

Word frequency effects in phonological retrieval in Modern Standard Arabic

Aljasser, Faisal ✉

Qassim University, Saudi Arabia (jasr@qu.edu.sa)

Received: 31 May 2019

Available Online: 3 September 2019

Revised: 3 July 2019

DOI: 10.5861/ijrsl.2019.4007

Accepted: 8 August 2019

ISSN: 2243-7754

Online ISSN: 2243-7762

OPEN ACCESS



Abstract

To produce a spoken word both a wholistic lexical representation, a “lemma” and its phonological representation, a “lexeme” are accessed. The retrieval of these components is influenced by various lexical factors. Word frequency has the most empirically established effect. This effect is assumed to arise at the phonological retrieval step of lexical access. The current study asked if this word frequency effect on phonological retrieval can also be evident in the context of a diglossic language, where two varieties of a language are used by native speakers. Native Arabic speakers were asked to shadow words that varied in their frequency in Modern Standard Arabic. Latency and accuracy rates showed that shadowing was facilitated by word frequency. Implication of these finding to theories of lexical access in speech production is discussed.

Keywords: word frequency; shadowing; speech production; lexical access; Arabic

Word frequency effects in phonological retrieval in Modern Standard Arabic

1. Introduction

Various lexical characteristics have been empirically shown to have some effects on lexical access in both spoken word recognition and spoken word production. These include Age of Acquisition (Cortese & Khanna, 2007), Phonotactic Probability (Vitevitch & Luce, 1998) and neighborhood density (Vitevitch, 2002). However, Word Frequency effect has been the most empirically established effect. That is, frequent words are processed faster and more accurately than infrequent ones and this effect may overshadow all other effects (Brybaert et al., 2011); but see Brown and Watson (1987).

The WF effect is so reliable to the extent that it has been shown in almost all tasks investigating lexical processing. This is true both in word recognition and word production. In word recognition for example, The Lexical Decision Task (LDT) (see Goldinger (1996) for full description) has been widely used to investigate the effect of different lexical (e.g. WF) and sub-lexical (e.g. phonotactic probability) variables on lexical processing. Using this task, early studies have shown that the effect of WF was very strong in both visual (Brybaert, Lange, & Wijnendaele, 2000; Whaley, 1978) and auditory (Taft & Hambly, 1986) word recognition.

Recently, Aljasser (in press) used the auditory LDT to investigate the effect of WF on the spoken word recognition in Modern Standard Arabic (MSA). His aim was to study the effect of WF on the spoken word recognition in the context of a diglossic language. Arabic language has a diglossic situation (see Ferguson (1959) for a discussion of diglossia) where native speakers learn two varieties of the language: an informal spoken Arabic vernacular (SAV) learned at home as a first language and a standard formal variety (i.e. MSA) that is learned later, typically at pre-school age. Using reaction times (RTs) and error rates of the auditory LDT, Aljasser found out a strong WF effect on the spoken word recognition in MSA. High frequency words in MSA were perceived faster and more accurately than low frequency ones.

In the present study, the question whether the WF effect also stands in the spoken word production of MSA is addressed. More specifically we ask whether WF effects on the phonological retrieval stage in speech production observed in other languages can be obtained in MSA.

The question of how WF affects lexical access in speech production has been and remains to be the subject of intense debate. Before we discuss the accounts of this debate, however, we need to shed light on the theoretical accounts of the stages involved in spoken word production and then how the locus of WF effects can be integrated within these accounts.

The most influential models of lexical access in speech production (e.g. Dell, 1986; Levelt, 1989; Levelt, Roelofs & Meyer, 1999) encompass three processes: Conceptualization-Formulation-Articulation. The conceptualization stage in speech production includes selecting the concept to be expressed. The concept is then formulated by selecting the lexical nodes that correspond to the concept and encoding them phonologically. The final stage entails converting the output of the formulation stage to articulatory movements. Moreover, there is a general agreement amongst these models (but see Caramazza, 1997) that there are two distinct layers of processing in lexical access; namely lemma selection or access (Levelt et al., 1999) and lexeme retrieval (i.e. accessing the phonological code or characteristics of the lemma or as Kittredge, Dell, Verkuilen, and Schwartz (2008) call it, phonological retrieval).

Caramazza (1997, p. 177) comments on this theory of lexical access before offering an alternative view by stating that

The dominant view is that lexical access involves at least two distinct stages of processing. The

first stage involves the selection of a semantically and syntactically specified lexical representation or lemma; the second stage involves the selection of its corresponding lexical-phonological representation or lexeme.

Supporting evidence for this two-step theory comes from research investigating the tip-of-the-tongue states (e.g. Burke, MacKay, Worthley, & Wade, 1991). Tip-of-the-tongue states are “those instances in which the fluent retrieval of a lexical item fails to occur. Typically, information regarding the meaning, gender, or syntactic class of the word may be accessible, but not the complete phonological form of the word” (Vitevitch & Sommers, 2003, p. 491). Following Kittredge et al. (2008), the terms Lemma-retrieval and phonological retrieval will be used throughout the current paper to refer to the two steps, respectively.

Dell’s (1986) model of speech production differs from the model of Levelt (1989) and Levelt et al. (1999) in the way the information flows between different levels. Dell’s model is interactive in that information flows between levels forward and backward to affect the mapping of the information at the other level. In Levelt’s model however, the levels are independent and serially ordered. In other words, the flow of information is only feedforward from the lemma level to phonological level without any backward flow.

The WF effect has been very widely documented in speech production using a variety of tasks. The discovery of the WF effect in speech production was pioneered by Oldfield and Wingfield (1965) using the picture naming task. In the picture naming task subjects are asked to name randomly presented pictures as quickly and as accurately as possible. This task is believed to tap into lexical access processes. Oldfield and Wingfield (1965) found that pictures of high-frequency words were named significantly faster than pictures of low-frequency ones. Moreover, WF has also been shown to influence the production of speech errors with low-frequency words resulting in more speech errors than high-frequency ones (e.g. Dell, 1990) and more tip-of-the -tongue states (Vitevitch & Sommers, 2003).

A hotly debated issue in the literature has been concerned with locating the locus of the WF effect. Does frequency effect arise in the conceptualization stage, the formulation stage (i.e. either of the two-step process of lexical access) or the articulation stage? Using a recognition task, where subjects, after being presented with a lexical concept, have only to verify the picture by providing “yes” or “no” answers (but not name it as in the picture naming task), the WF locus at the conceptualization stage was ruled out since there were no WF effects (Wingfield, 1968) (see also Jescheniak and Levelt (1994)). In other words, it was hypothesized that word-to-picture matching which is a representation of the conceptualization stage in speech production is not the locus of WF effects observed in the picture naming task.

Similarly, the seminal study by Jescheniak and Levelt (1994) provided evidence which arguably ruled out the articulation stage as the locus of the WF. Their rationale was that if the articulation stage was the locus of WF effect it should still be present in a delayed naming task. That is, after the delay lexical access would have taken place and any remaining WF effects would indicate that the articulation stage also contribute to the WF effects. Jescheniak and Levelt’s hypothesis was confirmed as frequency effects disappeared when the pictures were named with a delay. Based on this evidence, they concluded that frequency effects arise at the lexicalization stage with no substantial influence from the conceptualization or the articulation stages.

Having reached this conclusion, Jescheniak and Levelt (1994) decided to specifically determine the locus of frequency effects at the lexicalization stage (i.e. at what point, lemma-retrieval or phonological retrieval?). In other words, they attempted to locate the locus of WF within the framework of the two-step theory of lexical access. To this end, in their experiment 6, they used homophones in a translation task from English to Dutch. They reasoned that homophones such as *we* and *wee* share their lexeme but have different lemmas and therefore if the locus of WF is at the stage of lexeme access (i.e. phonological retrieval) low frequency homophones (e.g. *wee*) should be processed as efficiently as the high frequency twins (e.g. *we*) because they would inherit their frequency. Their results provided support to this hypothesis. Their English proficient Dutch subjects were faster in translating an English word into a Dutch low-frequency homophone that has a high frequency twin than

translating a low frequency non-homophone controls. Moreover, these low frequency homophones were actually translated as fast as their high frequency homophones and high frequency non-homophone controls. This was taken as evidence to conclude that the locus of WF effect is at the stage of phonological retrieval (see also Nickels (1995) for supporting Evidence from speech errors in aphasic subjects; but see Cuetos, Bonin, Alameda, and Caramazza (2010)). This conclusion was integrated into the theory of lexical access of Levelt et al. (1999).

2. The current study

How is this conclusion relevant to the current study? The current study aims to investigate the WF effect on phonological retrieval in speech production in MSA. Therefore, a task that taps into the second stage only (i.e. phonological retrieval) of lexical access is needed so that we can confidently attribute the WF effects to this step. Picture naming does not serve our purpose. It includes picture recognition and it necessarily requires the semantic lemma retrieval stage as well (Dell, Martin, & Schwartz, 2007; Schriefers, Meyer, & Levelt, 1990). On the other hand, word shadowing (also known as auditory word repetition) can perfectly fit the current study. In this task, subjects are required to repeat an auditorily-presented stimulus as quickly as possible. Unlike picture naming which requires step one of lexical access (lemma retrieval), shadowing does not require it because it is auditory input that activates the word and therefore the shadowing task starts at step two (phonological retrieval). In other words, the auditorily-presented word is first mapped into its phonological representation and then its output phonology is retrieved (see Nozari, Kittredge, Dell, and Schwartz (2010) for a detailed discussion of the difference between the lexical access processes involved in both tasks).

But does shadowing necessarily require step two of lexical access (i.e. phonological retrieval) (a lexical route)? In other words, couldn't it just be performed by bypassing lexical access altogether (a non-lexical route) (i.e. direct mapping of input phonology to output phonology similar to repeating non-words)? Comparing errors made in picture naming and shadowing data from aphasic subjects, Nozari et al. (2010) have concluded that the best model that can account for the data is a dual route model in which the lexical route and the non-lexical route are used together in word repetition (Hanley, Dell, Kay, & Baron, 2004).

The current study is neither particularly concerned with testing the accounts regarding which routes are used in word repetition (shadowing) nor does it dismiss the evidence that WF can also affect to some degree step one (lemma retrieval) (e.g. Kittredge et al., 2008). Rather, following Nozari et al. (2010) and based on the evidence reviewed, it is believed that WF effect on phonological retrieval in speech production is stronger. Therefore, the view taken in the current study is that if WF effects are observed in the shadowing task they are mainly, but not exclusively, arising in phonological retrieval.

2.1 Method

The shadowing task was used in the current experiment to investigate whether the effect of WF on diglossic Arabic language spoken word recognition also stands in spoken word production and more specifically on the phonological retrieval stage. In the shadowing task, subjects listen to an auditorily-presented stimulus item and their task is to repeat the item as quickly as possible. Depending on the purpose of the study, either response latency or error rate or both are collected and considered as the dependent variable(s). In the current study both were collected to better understand the WF effect on subjects' performance. As discussed above, this task is particularly suited for the current investigation and has widely been used to investigate the effect of different lexical factors on speech production in both normal children (e.g. Budd, Hanley, & Nozari, 2012), adults (e.g. Marslen-Wilson, 1985; Mitterer & Ernestus, 2008) and aphasic subjects (e.g. Nozari, Dell, & Schwartz, 2012).

2.2 Participants

Twenty-four native Arabic speakers, all students at an American university, took part voluntarily in the experiment. All participants were born and studied elementary/secondary school in their home countries (in Arabic). They were all male (mean age = 27 years). None of the participants reported a history of speech or

hearing problems.

2.3 Stimuli

Importantly, to create the stimuli items, a reliable frequency of occurrence wordlist of MSA was needed. This wordlist came from Ar ten ten (Arts, Belinkov, Habash, Kilgarriff, & Suchomel, 2014), an Arabic corpus that contains 5.8 billion words. The huge number of words this database has can provide reliable frequency counts. Equally important is the fact that a subset of Ar ten ten containing 115 million words was processed with MADA (Habash, Rambow, & Roth, 2009) which is a language-processing tool that uses a morphological analyzer for MSA to disambiguate undiacritized MSA words in context by reaching a preferred analysis. This tool helps in deciding between Arabic homographs when diacritics for short vowels and gemination are not used which is typical of most Arabic scripts.

A frequency wordlist containing 100000 words was purchased from Sketch Engine (Kilgarriff, Rychly, Smrz, & Tugwell, 2004) (<https://www.sketchengine.co.uk/>) which is the corpus manager holding the MADA processed 115-million-word subset. Fully vowelized Buckwalter transliteration (Buckwalter, 2002) was used for the transcription of the words in the frequency wordlist. While, the stimuli consisted of two lists of 24 words each. One list contained high frequency words (mean frequency = 113) and the other list contained low frequency ones (mean frequency = 3.3). The frequency counts reported here are given per-million calculated on the 115-million-word subset corpus.

To produce very well controlled stimuli items, both lists of 24 words were matched on a number of variables. These include word class and number of phonemes and syllables. Specifically, all words were nouns and had five phonemes and two syllables and the form /CVCVC/. The initial phonemes were also closely matched across the lists as initial phonemes have been shown to affect voice key measurements leading to difference in RTs (Kessler, Treiman, & Mullennix, 2002). Moreover, other variables that have been shown to affect spoken word production were also controlled. These include Phonotactic Probability, acoustic duration (i.e. duration of the word stimulus sound recording in millisecond), neighborhood density, and uniqueness point (i.e. the phoneme at which the word becomes the only possible word candidate (e.g. phoneme number four /ʕ/ in the word /ʃa:ʕir/ (poet). Furthermore, to encourage subjects to process the 48 stimuli items in an MSA mode rather than an SAV mode an important measure was taken. Namely, most of the stimuli words were either a pure MSA word (i.e. SAV uses a different word to express the same meaning; e.g. /kisaʔ/) or an MSA word that is pronounced differently in SAV either because it has an MSA only phoneme (e.g. /dʕ/ or /q/) or because of the difference in phonotactic constraints between the two varieties (e.g. / buhaʔr/). Table 1 shows the means and standard deviations for all aforementioned variables.

Table 1

Means and standard deviations of the various factors for high versus low frequency stimuli words

	Lexical characteristics of stimuli					
	Mean frequency <i>SD</i>	Mean duration (ms) <i>SD</i>	Mean ND <i>SD</i>	Mean phoneme PP <i>SD</i>	Mean biophone PP <i>SD</i>	Mean UP <i>SD</i>
High Frequency words (N=24)	113 103	526.2 48.9	2.9 1.5	1.47 0.16	1.03 0.013	4.9 0.88
Low Frequency words (N= 24)	3.3 3.6	534 74.3	3.3 1.6	1.44 0.18	1.03 0.017	4.7 0.53

Analysis of Variance (ANOVA) showed that WF for high frequency words was significantly greater than for low frequency words $p < .0001$. However, the analysis showed that the stimuli were equivalent in all other lexical characteristics. These characteristics were not significantly different $p > .05$ showing that all these lexical variables were well-controlled. The stimuli items were spoken in isolation and recorded by a male native Arabic speaker in an anechoic chamber using a high-quality microphone on to digital-audio-tape at a sampling rate of

44.1 kHz. The recordings were then saved as digital 16-bit files on a computer disk.

2.4 Procedure

An iMac computer running PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) was used for the experiment presentation and randomization. Participants were tested individually and one at a time. Each participant was seated in a computer booth equipped with headphones, a microphone connected to a PsyScope button box to record RTs. Prior to running the experiment, the instructions appeared on the computer screen in Arabic. Participants were instructed that they will listen to MSA words and that their task is to repeat these words as quickly and as accurately as possible. All participants' inquiries were answered prior to the start of the experiment.

Each trial started with the word مستعد "READY" appearing in the center of the computer screen for 500 ms to indicate the beginning of the trial. The participant was then presented with one of the randomly selected stimuli at a comfortable listening level over the headphones. The participants responded by repeating the word into the microphone as quickly and accurately as possible. When a response was made, the word مستعد "READY" in Arabic script appeared on the screen and the next item began. A response had to be detected for the next item to be presented. RT was measured from the onset of the stimulus file to the onset of the response. Responses were also recorded for error analysis.

Prior to the experimental trials, each participant received five practice trials. These trials were used to familiarize the participants with the task and were not included in the final data analysis.

3. Results

Latencies for correct responses measured from the stimulus onset and error rates were collected. An error was counted when any phonological segment in the verbal response of the participant was mispronounced. This was performed (as in Vitevitch and Luce (2005) by listening to subjects' responses and comparing them to a written transcription of the target words. T test on latencies was conducted and revealed that high frequency words were responded to significantly faster ($M=813\text{ms}$, $SD=68$) than low frequency ones ($M=845\text{ms}$, $SD=65$); $p < .0001$. Moreover, T test carried out on error rate showed that high frequency words were responded to significantly more accurately ($M=0.175\%$, $SD=0.857$) than low frequency ones ($M=2.45\%$, $SD=3.013$); $p < .0001$.

3.1 Discussion

The current study aimed to investigate whether WF can affect phonological retrieval in MSA. A shadowing task showed a strong effect of WF. Native Arabic speakers were faster and more accurate in shadowing high frequency words than low frequency ones. As have been discussed, the shadowing task is assumed to originate at the second step of lexical access (i.e. phonological retrieval). Therefore, it is reasonable to conclude that these WF effects have mainly arisen at this step. This is particularly plausible given that all possible confounding variables were controlled. These included word class, number of phonemes and syllables, duration, initial phonemes, phonotactic probability, neighborhood density and uniqueness point.

Given the empirical evidence from other studies (Nozari et al., 2012; Nozari et al., 2010), the current findings, although not particularly designed to test this, provide support to a lexical route of speech shadowing in MSA. In shadowing single words, our subjects are clearly accessing the phonological representations of these words in memory and then transfer them to output phonology. In other words, they are not bypassing lexical access altogether (a non-lexical route) by directly mapping input phonology to output phonology. Had our native speakers used a non-lexical route to repeating MSA words no WF effects should have been observed. Moreover, our error rate data provide interesting results. Although our subjects made generally few errors in both conditions, as expected in a shadowing task, all errors were phonologically similar to the target. More interestingly, all errors

resulted in real words rather than non-words. The latter result is what is often referred to as the “lexical bias effect” (Baars, Motley, & MacKay, 1975; Nooteboom & Quené, 2008; Nozari & Dell, 2009).

Models of speech production can account for this effect in different ways. Recall that whereas the model proposed by Levelt et al. (1999) contains serially ordered independent stages with feedforward only flow of information, Dell (1986) proposes an interactive model with bidirectional flow of information between the lemma level and the phonological level. In the serial model, the lexical bias effect can be accounted for by proposing the existence of an editor in the production system which checks for lexicality status in the output of the formulation stage (Baars et al., 1975; Levelt, 1989; Levelt et al., 1999). On the other hand, in the interactive model the lexical bias effect arises as a result of the interactive feedback between the phoneme level and the lexeme level. Although the “lexical editor” account is theoretically plausible, experimental evidence shows that it may not be as fast and efficient as needed in fluent speech (Nazbanou & Dell, 2009).

In addition, a serial model may fail to convincingly account for the other current finding that most of the error (14 out of 15 errors) were phonologically similar. Specifically, the errors resulted from the substitution of a phoneme in the target word with another similar phoneme that is only different by one phonetic feature (e.g. the target /h/ in /kaahil/ is substituted with /ħ/ resulting in the word /kaaħil/). This phonological similarity effect has usually been observed in errors elicited using the “tongue twister” paradigm (e.g. Shattuck-Hufnagel, 1992; Wilshire, 1999). Again, the serial model attempts to account for this effect using a monitoring mechanism. In this model, phonologically similar errors are likely to be unnoticed by the monitor.

The interactive model, however, is providing a more empirically supported account. Namely, the phonological similarity effect is explained in terms of increased activation and competition between similar phonemes as a result of feedback from the sub-phonemic level. This strong competition between similar phonemes can in turn increase the likelihood of phonologically similar errors.

In addition, a more careful look at the error data reveals another interesting result. That is, the phoneme /d^ɟ/ caused the highest number of substitution errors (six out of fifteen). Why has this particular phoneme been involved in more errors than any other phoneme? A potential cause lies within the distribution of this phoneme. Most spoken Arabic vernaculars do not have this sound. Specifically, it is not available in the SAV that our subjects speak which is commonly referred to as Gulf Arabic. Therefore, this sound is learned only as an MSA sound. In second language learning, it is well attested that L2 phonemes not available in L1 can impair the recognition of L2 words (e.g. Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Cutler & Otake, 2004; Cutler, Weber, & Otake, 2006). The same can also lead to more speech errors in L2 word production (Gollan & Goldrick, 2012; e.g. Aljasser, Jackson, Vitevitch, & Sereno, 2018; Li, Goldrick, & Gollan, 2017).

Recently, Aljasser et al. (2018) have found that L2 English minimal word pairs containing phonemic contrasts that are not available in L1 Arabic (e.g. p-b pack back) induced more errors in tongue twisters than those available in L1 Arabic (e.g. t-d tuck duck). They attributed this finding to weak sub-lexical representation of the phonemes that do not exist in Arabic (see also Li, Goldrick and Gollan (2017) for similar conclusions from Spanish and Mandarin speakers). This weak sub-lexical representation is claimed to be mainly due to a frequency-lag of the phonemes not found in the native language (Li et al., 2017). In the current study, the finding that the MSA phoneme which does not exist in the SAV produced more substitution errors may tentatively point to similar conclusions about weak sub-lexical representations of MSA phonemes.

4. Conclusion

The current study provided evidence of strong WF effects in speech production in the formal variety of a diglossic language (i.e. MSA). The findings also indicate that phonemes only available in MSA and not in SAV may be weakly represented. However, further carefully controlled speech error studies comparing native Arabic speakers' performance in MSA and SAV are required. If such studies show similar findings (i.e. non SAV phonemes causing more errors) then this might be further evidence that at least at the sub-lexical level MSA

shows signs of processing inefficiency comparable to an L2 (e.g. Ibrahim & Aharon-Peretz, 2005).

Acknowledgment: I am greatly indebted to Professor Michael Vitevitch for granting full access to his Spoken Language Laboratory during data collection of the present study.

5. References

- Aljasser, F. (in press). Frequency effects in auditory lexical decision in Arabic. *Journal of Arabic and Human Sciences*.
- Aljasser, F. M., Jackson, K. T., Vitevitch, M. S., & Sereno, J. A. (2018). The influence of phoneme inventory on elicited speech errors in Arabic speakers of English. *The Mental Lexicon*, 13(1), 26–37. <https://doi.org/10.1075/ml.17014.alj>
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, 32(2), 233–250. [https://doi.org/10.1016/S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)
- Arts, T., Belinkov, Y., Habash, N., Kilgarriff, A., & Suchomel, V. (2014). ArTenTen: Arabic corpus and word sketches. *Journal of King Saud University - Computer and Information Sciences*, 26(4), 357–371. <https://doi.org/10.1016/j.jksuci.2014.06.009>
- Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, 14(4), 382–391. [https://doi.org/10.1016/S0022-5371\(75\)80017-X](https://doi.org/10.1016/S0022-5371(75)80017-X)
- Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, 15(3), 208–216. <https://doi.org/10.3758/BF03197718>
- Brybaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. <https://doi.org/10.1027/1618-3169/a000123>
- Brybaert, M., Lange, M., & Wijnendaele, I. Van. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, 12(1), 65–85. <https://doi.org/10.1080/095414400382208>
- Buckwalter, T. (2002). Arabic transliteration. Retrieved from <http://www.qamus.org/transliteration.htm>
- Budd, M. J., Hanley, J. R., & Nozari, N. (2012). Evidence for a non-lexical influence on children's auditory repetition of familiar words. *Journal of Psycholinguistic Research*, 41(4), 253–266. <https://doi.org/10.1007/s10936-011-9189-8>
- 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. [https://doi.org/https://doi.org/10.1016/0749-596X\(91\)90026-G](https://doi.org/https://doi.org/10.1016/0749-596X(91)90026-G)
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14(1), 177–208. <https://doi.org/10.1080/026432997381664>
- Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25(2), 257–271. <https://doi.org/10.3758/BF03204507>
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, 60(8), 1072–1082. <https://doi.org/10.1080/17470210701315467>
- Cuetos, F., Bonin, P., Alameda, J. R., & Caramazza, A. (2010). The specific-word frequency effect in speech production: Evidence from Spanish and French. *The Quarterly Journal of Experimental Psychology*, 63(4), 750–771. <https://doi.org/10.1080/17470210903121663>

- Cutler, A., & Otake, T. (2004). Pseudo-homophony in non-native listening. *The Journal of the Acoustical Society of America*, *115*(5), 2392–2392. <https://doi.org/10.1121/1.4780547>
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. <https://doi.org/10.1016/j.wocn.2005.06.002>
- Dell, G. (1986). A spreading activation model of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349. <https://doi.org/10.1080/01690969008407066>
- Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language*, *56*(4), 490–520. <https://doi.org/10.1016/j.jml.2006.05.007>
- Ferguson, C. A. (1959). Diglossia. *Word*, *15*, 325–340. <https://doi.org/10.1080/00437956.1959.11659702>
- Goldinger, S. D. (1996). Auditory Lexical Decision. *Language and Cognitive Processes*, *11*(6), 559–568. <https://doi.org/10.1080/016909696386944>
- Gollan, T. H., & Goldrick, M. (2012). Does bilingualism twist your tongue? *Cognition*, *125*(3), 491–497. <https://doi.org/10.1016/j.cognition.2012.08.002>
- Habash, N., Rambow, O., & Roth, R. (2009). MADA+TOKAN: A Toolkit for Arabic Tokenization, Diacritization, Morphological Disambiguation, POS Tagging, Stemming and Lemmatization. In the *Proceedings of the second international conference on Arabic language resources and tools* (pp. 102–109). Retrieved from <https://doi.org/10.1080/016909696386944>
- Hanley, R., Dell, G., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words: A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology*, *21*(2–4), 147–158.
- Ibrahim, R., & Aharon-Peretz, J. (2005). Is literary Arabic a second language for native Arab speakers?: Evidence from semantic priming study. *Journal of Psycholinguistic Research*, *34*(1), 51–70. <https://doi.org/10.1007/s10936-005-3631-8>
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843. <https://doi.org/10.1037/0278-7393.20.4.824>
- Kessler, B., Treiman, R., & Mullenix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language*, *47*(1), 145–171. <https://doi.org/10.1006/jmla.2001.2835>
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The sketch engine. In *Proceedings of Euralex* (pp. 105–116). Lorient.
- Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492. <https://doi.org/10.1080/02643290701674851>
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38.
- Li, C., Goldrick, M., & Gollan, T. H. (2017). Bilinguals' twisted tongues: Frequency lag or interference? *Memory and Cognition*, *45*(4), 600–610. <https://doi.org/10.3758/s13421-017-0688-1>
- Marslen-Wilson, W. D. (1985). Speech Shadowing and Speech Comprehension. *Speech Communication*, *4*, 55–73.
- Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. <https://doi.org/10.1016/j.cognition.2008.08.002>
- Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Language and Cognitive Processes*, *10*(1), 13–45. <https://doi.org/10.1080/01690969508407086>

- Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language*, 58(3), 837–861. <https://doi.org/10.1016/j.jml.2007.05.003>
- Nozari, N., & Dell, G. S. (2009). More on lexical bias: How efficient can a “lexical editor” be? *Journal of Memory and Language*, 60(2), 291–307. <https://doi.org/10.1016/j.jml.2008.09.006>
- Nozari, N., Dell, G., & Schwartz, M. (2012). Who are the lexical-routers? An investigation into the nature of word repetition in aphasia. *Procedia - Social and Behavioral Sciences*, 61, 104–105. <https://doi.org/10.1016/j.sbspro.2012.10.100>
- Nozari, N., Kittredge, A. K., Dell, G. S., & Schwartz, M. F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. *Journal of Memory and Language*, 63(4), 541–559. <https://doi.org/10.1016/j.jml.2010.08.001>
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology*, 17(4), 273–281.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29(1), 86–102. [https://doi.org/10.1016/0749-596X\(90\)90011-N](https://doi.org/10.1016/0749-596X(90)90011-N)
- Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42(1–3), 213–259. [https://doi.org/10.1016/0010-0277\(92\)90044-I](https://doi.org/10.1016/0010-0277(92)90044-I)
- Taft, M., & Hambly, G. (1986). Exploring the Cohort Model of spoken word recognition. *Cognition*, 22, 259–282.
- Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 735–747.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325–329. <https://doi.org/10.1111/1467-9280.00064>
- Vitevitch, M. S., & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, 52(2), 193–204. <https://doi.org/10.1016/j.jml.2004.10.003>
- Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory and Cognition*, 31(4), 491–504. <https://doi.org/10.3758/BF03196091>
- Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17(2), 143–154. [https://doi.org/10.1016/S0022-5371\(78\)90110-X](https://doi.org/10.1016/S0022-5371(78)90110-X)
- Wilshire, C. E. (1999). The “tongue twister” paradigm as a technique for studying phonological encoding. *Language and Speech*, 42(1), 57–82. <https://doi.org/10.1177/00238309990420010301>
- Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *The American Journal of Psychology*, 81(2), 226–34.