1.000	1.000
pasta	match	Spaghetti and penne are types of ...	pasta	1.000	0.917
pasta	semantic	Nothing helps a cold like a bowl of chicken noodle ...	soup	1.000	0.917
pasta	unrelated	The boy swept up his mess with the ...	broom	1.000	0.917
peacock	match	The bird whose tail feathers make a colorful fan is a ...	peacock	1.000	1.000
peacock	semantic	The bird that says "hoo" is the ...	owl	1.000	1.000
peacock	unrelated	He laid a trap with cheese to catch the ...	mouse	1.000	1.000
peanuts	match	At the ballpark, the boys enjoyed a bag of salty ...	peanuts	0.737	1.000
peanuts	semantic	On Halloween kids trick-or-treat to collect ...	candy	0.737	1.000
peanuts	unrelated	The animal with a long neck and long legs is a ...	giraffe	0.737	1.000
peas	match	The vegetables that come in pods are ...	peas	0.842	1.000
peas	semantic	The vegetable that comes on a cob is ...	corn	0.842	1.000
peas	unrelated	To start a car, you need the ...	key	0.842	1.000
peeler	match	Cooks remove skins from vegetables using a ...	peeler	0.789	0.667
peeler	semantic	Peter serves the soup out of the pot with a ...	ladle	0.789	0.667
peeler	unrelated	Matt grabbed the swatter to kill the ...	fly	0.789	0.667
pen	match	The ink ran out in my ballpoint ...	pen	1.000	1.000
pen	semantic	John joined the pieces of paper together by pushing down hard on the ...	stapler	1.000	1.000
pen	unrelated	The dog ran in circles chasing his own ...	tail	1.000	1.000
pencil	match	The lead broke so she sharpened the ...	pencil	0.947	0.833
pencil	semantic	The navigation device that points north is a ...	compass	0.947	0.833
pencil	unrelated	With breakfast Julie always drinks some orange ...	juice	0.947	0.833
penguin	match	The bird that looks like it's wearing a tuxedo is a ...	penguin	0.842	0.947
penguin	semantic	The children loved to play Duck-Duck ...	goose	0.842	0.947
penguin	unrelated	He kicked the ball with his left ...	foot	0.842	0.947
penny	match	A one cent coin is called a ...	penny	1.000	1.000
penny	semantic	The smallest bank note is one ...	dollar	1.000	1.000
penny	unrelated	The ornate shade covered the bulb of the ...	lamp	1.000	1.000
pickle	match	The burger came with a side of chips and a dill ...	pickle	0.947	0.833

pickle	semantic	He garnished the martini with a green ...	olive	0.947	0.833
pickle	unrelated	Before the movie started everyone bought some buttery ...	popcorn	0.947	0.833
pig	match	The animal bacon and ham come from is the ...	pig	1.000	1.000
pig	semantic	Susan hates dogs, but loves Garfield, her fat tabby ...	cat	1.000	1.000
pig	unrelated	Honey is produced by an insect called a ...	bee	1.000	1.000
pillow	match	He lies his head down to sleep on his ...	pillow	0.579	1.000
pillow	semantic	Bob sleeps in a king-sized ...	bed	0.579	1.000
pillow	unrelated	His mouth puckered when he ate the sour yellow ...	lemon	0.579	1.000
pizza	match	Chicago is famous for deep dish ...	pizza	1.000	0.667
pizza	semantic	At the Mexican restaurant, he ordered one hard and one soft ...	taco	1.000	0.667
pizza	unrelated	The joint connecting the thigh and shin is the ...	knee	1.000	0.667
plate	match	She put her salad on a large ...	plate	0.737	1.000
plate	semantic	He ate his cereal in a bowl with a ...	spoon	0.737	1.000
plate	unrelated	The nocturnal animal that looks like a masked robber is a ...	raccoon	0.737	1.000
pool	match	She likes to swim laps at the ...	pool	0.947	1.000
pool	semantic	Mary went to Hawaii to tan on a sandy ...	beach	0.947	1.000
pool	unrelated	The farmer shaved the wool off of the ...	sheep	0.947	1.000
popcorn	match	Before the movie started everyone bought some buttery ...	popcorn	1.000	1.000
popcorn	semantic	At the party, we had some salsa and tortilla ...	chips	1.000	1.000
popcorn	unrelated	Many people like the Grand Cherokee, but the Wrangler is my favorite kind of ...	jeep	1.000	1.000
pot	match	She heated the stew in a large metal ...	pot	0.947	1.000
pot	semantic	The chef chopped the vegetables with a ...	knife	0.947	1.000
pot	unrelated	The lead broke so she sharpened the ...	pencil	0.947	1.000
printer	match	He hooked up his computer and discovered the ink had run dry in the ...	printer	0.842	1.000
printer	semantic	A portable computer is called a ...	laptop	0.842	1.000
printer	unrelated	He keeps track of time by looking to the wall at the ...	clock	0.842	1.000
rabbit	match	Alice ended up in Wonderland when she followed the ...	rabbit	0.842	1.000
rabbit	semantic	The salesman was as sly as a ...	fox	0.842	1.000
rabbit	unrelated	The explosion came from a homemade pipe ...	bomb	0.842	1.000
raccoon	match	The nocturnal animal that looks like a masked robber is a ...	raccoon	0.842	1.000
raccoon	semantic	The campers were frightened by a large grizzly ...	bear	0.842	1.000
raccoon	unrelated	She put her salad on a large ...	plate	0.842	1.000
rainbow	match	He found a pot of gold at the end of the ...	rainbow	0.947	0.667
rainbow	semantic	Kids are often frightened by thunder and ...	lightning	0.947	0.667
rainbow	unrelated	The parents bought a stroller for their newborn ...	baby	0.947	0.667
rake	match	The gardener collects the leaves in a pile using a ...	rake	0.947	1.000
rake	semantic	Sandy watered her garden using a rubber ...	hose	0.947	1.000
rake	unrelated	The directions did not match any roads on the ...	map	0.947	1.000
razor	match	Every morning the man shaves using a ...	razor	0.947	1.000
razor	semantic	After brushing his teeth, Mike also uses dental ...	floss	0.947	1.000
razor	unrelated	The bird that looks like it's wearing a tuxedo is a ...	penguin	0.947	1.000
record	match	Before the CD existed, music was listened to off of a black vinyl ...	record	0.789	0.667
record	semantic	Apple's mp3 player is called an ...	ipod	0.789	0.667
record	unrelated	During work Danny sits all day long at his ...	desk	0.789	0.667
ring	match	The man gave his fiancé an engagement ...	ring	1.000	0.750
ring	semantic	A Rolex is an expensive type of ...	watch	1.000	0.750

ring	unrelated	The bike was protected from theft by an expensive ...	lock	1.000	0.750
rock	match	This frozen turkey is as hard as a ...	rock	0.947	1.000
rock	semantic	The Canadian flag features a maple ...	leaf	0.947	1.000
rock	unrelated	To protect his head from sunburn, the bald man wore a wide-brimmed ...	hat	0.947	1.000
rooster	match	The farmer woke up to the cock-a-doodle-doo of the ...	rooster	0.947	1.000
rooster	semantic	Every day the farmer goes to milk his only ...	cow	0.947	1.000
rooster	unrelated	Her foot was cold even though she wore a sock and a winter ...	boot	0.947	1.000
rope	match	They tied the ship to the dock with a ...	rope	0.895	1.000
rope	semantic	The sailor stopped the ship and dropped the ...	anchor	0.895	1.000
rope	unrelated	Snow White was poisoned when she bit into an ...	apple	0.895	1.000
ruler	match	Sam measured the length of the paper using a ...	ruler	0.947	0.750
ruler	semantic	To fix his torn paper he needs some ...	tape	0.947	0.750
ruler	unrelated	Peter serves the soup out of the pot with a ...	ladle	0.947	0.750
sandal	match	The footwear typically worn in the summer is a ...	sandal	0.789	0.917
sandal	semantic	She needed new laces for just one ...	shoe	0.789	0.917
sandal	unrelated	They paddled down the river in a wooden ...	canoe	0.789	0.917
sandwich	match	She snacks on a peanut butter and jelly ...	sandwich	0.947	0.417
sandwich	semantic	He melted cheese over tortilla chips to make ...	nachos	0.947	0.417
sandwich	unrelated	Swimmers protect their eyes by wearing ...	goggles	0.947	0.417
scissors	match	She cut the paper using ...	scissors	1.000	1.000
scissors	semantic	Emily fixed the broken mug with some ...	glue	1.000	1.000
scissors	unrelated	It was windy enough to fly a ...	kite	1.000	1.000
scooter	match	A Vespa is a type of motorized ...	scooter	0.526	0.417
scooter	semantic	The boy went to the half pipe and practiced tricks on his ...	skateboard	0.526	0.417
scooter	unrelated	John joined the pieces of paper together by pushing down hard on the ...	stapler	0.526	0.417
seal	match	The activists scolded the poacher for clubbing a baby ...	seal	0.842	0.583
seal	semantic	The fisherman came upon a pod with a baby bottlenose ...	dolphin	0.842	0.583
seal	unrelated	The boy swept up his mess with the ...	broom	0.842	0.583
sheep	match	The farmer shaved the wool off of the ...	sheep	0.947	1.000
sheep	semantic	At the petting zoo Suzie's snack was stolen by a pesky billy ...	goat	0.947	1.000
sheep	unrelated	The bird whose tail feathers make a colorful fan is a ...	peacock	0.947	1.000
shoe	match	She needed new laces for just one ...	shoe	0.895	0.833
shoe	semantic	The footwear typically worn in the summer is a ...	sandal	0.895	0.833
shoe	unrelated	For his birthday, his mom baked him a chocolate ...	cake	0.895	0.833
skateboard	match	The boy went to the half pipe and practiced tricks on his ...	skateboard	0.947	0.250
skateboard	semantic	A Vespa is a type of motorized ...	scooter	0.947	0.250
skateboard	unrelated	The ink ran out in my ballpoint ...	pen	0.947	0.250
snake	match	In the desert, he got bitten by a rattle ...	snake	1.000	0.750
snake	semantic	The animal with antlers that is much larger than a deer is a ...	moose	1.000	0.750
snake	unrelated	The man happily sat down in the comfortable ...	chair	1.000	0.750
soap	match	In the shower, he washed with a bar of ...	soap	1.000	1.000
soap	semantic	He dried his wet hands with a ...	towel	1.000	1.000
soap	unrelated	Levi makes high quality denim blue ...	jeans	1.000	1.000
sock	match	After doing laundry Derek noticed he was missing just one ...	sock	0.947	1.000
sock	semantic	To protect his head from sunburn, the bald man wore a wide-brimmed ...	hat	0.947	1.000

sock	unrelated	Susan hates dogs, but loves Garfield, her fat tabby ...	cat	0.947	1.000
soup	match	Nothing helps a cold like a bowl of chicken noodle ...	soup	1.000	1.000
soup	semantic	Spaghetti and penne are types of ...	pasta	1.000	1.000
soup	unrelated	Mary cleaned the floor using a bucket and ...	mop	1.000	1.000
speaker	match	The message was broadcast to the students over a loud ...	speaker	0.947	0.250
speaker	semantic	Others won't hear your music when you listen to it wearing ...	headphones	0.947	0.250
speaker	unrelated	Mexican food comes with a side of rice and ...	beans	0.947	0.250
spoon	match	He ate his cereal in a bowl with a ...	spoon	1.000	0.833
spoon	semantic	She put her salad on a large ...	plate	1.000	0.833
spoon	unrelated	The campers were frightened by a large grizzly ...	bear	1.000	0.833
stapler	match	John joined the pieces of paper together by pushing down hard on the ...	stapler	0.684	1.000
stapler	semantic	The ink ran out in my ballpoint ...	pen	0.684	1.000
stapler	unrelated	A Vespa is a type of motorized ...	scooter	0.684	1.000
star	match	The hopeful girl wished upon a ...	star	1.000	1.000
star	semantic	Neil Armstrong was the first man to walk on the ...	moon	1.000	1.000
star	unrelated	The national bird of the United States is the ...	eagle	1.000	1.000
stool	match	I like to sit at the bar on a tall wooden ...	stool	0.842	0.917
stool	semantic	The man happily sat down in the comfortable ...	chair	0.842	0.917
stool	unrelated	It was windy enough to fly a ...	kite	0.842	0.917
suit	match	For his interview Mr. Jones needed a new ...	suit	0.632	0.750
suit	semantic	Raymond needed a belt to hold up his ...	pants	0.632	0.750
suit	unrelated	The wild cat that runs the fastest is the ...	cheetah	0.632	0.750
swan	match	Every ugly duckling eventually becomes a beautiful ...	swan	0.895	0.917
swan	semantic	The national bird of the United States is the ...	eagle	0.895	0.917
swan	unrelated	Neil Armstrong was the first man to walk on the ...	moon	0.895	0.917
sweater	match	Each year his grandmother knit him an ugly Christmas ...	sweater	0.947	0.917
sweater	semantic	Her foot was cold even though she wore a sock and a winter ...	boot	0.947	0.917
sweater	unrelated	The rodent famous for building dams in rivers is called a ...	beaver	0.947	0.917
table	match	For dinner, the family gathers at the dining room ...	table	0.947	0.750
table	semantic	During work Danny sits all day long at his ...	desk	0.947	0.750
table	unrelated	Bob sleeps in a king-sized ...	bed	0.947	0.750
taco	match	At the Mexican restaurant, he ordered one hard and one soft ...	taco	0.895	1.000
taco	semantic	Chicago is famous for deep dish ...	pizza	0.895	1.000
taco	unrelated	The joint connecting the forearm to the bicep is the ...	elbow	0.895	1.000
tail	match	The dog ran in circles chasing his own ...	tail	0.947	1.000
tail	semantic	The bird could not fly because of an injured ...	wing	0.947	1.000
tail	unrelated	The boy went to the half pipe and practiced tricks on his ...	skateboard	0.947	1.000
tank	match	The soldiers were protected by the armored fighting vehicle called a ...	tank	0.947	0.917
tank	semantic	The explosion came from a homemade pipe ...	bomb	0.947	0.917
tank	unrelated	The student took her books home in her ...	backpack	0.947	0.917
tape	match	To fix his torn paper he needs some ...	tape	1.000	1.000
tape	semantic	Sam measured the length of the paper using a ...	ruler	1.000	1.000
tape	unrelated	The boy took a pole to the lake to catch a ...	fish	1.000	1.000
tie	match	He wore a suit with a Windsor knot in his ...	tie	0.842	0.917
tie	semantic	Levi makes high quality denim blue ...	jeans	0.842	0.917

tie	unrelated	The king of the jungle is the ...	lion	0.842	0.917
tire	match	During their bike tour Mike had to fix a flat ...	tire	1.000	0.917
tire	semantic	The stagecoach wasn't moving because of the broken wagon ...	wheel	1.000	0.917
tire	unrelated	Matt grabbed the swatter to kill the ...	fly	1.000	0.917
toast	match	John spread butter and grape jelly on his morning ...	toast	0.842	0.917
toast	semantic	Angela likes cream cheese on a ...	bagel	0.842	0.917
toast	unrelated	The directions did not match any roads on the ...	map	0.842	0.917
toothbrush	match	Proper dental hygiene includes cleaning each day with a bristled ...	toothbrush	0.526	1.000
toothbrush	semantic	The man brushed his hair using a fine-toothed ...	comb	0.526	1.000
toothbrush	unrelated	She put her salad on a large ...	plate	0.526	1.000
towel	match	He dried his wet hands with a ...	towel	1.000	1.000
towel	semantic	In the shower, he washed with a bar of ...	soap	1.000	1.000
towel	unrelated	The king of the jungle is the ...	lion	1.000	1.000
tree	match	The bird built a nest high up in an elm ...	tree	1.000	0.917
tree	semantic	In the desert, he pricked his finger on the spine of a ...	cactus	1.000	0.917
tree	unrelated	The joint connecting the forearm to the bicep is the ...	elbow	1.000	0.917
trophy	match	The team that won the tournament took home a ...	trophy	0.947	0.917
trophy	semantic	In the Olympic games, she won a gold ...	medal	0.947	0.917
trophy	unrelated	The homeless man slept on a park ...	bench	0.947	0.917
tweezers	match	Andy removed a splinter with some ...	tweezers	0.842	0.500
tweezers	semantic	He applied the medicine to his eye using a ...	dropper	0.842	0.500
tweezers	unrelated	To protect his head from sunburn, the bald man wore a wide-brimmed ...	hat	0.842	0.500
van	match	We couldn't get a truck, but we managed to pack everything in a moving ...	van	0.895	0.420
van	semantic	Many people like the Grand Cherokee, but the Wrangler is my favorite kind of ...	jeep	0.895	0.420
van	unrelated	Nick sneezed and blew his ...	nose	0.895	0.420
vase	match	She placed the flowers in a glass ...	vase	0.947	0.833
vase	semantic	When the power went out, they lit a ...	candle	0.947	0.833
vase	unrelated	The bird built a nest high up in an elm ...	tree	0.947	0.833
vest	match	Between his jacket and shirt, the usher had buttoned his ...	vest	0.579	0.917
vest	semantic	Under their shirts, most women wear a supportive ...	bra	0.579	0.917
vest	unrelated	The lead broke so she sharpened the ...	pencil	0.579	0.917
waffle	match	At brunch, Maggie either eats pancakes or a Belgian ...	waffle	0.895	0.417
waffle	semantic	Chocolate glazed with sprinkles is his favorite kind of ...	donut	0.895	0.417
waffle	unrelated	Mary went to Hawaii to tan on a sandy ...	beach	0.895	0.417
watch	match	A Rolex is an expensive type of ...	watch	1.000	1.000
watch	semantic	The man gave his fiancé an engagement ...	ring	1.000	1.000
watch	unrelated	To start a car, you need the ...	key	1.000	1.000
wheel	match	The stagecoach wasn't moving because of the broken wagon ...	wheel	0.895	1.000
wheel	semantic	During their bike tour Mike had to fix a flat ...	tire	0.895	1.000
wheel	unrelated	The blind, flying rodent that lives in caves is the ...	bat	0.895	1.000
window	match	To let some cool air in the apartment they opened a ...	window	1.000	0.917
window	semantic	She felt hot in the office and plugged in the ...	fan	1.000	0.917
window	unrelated	Under their shirts, most women wear a supportive ...	bra	1.000	0.917
wine	match	The old girlfriends chatted over a bottle of red ...	wine	1.000	1.000

wine	semantic	With breakfast Julie always drinks some orange ...	juice	1.000	1.000
wine	unrelated	Apple's mp3 player is called an ...	ipod	1.000	1.000
wing	match	The bird could not fly because of an injured ...	wing	1.000	1.000
wing	semantic	The dog ran in circles chasing his own ...	tail	1.000	1.000
wing	unrelated	Emily fixed the broken mug with some ...	glue	1.000	1.000
zebra	match	The horse-like animal with black and white stripes is the ...	zebra	1.000	0.833
zebra	semantic	The animal with a long neck and long legs is a ...	giraffe	1.000	0.833
zebra	unrelated	On Halloween kids trick-or-treat to collect ...	candy	1.000	0.833

Appendix C: Full model coefficient tables and random effect structure

In all tables, significance, as assessed with chi-square tests of nested models with and without predictor, is reported in the last column. Significant effects ($p < 0.05$) are bolded; marginal effects ($p < 0.10$) are italicized. When chi-square models with the fixed effect held out did not converge (DNC), preventing nested model comparison, the absolute value of the t-statistic was used as a proxy. A (*) is used to indicate significance as assessed with a t-statistic > 2, (.) indicates a marginal effect for a t-statistic > 1.5, and n.s. indicates a non-significant effect for a t-statistic < 1.5. For logistic regressions, the z-statistic was used similarly.

Table A2: Summary of single-experiment error models. There was insufficient variance in the distribution of completion errors in experiments 1 and 3; as a result, a model is reported for experiment 2 only.

Experiment 2: Young adults, time pressure					
	Estimate (β)	Std. error	Z-value	χ^2	P (chisq.)
Block	*-1.308*	*0.757*	*-1.728*	*3.83*	*0.051*
Semantic relatedness	**5.771**	**1.782**	**3.238**	**9.25**	**0.002**
Block:semantic	**-1.143**	**0.820**	**-1.394**	**6.25**	**0.012**

<u>Random effect structure</u>

Experiment 2: correlated subject and item slopes for semantic relatedness, block

Table A3: Summary of single-experiment response time models.

Experiment 1: Young adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	**-0.060**	**0.007**	**-8.74**	**30.39**	**< 0.001**
Match status	**0.247**	**0.020**	**12.44**	**43.95**	**< 0.001**
Semantic relatedness	*-0.014*	*0.008*	*-1.70*	*DNC*	*(.)*
Shipley score	-0.004	0.007	-0.53	0.20	0.655
Block:match	-0.011	0.014	-0.77	0.58	0.445
Block:semantic	**-0.019**	**0.012**	**-1.68**	**6.22**	**0.012**
Experiment 2: Young adults, time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	**-0.062**	**0.007**	**-9.37**	**33.02**	**< 0.001**
Match status	**0.307**	**0.024**	**12.83**	**43.17**	**< 0.001**
Semantic relatedness	0.001	0.008	0.07	0.00	0.944
Shipley score	-0.003	0.003	-1.10	1.17	0.279
Block:match	**0.031**	**0.014**	**2.18**	**4.25**	**0.039**
Block:semantic	*-0.017*	*0.009*	*-1.91*	*3.63*	*0.057*
Experiment 3: Older adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	**-0.054**	**0.008**	**-6.78**	**DNC**	**(*)**
Match status	**0.232**	**0.019**	**12.37**	**DNC**	**(*)**
Semantic relatedness	0.003	0.011	0.29	DNC	
Shipley score	0.002	0.007	0.34	DNC	
Block:match	-0.006	0.015	-0.37	DNC	
Block:semantic	0.010	0.012	0.82	DNC	

Random effect structure

Experiment 1: correlated subject slopes for block, match status, block:match, block:semantic; ; correlated item slopes for Shipley score, match status, block, semantic relatedness, block:semantic

Experiment 2: decorrelated subject slopes for block, match status, block:match; decorrelated item slopes for Shipley score, semantic relatedness, block, match status, block:match

Experiment 3: correlated subject slopes for block, match status, semantic relatedness, block:match; correlated item slopes for Shipley score, block, match status, semantic relatedness, block:match, block:semantic

Table A4: Summary of cross-experiment comparison models, response time.

Experiment 1 vs. experiment 2					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**-0.083**	**0.034**	**-2.40**	**5.45**	**0.020**
Block	**-0.061**	**0.005**	**-12.70**	**65.34**	**< 0.001**
Match status	**0.279**	**0.017**	**16.60**	**86.42**	**< 0.001**
Semantic relatedness	-0.006	0.007	-0.90	0.84	0.359
Shipley score	-0.003	0.003	-1.00	1.03	0.310
Experiment:block	-0.001	0.008	-0.20	0.03	0.861
Experiment:match	**0.065**	**0.029**	**2.20**	**4.70**	**0.030**
Experiment:semantic	0.013	0.010	1.30	1.73	0.189
Block:match	0.011	0.010	1.10	1.18	0.278
Block:semantic	**-0.019**	**0.007**	**-2.90**	**7.97**	**0.005**
Experiment:block:match	**0.039**	**0.019**	**2.00**	**3.92**	**0.048**
Experiment:block:semantic	0.005	0.012	0.40	0.19	0.662
Experiment 1 vs. experiment 3					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**0.222**	**0.058**	**3.83**	**12.34**	**< 0.001**
Block	**-0.057**	**0.005**	**-10.57**	**51.89**	**< 0.001**
Match status	**0.234**	**0.014**	**16.24**	**85.62**	**< 0.001**
Semantic relatedness	-0.007	0.007	-0.99	0.98	0.323
Shipley score	0.000	0.006	-0.07	0.01	0.942
Experiment:block	0.005	0.010	0.44	0.19	0.661
Experiment:match	-0.030	0.024	-1.21	1.45	0.229
Experiment:semantic	0.016	0.011	1.46	2.03	0.154
Block:match	-0.013	0.009	-1.44	2.02	0.156
Block:semantic	0.000	0.006	-0.04	0.00	0.967
Experiment:block:match	-0.001	0.019	-0.08	0.01	0.936
Experiment:block:semantic	**0.031**	**0.014**	**2.16**	**4.53**	**0.033**

Random effect structure

Model comparing experiment 1 and experiment 2: decorrelated subject slopes for block, match status, block:match, decorrelated item slopes for Shipley score, match status, experiment, block, semantic relatedness, experiment:match, block:match, block:semantic, experiment:block:match

Model comparing experiment 1 and experiment 3: decorrelated subject slopes for block, match status, semantic relatedness, block:match; decorrelated item slopes for Shipley score, experiment, block, match status, semantic relatedness, experiment:match, block:semantic, experiment:block:match, experiment:block:unrelated

Table A5: Summary of single-experiment word duration models.

Experiment 1: Young adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	4.145	4.346	0.95	0.90	0.347
Match status	**18.783**	**3.405**	**5.52**	**30.09**	**< 0.001**
Semantic relatedness	-4.268	2.749	-1.55	2.33	0.127
Trial-level RT	**-33.281**	**13.274**	**-2.51**	**5.73**	**0.017**
BLUP from RT model	*146.588*	*76.192*	*1.92*	*2.75*	*0.097*
Shipley score	-0.098	2.470	-0.04	0.00	0.969
Block:match	1.308	3.793	0.34	0.12	0.731
Block:semantic	2.108	3.739	0.56	0.31	0.576
RT:match	-19.718	15.563	-1.27	1.56	0.212
RT:semantic	**50.667**	**15.344**	**3.30**	**9.84**	**0.002**
Experiment 2: Young adults, time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	-4.577	2.863	-1.60	2.39	0.122
Match status	**12.717**	**3.612**	**3.52**	**10.12**	**0.002**
Semantic relatedness	*-5.298*	*2.767*	*-1.92*	*3.66*	*0.056*
Trial-level RT	-10.080	8.258	-1.22	1.47	0.225
BLUP from RT model	427.909	471.328	0.91	0.81	0.369
Shipley score	*-3.103*	*1.548*	*-2.01*	*3.62*	*0.056*
Block:match	-7.161	4.337	-1.65	2.69	0.101
Block:semantic	2.987	4.302	0.69	0.48	0.489
RT:match	*27.527*	*14.706*	*1.87*	*3.45*	*0.063*
RT:semantic	12.328	14.092	0.88	0.76	0.382
Experiment 3: Older adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	-4.664	3.655	-1.28	1.56	0.212
Match status	5.105	4.451	1.15	1.31	0.253
Semantic relatedness	**-11.304**	**3.711**	**-3.05**	**8.86**	**0.003**
Trial-level RT	-10.255	9.019	-1.14	1.27	0.259
BLUP from RT model	**179.651**	**82.441**	**2.18**	**4.22**	**0.040**
Shipley score	**-6.711**	**3.107**	**-2.16**	**4.19**	**0.041**
Block:match	**-19.786**	**7.558**	**-2.62**	**5.83**	**0.016**
Block:semantic	5.570	5.452	1.02	1.04	0.308
RT:match	-20.459	15.402	-1.33	1.75	0.186
RT:semantic	7.360	17.078	0.43	0.19	0.667

Random effect structure

Experiment 1: correlated subject slopes for block, response time, RT:semantic; correlated item slopes for Shipley score, response time, BLUPs from RT model, block:semantic

Experiment 2: decorrelated subject slopes for block, match, response time, block:semantic; decorrelated item slopes for Shipley score, match status, BLUPs from RT model, block:match, block:semantic

Experiment 3: decorrelated subject slopes for block, response time, block:match, block:semantic, response time:semantic; decorrelated item slopes for Shipley score, response time, BLUPs, block:match, block:semantic, RT:match, RT:semantic

Table A6: Cross-experiment comparison models, word duration.

Experiment 1 vs. experiment 2					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**-53.213**	**15.479**	**-3.44**	**10.24**	**0.001**
Block	-0.320	2.654	-0.12	DNC	
Match status	**16.310**	**2.442**	**6.68**	**44.18**	**< 0.001**
Semantic relatedness	**-5.564**	**1.983**	**-2.81**	**7.85**	**0.005**
Trial-level RT	**-21.480**	**7.563**	**-2.84**	**7.66**	**0.006**
BLUP from RT model	-69.426	299.063	-0.23	0.05	0.817
Shipley score	*-2.340*	*1.352*	*-1.73*	*2.88*	*0.090*
Experiment:block	-8.629	5.302	-1.63	2.55	0.110
Experiment:match	1.061	4.821	0.22	0.05	0.826
Experiment:unrelated	0.605	3.928	0.15	0.02	0.878
Block:match	-4.295	2.654	-1.62	2.60	0.107
Block:semantic	2.531	2.957	0.86	0.73	0.393
Experiment:RT	11.771	14.258	0.83	0.67	0.411
Match:RT	7.070	10.322	0.68	0.47	0.494
Unrelated:RT	17.762	10.525	1.69	2.82	0.093
Experiment:block:match	-5.612	5.193	-1.08	1.16	0.281
Experiment:block:unrelated	1.623	4.720	0.34	0.12	0.731
Experiment:RT:match	**55.878**	**20.317**	**2.75**	**DNC**	**(*)**
Experiment:RT:unrelated	-17.517	19.453	-0.90	0.81	0.369
Experiment 1 vs. experiment 3					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**54.658**	**23.563**	**2.32**	**5.02**	**0.025**
Block	-0.764	2.804	-0.27	0.07	0.785
Match status	**14.856**	**3.050**	**4.87**	**23.58**	**< 0.001**
Semantic relatedness	**-5.451**	**2.455**	**-2.22**	**4.92**	**0.027**
Trial-level RT	**-24.785**	**8.425**	**-2.94**	**8.25**	**0.004**
BLUP from RT model	-285.849	456.564	-0.63	0.36	0.550
Shipley score	*-4.191*	*2.358*	*-1.78*	*2.82*	*0.093*
Experiment:block	-8.503	5.605	-1.52	2.23	0.136
Experiment:match	-2.320	6.083	-0.38	0.14	0.704
Experiment:unrelated	**-10.309**	**4.904**	**-2.1**	**4.41**	**0.038**
Block:match	-4.368	3.156	-1.38	1.91	0.167
Block:semantic	2.814	2.829	0.99	0.98	0.321
Experiment:RT	-1.668	15.298	-0.11	0.01	0.914
Match:RT	-9.736	10.881	-0.89	0.79	0.375
Unrelated:RT	*17.368*	*9.539*	*1.82*	*3.29*	*0.070*
Experiment:block:match	**-20.408**	**6.293**	**-3.24**	**10.48**	**0.001**
Experiment:block:unrelated	5.289	5.660	0.93	0.87	0.351
Experiment:RT:match	-19.191	21.651	-0.89	0.78	0.378
Experiment:RT:unrelated	-25.535	19.094	-1.34	1.77	0.183

Random effect structure

Model comparing experiment 1 and experiment 2: decorrelated subject slopes for block, response time, block:semantic; decorrelated item slopes for experiment, block, response time, BLUPs from RT model, block:match, block:semantic, RT:semantic, RT:experiment

Model comparing experiment 1 and experiment 3: correlated subject slopes for block, response time; correlated item slopes for Shipley score, experiment, response time, BLUPs from RT model

Table A7: Summary of single-experiment models of initial consonant duration.

Experiment 1: Young adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	**-2.245**	**0.819**	**-2.74**	**6.33**	**0.012**
Match status	**10.852**	**1.895**	**5.73**	**16.90**	**< 0.001**
Semantic relatedness	-2.397	1.759	-1.36	1.80	0.180
Trial-level RT	**-35.532**	**4.728**	**-7.52**	**30.07**	**< 0.001**
BLUP from RT model	**44.145**	**20.043**	**2.20**	**4.34**	**0.037**
Number of consonants	**39.786**	**7.322**	**5.43**	**26.17**	**< 0.001**
Shipley score	0.125	0.651	0.19	0.04	0.849
Block:match	0.621	2.079	0.30	0.09	0.768
Block:semantic	3.139	2.103	1.49	2.19	0.139
RT:match	**-31.232**	**8.693**	**-3.59**	**9.87**	**0.002**
RT:semantic	5.651	8.045	0.70	0.46	0.499

Experiment 2: Young adults, time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	**-2.315**	**0.736**	**-3.14**	**7.96**	**0.005**
Match status	**8.621**	**1.953**	**4.42**	**15.36**	**< 0.001**
Semantic relatedness	**-3.759**	**1.478**	**-2.54**	**6.44**	**0.011**
Trial-level RT	**-14.980**	**3.592**	**-4.17**	**14.61**	**< 0.001**
BLUP from RT model	-11.917	76.347	-0.16	0.02	0.876
Number of consonants	**27.276**	**5.810**	**4.70**	**20.42**	**< 0.001**
Shipley score	*-0.498*	*0.251*	*-1.99*	*3.59*	*0.058*
Block:match	0.122	1.796	0.07	0.00	0.946
Block:semantic	-1.534	1.885	-0.81	0.62	0.431
RT:match	-0.866	7.570	-0.11	0.01	0.912
RT:semantic	*16.658*	*9.217*	*1.81*	*3.18*	*0.075*

Experiment 3: Older adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	-0.519	0.907	-0.57	0.32	0.569
Match status	*3.381*	*2.006*	*1.69*	*2.72*	*0.099*
Semantic relatedness	-2.486	1.604	-1.55	2.39	0.122
Trial-level RT	*-19.745*	*6.219*	*-3.18*	*8.54*	*0.004*
BLUP from RT model	**35.899**	**20.584**	**1.74**	**2.80**	**0.094**
Number of consonants	**40.337**	**9.971**	**4.05**	**15.56**	**< 0.001**
Shipley score	*-1.544*	*0.770*	*-2.01*	*3.71*	*0.054*
Block:match	-0.909	2.936	-0.31	0.09	0.759
Block:semantic	-1.177	2.197	-0.54	0.28	0.594
RT:match	5.522	7.343	0.75	0.55	0.458
RT:semantic	-5.706	6.934	-0.82	0.67	0.413

Random effect structure

Experiment 1: correlated subject slopes for block, match status, semantic relatedness, response time, number of consonants, RT:match, RT:semantic; correlated item slopes for match, RT, BLUP from RT model, Shipley score, block:match, block:semantic, RT:match

Experiment 2: decorrelated subject slopes for block, match status, response time, number of consonants, block:semantic, RT:semantic; decorrelated item slopes for block, match, semantic, block:semantic, RT:semantic

Experiment 3: decorrelated subject slopes for block, response time, number of consonants, block:match, RT:match; decorrelated item slopes for RT, BLUP from RT model, Shipley score, block:semantic, RT:semantic

Table A8: Cross-experiment comparison models, first consonant duration.

Experiment 1 vs. experiment 2

	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**-11.380**	**2.996**	**-3.80**	**12.25**	**< 0.001**
Block	**-2.081**	**0.573**	**-3.63**	**11.43**	**< 0.001**
Match status	**10.418**	**1.305**	**7.98**	**43.78**	**< 0.001**
Semantic relatedness	**-2.562**	**1.038**	**-2.47**	**5.95**	**0.015**
Trial-level RT	**-25.166**	**2.997**	**-8.40**	**47.38**	**< 0.001**
BLUP from RT model	**42.837**	**18.177**	**2.36**	**5.23**	**0.022**
Number of consonants	**34.291**	**6.306**	**5.44**	**27.36**	**< 0.001**
Shipley score	*-0.428*	*0.251*	*-1.70*	*2.81*	*0.094*
Experiment:block	0.266	1.141	0.23	0.05	0.816
Experiment:match	-3.557	2.475	-1.44	2.05	0.152
Experiment:unrelated	-1.325	2.073	-0.64	0.41	0.523
Block:match	0.751	1.388	0.54	0.29	0.589
Block:semantic	0.592	1.398	0.42	0.18	0.673
Experiment:RT	**19.451**	**5.818**	**3.34**	**DNC**	**(*)**
Match:RT	**-14.512**	**5.492**	**-2.64**	**6.79**	**0.009**
Unrelated:RT	6.538	5.833	1.12	1.20	0.274
Experiment:block:match	-0.068	2.662	-0.03	0.00	0.980
Experiment:block:unrelated	*-4.844*	*2.657*	*-1.82*	*3.30*	*0.069*
Experiment:RT:match	**25.716**	**10.724**	**2.40**	**5.72**	**0.017**
Experiment:RT:unrelated	5.693	11.839	0.48	0.23	0.634

Experiment 1 vs. experiment 3

	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**12.957**	**4.996**	**2.59**	**DNC**	**(*)**
Block	*-1.205*	*0.647*	*-1.86*	*DNC*	*(.)*
Match status	**5.565**	**1.482**	**3.75**	**DNC**	**(*)**
Semantic relatedness	-1.543	1.365	-1.13	DNC	
Trial-level RT	**-24.060**	**4.025**	**-5.98**	**28.72**	**< 0.001**
BLUP from RT model	**38.690**	**11.159**	**3.47**	**DNC**	**(*)**
Number of consonants	**35.186**	**6.996**	**5.03**	**DNC**	**(*)**
Shipley score	-0.042	0.381	-0.11	0.01	0.923
Experiment:block	1.243	1.264	0.98	DNC	
Experiment:match	**-6.284**	**2.969**	**-2.12**	**DNC**	**(*)**
Experiment:unrelated	0.145	2.653	0.06	DNC	

Block:match	0.344	2.012	0.17	0.03	0.865
Block:semantic	1.130	1.652	0.68	DNC	
Experiment:RT	9.926	7.426	1.34	1.67	0.197
Match:RT	-6.910	5.447	-1.27	DNC	
Unrelated:RT	-0.566	5.512	-0.10	0.01	0.921
Experiment:block:match	-4.458	3.606	-1.24	DNC	
Experiment:block:unrelated	*-5.126*	*2.715*	*-1.89*	*3.50*	*0.061*
Experiment:RT:match	**30.276**	**10.682**	**2.83**	**DNC**	**(*)**
Experiment:RT:unrelated	*-18.670*	*10.489*	*-1.78*	*2.83*	*0.093*

Experiment 1 vs. experiment 3	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**12.95734**	**4.99625**	**2.593**	**DNC**	**(*)**
Block	*-1.20542*	*0.64715*	*-1.863*	*DNC*	*(.)*
Match status	**5.56521**	**1.48246**	**3.754**	**DNC**	**(*)**
Semantic relatedness	-1.54258	1.36484	-1.13	DNC	
Trial-level RT	**-24.06**	**4.02537**	**-5.977**	**28.72**	**<0.001**
BLUP from RT model	**38.68961**	**11.15915**	**3.467**	**DNC**	**(*)**
Number of consonants	**35.1864**	**6.99577**	**5.03**	**DNC**	**(*)**
Shipley score	-0.04195	0.38072	-0.11	0.01	0.923
Experiment:block	1.24325	1.26446	0.983	DNC	
Experiment:match	**-6.28383**	**2.96939**	**-2.116**	**DNC**	**(*)**
Experiment:unrelated	0.14536	2.65251	0.055	DNC	
Block:match	0.34448	2.01241	0.171	0.03	0.865
Block:semantic	1.13002	1.65178	0.684	DNC	
Experiment:RT	9.92626	7.42608	1.337	1.67	0.197
Match:RT	-6.90979	5.44739	-1.268	DNC	
Unrelated:RT	-0.56578	5.51224	-0.103	0.01	0.921
Experiment:block:match	-4.45767	3.60568	-1.236	DNC	
Experiment:block:unrelated	*-5.1259*	*2.71532*	*-1.888*	*3.50*	*0.061*
Experiment:RT:match	**30.27565**	**10.68203**	**2.834**	**DNC**	**(*)**
Experiment:RT:unrelated	*-18.6699*	*10.4892*	*-1.78*	*2.83*	*0.093*

Random effect structure

Model comparing experiment 1 and experiment 2: decorrelated subject slopes for block, match status, response time, number of consonants, RT:match, RT:semantic; decorrelated item slopes for group, match status, response time, BLUPs from Rt model, Shipley score, block:match, block:unrelated, RT:match, RT:unrelated, RT:experiment, RT:semantic:experiment, block:match:experiment, block:unrelated:experiment

Model comparing experiment 1 and experiment 3: decorrelated subject slopes for block, unrelated, response time, number of consonants, block:match, RT:unrelated; decorrelated item slopes for experiment, response time, BLUPs from RT model, Shipley score, block:match, block:unrelated, RT:unrelated

Table A9: Summary of single-experiment models of vowel duration.

Experiment 1: Young adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	**5.141**	**1.963**	**2.62**	**5.81**	**0.016**
Match status	-0.451	1.801	-0.25	0.06	0.802
Semantic relatedness	0.308	1.437	0.21	0.05	0.831
Trial-level RT	*7.026*	*4.119*	*1.71*	*2.89*	*0.089*
BLUP from RT model	*61.607*	*34.581*	*1.78*	*2.98*	*0.087*
Shipley score	-0.116	1.134	-0.10	0.01	0.918
Block:match	**-4.774**	**2.300**	**-2.08**	**4.24**	**0.040**
Block:semantic	-1.622	1.958	-0.83	0.68	0.409
RT:match	-5.604	7.936	-0.71	0.50	0.481
RT:semantic	**20.607**	**7.622**	**2.70**	**6.99**	**0.008**
Experiment 2: Young adults, time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	-0.608	1.148	-0.53	0.28	0.598
Match status	**3.395**	**1.723**	**1.97**	**3.85**	**0.050**
Semantic relatedness	-0.284	1.412	-0.20	0.04	0.841
Trial-level RT	7.840	4.725	1.66	2.64	0.104
BLUP from RT model	194.931	199.889	0.98	0.93	0.336
Shipley score	-1.050	0.656	-1.60	2.39	0.122
Block:match	**-7.130**	**2.294**	**-3.11**	**9.28**	**0.002**
Block:semantic	-0.960	2.036	-0.47	0.22	0.640
RT:match	7.391	7.536	0.98	0.96	0.328
RT:semantic	4.087	7.339	0.56	0.31	0.579
Experiment 3: Older adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	1.083	1.351	0.80	0.62	0.426
Match status	2.743	1.839	1.49	1.90	0.168
Semantic relatedness	-0.729	1.540	-0.47	0.22	0.636
Trial-level RT	**-8.231**	**3.896**	**-2.11**	**4.39**	**0.036**
BLUP from RT model	*37.901*	*21.962*	*1.73*	*2.77*	*0.096*
Shipley score	-0.667	0.824	-0.81	0.64	0.422
Block:match	-1.362	2.353	-0.58	0.33	0.564
Block:semantic	1.091	2.369	0.46	0.21	0.647
RT:match	-6.539	6.338	-1.03	1.06	0.303
RT:semantic	-0.173	5.544	-0.03	0.00	0.975

Random effect structure

Experiment 1: decorrelated subject slopes for block; decorrelated item slopes for Shipley score, response time, BLUP from RT model, block:match, block:unrelated, RT:unrelated

Experiment 2: decorrelated subject slopes for block, match, response time, block:match, block:semantic; decorrelated item slopes for Shipley score, RT, BLUPs from RT model, block:match, block:semantic

Experiment 3: decorrelated subject slopes for block, block:semantic, RT:match; decorrelated item slopes for Shipley score, block, semantic relatedness, response time, BLUPs from RT model, block:match, block:semantic

Table A10: Cross-experiment comparison models, vowel duration.

Experiment 1 vs. experiment 2	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**-15.270**	**6.270**	**-2.44**	**5.50**	**0.019**
Block	**2.337**	**1.116**	**2.09**	**4.14**	**0.042**
Match status	1.199	1.307	0.92	0.84	0.359
Semantic relatedness	-0.453	1.054	-0.43	0.18	0.668
Trial-level RT	**7.660**	**3.245**	**2.36**	**5.29**	**0.022**
BLUP from RT model	*65.306*	*36.915*	*1.77*	*3.00*	*0.083*
Shipley score	-0.807	0.531	-1.52	2.23	0.135
Experiment:block	**-5.640**	**2.230**	**-2.53**	**5.90**	**0.015**
Experiment:match	2.006	2.590	0.77	0.60	0.439
Experiment:unrelated	1.382	2.102	0.66	0.430	0.511
Block:match	**-5.599**	**1.390**	**-4.03**	**16.19**	**< 0.001**
Block:semantic	-0.862	1.266	-0.68	0.46	0.500
Experiment:RT	4.546	6.044	0.75	0.56	0.453
Match:RT	-0.750	5.613	-0.13	0.02	0.894
Unrelated:RT	**13.383**	**5.329**	**2.51**	**6.29**	**0.012**
Experiment:block:match	-2.530	2.779	-0.91	0.83	0.363
Experiment:block:unrelated	0.808	2.530	0.32	0.10	0.750
Experiment:RT:match	16.470	11.067	1.49	2.19	0.139
Experiment:RT:unrelated	-15.228	10.636	-1.43	2.04	0.153

Experiment 1 vs. experiment 3	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**15.550**	**6.512**	**2.39**	**5.31**	**0.021**
Block	**3.026**	**1.134**	**2.67**	**6.52**	**0.011**
Match status	0.844	1.437	0.59	0.35	0.557
Semantic relatedness	1.342	1.132	1.19	1.40	0.236
Trial-level RT	0.665	2.648	0.25	0.06	0.802
BLUP from RT model	**45.602**	**18.697**	**2.44**	**5.53**	**0.019**
Shipley score	-0.563	0.660	-0.85	0.72	0.396
Experiment:block	*-4.289*	*2.267*	*-1.89*	*3.41*	*0.065*
Experiment:match	4.596	2.867	1.60	2.56	0.109
Experiment:unrelated	-1.955	2.267	-0.86	0.74	0.389
Block:match	*-3.293*	*1.771*	*-1.86*	*3.40*	*0.065*
Block:semantic	-0.562	1.338	-0.42	0.18	0.675
Experiment:RT	**-15.369**	**5.876**	**-2.62**	**0.68**	**0.009**
Match:RT	-7.951	5.140	-1.55	2.39	0.122
Unrelated:RT	**10.594**	**4.660**	**2.27**	**5.15**	**0.023**
Experiment:block:match	1.516	2.916	0.52	0.27	0.604
Experiment:block:unrelated	3.319	2.607	1.27	1.62	0.204
Experiment:RT:match	-5.023	10.230	-0.49	0.24	0.624
Experiment:RT:unrelated	**-21.417**	**9.290**	**-2.31**	**5.29**	**0.021**

Random effect structure

Model comparing experiment 1 and experiment 2: decorrelated subject slopes for block, response time; decorrelated item slopes for Shipley score, experiment, response time, BLUPs from RT model

Model comparing experiment 1 and experiment 3: decorrelated subject slopes for block, block:semantic; decorrelated item slopes for Shipley score, experiment, response time, BLUPs from RT model, block:match, block:unrelated

Table A11: Summary of single-experiment models of vowel distance.

	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment 1: Young adults, no time pressure					
Block	*-3.708*	*1.922*	*-1.93*	*3.72*	*0.054*
Match status	7.479	4.857	1.54	2.36	0.124
Semantic relatedness	-3.849	3.902	-0.987	0.97	0.324
Trial-level RT	3.484	10.404	0.335	0.11	0.738
BLUP from RT model	71.482	81.217	0.88	0.76	0.384
Shipley score	2.574	2.721	0.946	0.87	0.349
Block:match	-3.317	7.281	-0.456	0.2	0.651
Block:semantic	0.599	6.105	0.098	0.01	0.922
RT:match	23.994	21.374	1.123	1.25	0.263
RT:semantic	-32.140	20.031	-1.605	2.56	0.109
Experiment 2: Young adults, time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	-5.966	2.719	-2.195	**4.37**	**0.037**
Match status	**12.921**	**5.261**	**2.456**	**6.01**	**0.014**
Semantic relatedness	1.172	4.391	0.267	0.07	0.790
Trial-level RT	-3.409	10.141	-0.336	0.11	0.738
BLUP from RT model	*503.629*	*268.912*	*1.873*	*3.13*	*0.077*
Shipley score	-0.623	0.884	-0.705	0.49	0.485
Block:match	-3.774	5.751	-0.656	0.51	0.513
Block:semantic	-0.365	6.010	-0.061	0.95	0.952
RT:match	12.462	21.787	0.572	0.57	0.569
RT:semantic	2.781	22.242	0.125	0.02	0.901
Experiment 3: Older adults, no time pressure					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Block	-0.200	2.098	-0.096	0.01	0.924
Match status	1.295	6.496	0.199	0.04	0.843
Semantic relatedness	-2.664	4.358	-0.611	0.36	0.547
Trial-level RT	**21.875**	**10.307**	**2.122**	**4.45**	**0.035**
BLUP from RT model	20.551	48.912	0.42	0.18	0.675
Shipley score	-2.223	1.814	-1.226	1.44	0.231
Block:match	-6.678	7.499	-0.89	0.77	0.381
Block:semantic	5.628	5.368	1.048	1.09	0.296
RT:match	*48.689*	*24.065*	*2.023*	*3.81*	*0.051*
RT:semantic	-19.634	18.771	-1.046	1.09	0.297

Random effect structure

Experiment 1: decorrelated subject slopes for block, block:match, block:semantic; decorrelated item slopes for Shipley score, BLUPs from RT model, block:match, block:semantic

Experiment 2: correlated subject slopes for block and trial-level RT; correlated item slopes for Shipley score, BLUP from RT model, block:semantic

Experiment 3: decorrelated subject slopes for match, trial-level RT, block:match, RT:match, RT:semantic; decorrelated item slopes for Shipley score, trial-level RT, BLUPs from RT model, block:match

Table A12: Cross-experiment comparison models, vowel distance.

Experiment 1 vs. experiment 2	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	12.185	12.551	0.971	0.930	0.335
Block	**-5.528**	**1.555**	**-3.556**	**11.310**	**< 0.001**
Match status	**9.815**	**3.838**	**2.558**	**6.520**	**0.011**
Semantic relatedness	-1.865	3.138	-0.594	0.350	0.552
Trial-level RT	-2.759	7.447	-0.371	0.140	0.711
BLUP from RT model	64.059	71.047	0.902	0.790	0.373
Shipley score	0.665	1.032	0.644	0.410	0.524
Experiment:block	-3.809	3.094	-1.231	1.500	0.220
Experiment:match	*14.348*	*7.607*	*1.886*	*3.550*	*0.060*
Experiment:unrelated	1.797	6.264	0.287	0.080	0.774
Block:match	0.124	4.100	0.030	DNC	
Block:semantic	0.815	3.748	0.217	0.050	0.828
Experiment:RT	*-25.000*	*14.390*	*-1.737*	*3.010*	*0.083*
Match:RT	7.838	15.909	0.493	0.240	0.623
Unrelated:RT	-15.544	15.523	-1.001	1.000	0.317
Experiment:block:match	-2.511	8.205	-0.306	0.090	0.760
Experiment:block:unrelated	4.157	7.502	0.554	0.310	0.580
Experiment:RT:match	-5.968	31.438	-0.190	0.040	0.850
Experiment:RT:unrelated	36.171	30.972	1.168	1.360	0.243

Experiment 1 vs. experiment 3					
	Estimate (β)	Std. error	T-value	χ^2	P (chisq.)
Experiment	**35.565**	**16.073**	**2.213**	**4.650**	**0.031**
Block	-1.970	1.503	-1.311	1.710	0.191
Match status	4.667	4.148	1.125	0.126	0.261
Semantic relatedness	*-6.109*	*3.269*	*-1.869*	*3.490*	*0.062*
Trial-level RT	11.031	8.266	1.334	1.760	0.184
BLUP from RT model	34.905	44.080	0.792	0.620	0.431
Shipley score	-0.252	1.561	-0.161	0.030	0.872
Experiment:block	4.894	3.002	1.630	2.650	0.103
Experiment:match	-8.582	8.263	-1.039	1.080	0.300
Experiment:unrelated	4.055	6.537	0.620	0.380	0.536
Block:match	-6.824	4.234	-1.612	2.590	0.108
Block:semantic	1.552	3.808	0.407	0.100	0.684
Experiment:RT	13.024	14.516	0.897	0.790	0.373
Match:RT	**29.358**	**14.670**	**2.001**	**3.990**	**0.046**
Unrelated:RT	*-23.701*	*13.314*	*-1.780*	*3.160*	*0.076*
Experiment:block:match	-5.954	8.373	-0.711	0.500	0.478
Experiment:block:unrelated	7.513	7.532	0.998	0.990	0.319
Experiment:RT:match	9.251	29.330	0.315	0.100	0.753
Experiment:RT:unrelated	33.853	26.595	1.273	1.610	0.204

Random effect structure

Model comparing experiment 1 and experiment 2: Correlated random subject slopes for block; correlated random item slopes for Shipley score, experiment, BLUPs from RT model

Model comparing experiment 1 and experiment 3: Subject intercept; correlated random item slopes for Shipley score, experiment, trial-level RT, BLUPs from RT model

Appendix D: Condition means by experiment

Table A13. Mean and standard errors for each dependent variable, broken down by experiment and condition, are presented below. Note that while reaction time models were run with a log-transformed dependent variable, the raw values are presented below for easier interpretability.

Reaction time (ms)	*Experiment 1*		*Experiment 2*		*Experiment 3*	
	Mean	s.e.	Mean	s.e.	Mean	s.e.
competitor	730.828	(6.139)	682.087	(5.320)	903.763	(8.936)
match	613.776	(5.146)	549.004	(4.566)	776.781	(5.559)
unrelated	745.887	(6.187)	684.229	(4.971)	924.553	(8.914)

Word duration (ms)	*Experiment 1*		*Experiment 2*		*Experiment 3*	
	Mean	s.e.	Mean	s.e.	Mean	s.e.
competitor	376.807	(3.762)	323.857	(3.663)	397.651	(4.105)
match	370.838	(3.479)	324.491	(3.295)	393.569	(3.815)
unrelated	380.109	(3.801)	331.083	(3.532)	421.518	(4.501)

Initial consonant duration (ms)	*Experiment 1*		*Experiment 2*		*Experiment 3*	
	Mean	s.e.	Mean	s.e.	Mean	s.e.
competitor	59.042	(1.846)	45.812	(1.627)	58.542	(1.678)
match	54.423	(1.653)	42.996	(1.349)	56.902	(1.603)
unrelated	58.291	(1.832)	48.281	(1.574)	61.645	(1.859)

Vowel duration (ms)	*Experiment 1*		*Experiment 2*		*Experiment 3*	
	Mean	s.e.	Mean	s.e.	Mean	s.e.
competitor	156.903	(1.918)	141.885	(1.773)	163.598	(1.685)
match	153.908	(1.774)	136.296	(1.605)	158.709	(1.656)
unrelated	154.248	(1.947)	141.776	(1.817)	167.668	(1.923)

Vowel distance (Hz)	*Experiment 1*		*Experiment 2*		*Experiment 3*	
	Mean	s.e.	Mean	s.e.	Mean	s.e.
competitor	245.888	(5.223)	264.282	(5.349)	268.639	(4.635)
match	244.025	(4.768)	254.768	(5.004)	274.651	(3.774)
unrelated	245.834	(5.090)	257.240	(5.171)	285.792	(4.952)