Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2015 January ; 41(1): 77–94. doi:10.1037/a0037778.

Friends *and* Foes in the Lexicon: Homophone Naming in Aphasia

Erica L. Middleton¹, Qi Chen^{1,2}, and Jay Verkuilen³

¹Moss Rehabilitation Research Institute

²Center for Studies of Psychological Application and School of Psychology, South China Normal University

³City University of New York

Abstract

The study of homophones—words with different meanings that sound the same—has great potential to inform models of language production. Of particular relevance is a phenomenon termed *frequency inheritance*, where a low frequency word (e.g., deer) is produced more fluently than would be expected based on its frequency characteristics, presumably because of shared phonology with a high-frequency homophone counterpart (e.g., dear). However, prior studies have been inconsistent in showing frequency inheritance. To explain this inconsistency, we propose a *dual nature account of homophony*: a high-frequency counterpart exerts two counterposing effects on a low frequency homophone target during the two main stages of naming: (1) a detrimental impact during semantically-driven lexical retrieval; (2) a beneficial impact during phonological retrieval. In a study of naming in participants with chronic aphasia followed by computational investigations, we find strong evidence for the dual nature account of homophony.

Keywords

homophone; naming; lexical access; aphasia; language production; word frequency

Homophones are pairs of words that have distinct meanings (and possibly distinct grammatical categories and spelling), but sound the same. In the study of spoken word production, much research has focused on determining how homophones are represented and accessed (Anton-Mendez, Schuetze, Champion, & Gollan, 2012; Biedermann, Coltheart, Nicekls, & Saunders, 2009; Biedermann & Nickels, 2008a; 2008b; Bonin & Fayol, 2002; Burke, Locantore, Austin, & Chae, 2004; Caramazza, Bi, Costa, & Miozzo, 2004; Caramazza, Costa, Miozzo, & Bi, 2001; Cueto, Bonin, Ramon Alameda, & Caramazza, 2010; Cutting & Ferreira, 1999; Dell, 1990; Ferreira & Griffin, 2003; Gahl,

Correspondence concerning this article can be addressed to either Erica Middleton, Research Department, Moss Rehabilitation Research Institute, 50 Township Line Rd., Elkins Park, PA, 19027. middleer@gmail.com; or Qi Chen, Center for Studies of Psychological Application and School of Psychology, South China Normal University, South China Normal University, 510631 Guangzhou, P. R. China. chenqi.research@gmail.com.

Erica L. Middleton, Research Department, Moss Rehabilitation Research Institute; Qi Chen, Research Department, Moss Rehabilitation Research Institute (now at South China Normal University); Jay Verkuilen, Educational Psychology Program and Center for Advanced Study in Education, City University of New York.

2008; Jescheniak & Levelt, 1994; Jescheniak, Meyer, & Levelt, 2003; Miozzo & Caramazza, 2005; Miozzo, Jacobs, & Singer, 2004; Shatzman & Schiller, 2004). A primary goal of this work has been to delineate whether and how having a homophonic counterpart impacts how a word is retrieved. These studies have focused on a phenomenon termed *frequency inheritance*, which is where low frequency words appear to “inherit” fluency and/or resistance to error from their high frequency homophonic counterparts (Dell, 1990; Jescheniak & Levelt, 1994; Jescheniak et al., 2003).

The possible special status of homophones with regard to frequency inheritance has received much attention because of its potential to inform models of speech production. The dominant view of how speakers produce a word from meaning (i.e., naming) is that it involves at least two stages of retrieval: (1) mapping from semantics to an intermediary lexical representation(s); (2) retrieval of a word’s phonological constituents, (for reviews, see Rapp & Goldrick, 2000; Vigliocco & Hartsuiker, 2002). Though most models of naming adopt this general framework, long-standing controversies surround how to characterize various key aspects within the framework. One topic of debate for which the study of homophones is particularly relevant concerns the nature of the lexical representations that mediate between semantics and phonological constituents. According to numerous accounts of naming, meaning first maps to a *lemma*, a modality-neutral representation that codes a word’s specific array of lexical and grammatical features (e.g., Dell, 1986; Kempen & Huijibers, 1983; Levelt, Roelofs, & Meyer, 1999). In one such model (WEAVER++; Levelt et al., 1999), retrieval of a word’s lemma is followed by retrieval of its *lexeme*, which may be thought of as a holistic form-based lexical representation specific to the modality of output (i.e., phonological or orthographic). In WEAVER++, homophonous words are proposed to share a lexeme, with their distinct meanings represented with distinct lemmas. Frequency inheritance in homophones, coupled with an assumption that frequency impacts word form access in production, follows directly from such a view: a low frequency homophone will experience benefit because it shares a lexeme with a high frequency homophonic counterpart. Note, in WEAVER++, the prediction of frequency inheritance hinges on the assumption that frequency impacts word form retrieval. Though evidence points to phonological form retrieval as the predominant locus of frequency effects (Kittredge et al., 2008; Jescheniak & Levelt, 1994) it is possible that frequency could also impact lemma retrieval (for evidence see Kittredge et al., 2008), in which case WEAVER++ would expect partial frequency inheritance. Either way, what is important is that this theoretical framework predicts some degree of frequency inheritance, which would implicate shared phonological form of homophones.

However, a number of researchers have questioned the theoretical and empirical grounds for invoking a modality-neutral stage of lexical retrieval (Caramazza & Miozzo, 1997; Caramazza, 1997; Harley, 1999). An alternative is that semantics maps directly to modality-specific lexical form representations, and these representations license access to a word’s lexical-grammatical information (Independent Network Model; Caramazza, 1997; Caramazza & Miozzo, 1997). This account predicts no frequency inheritance in that the distinct lexical-grammatical properties of the different meanings of a homophone are associated with distinct form representations.¹

Thus, the study of homophones, and frequency inheritance in particular, has great potential for informing models of speech production. Unfortunately, however, the empirical literature regarding frequency inheritance (reviewed below) has been perplexing, with many contradictory results. In an investigation of homophone naming in chronic aphasia, we revisit the issue of frequency inheritance and explore factors that may mitigate its impact. We provide evidence for an explanatory framework that sheds light on the host of contradictory results regarding frequency inheritance.

Frequency Inheritance in Homophone Production

An initial demonstration of frequency inheritance was reported by Dell (1990), who found that low-frequency open class words (e.g., *witch*) were no more vulnerable to experimentally-induced phonological errors than high frequency homophonic function words (e.g., *which*). Dell provided additional evidence for frequency inheritance by showing in a regression analysis that the frequency of the high frequency homophonic counterparts predicted error rates on the low-frequency homophones.

Jescheniak and Levelt (1994) studied frequency inheritance in bilinguals, using an English to Dutch translation task, where the Dutch words were homophones or control words without homophonic counterparts. Jescheniak and Levelt compared translation times for three word types: (1) low frequency words that had a high frequency homophone counterpart (hereafter, *homophone targets*); (2) high frequency words (without homophonic counterparts) matched to the summed frequency of the homophone targets and their counterparts (*high frequency* or *HF* controls); (3) low frequency words (also without homophonic counterparts) matched to the low frequency meaning of the homophone targets (*low frequency* or *LF* controls). While correcting for any differences in semantic processing across words, Jescheniak and Levelt found equal translation times for the homophone targets and the HF controls; and, both sets were faster than the LF controls. In follow-up studies, Jescheniak et al. (2003) replicated the finding of faster production of homophone targets compared to LF controls in an English-Dutch translation task as well in an English-German translation task.

A number of studies of naming treatment in aphasia have reported a different kind of inheritance effect with homophones (Biedermann, Blanken, & Nickels, 2002; Biedermann & Nickels, 2008a, 2008b). Across these studies, phonological-based treatment of one meaning of a homophone generalized to improved naming of an untreated homophonic counterpart. Biedermann and colleagues were able to rule out an articulatory basis of the generalization because their participants could accurately repeat the target forms. Furthermore, untreated control words that were only phonologically related to treated items did not experience benefit, suggesting the basis for the generalization did not arise from improvement in the production of overlapping but nonidentical segmental sequences. Biedermann et al. concluded the treatment worked by facilitating retrieval of a shared phonological form between distinct homophone meanings.

¹As Caramazza et al. (2001) note, the Independent Network Model as formulated in Caramazza (1997) predicts no frequency inheritance. However, Caramazza et al. described a modified framework permitting interactive activation between lexemes and constituent phonemes, which may be compatible with an observation of frequency inheritance.

The studies reviewed thus far point to a processing advantage conferred on low frequency homophones with high frequency counterparts. However, the literature on homophone production also features a number of failures to replicate the frequency inheritance effect. In English (Experiment 1a), Mandarin-Chinese (Experiment 2a) and in a Spanish-English translation task (Experiment 3a), Caramazza et al. (2001) found that latencies of the homophone targets were slower than HF controls, and that they patterned similarly to LF controls. However, in a critique of this work, Jescheniak et al. (2003) cited a number of potential methodological shortcomings, such as possible power issues and differences in object recognition difficulty between conditions. Yet Cuetos et al. (2010) addressed these issues in a series of studies and found no frequency inheritance in two naming studies in Spanish and one in French. Specifically, low frequency and high frequency homophone counterparts differed in naming times and patterned similarly to LF and HF control conditions, respectively (for related findings, see Anton-Mendez et al., 2012; Bonin & Fayol, 2002; Schiller & Schatzman, 2004). In a related study, Miozzo et al., (2004) studied homophone naming from pictures and descriptions in an individual with stroke aphasia who demonstrated a word-finding impairment. For this individual, accuracy on (low frequency) homophone targets was similar to LF controls but less than HF controls and high frequency homophonic counterparts. Lastly, in a speech-corpus analysis by Gahl (2008), high frequency homophones (e.g., time) were shorter in duration than their low frequency mates (e.g., thyme). However, Gahl noted as a caveat that the pattern could be consistent with partial inheritance as LF controls were not included in the study.

The Dual Nature of Homophony

This review has shown that the study of frequency inheritance in homophone production comprises many contradictory findings—from partial or full inheritance (e.g., Jescheniak & Levelt, 1994; Jescheniak et al., 2003) to a complete absence of frequency inheritance (e.g., Caramazza et al., 2001; Cuetos et al., 2010). In trying to make sense of this inconsistency, a common strategy in prior work has been to focus on potential methodological weaknesses or differences between studies. However, an intriguing possibility is that the inconsistency points to a needed shift in how the theoretical debate surrounding frequency inheritance has been framed. The naming models discussed in studies on frequency inheritance have predominantly assumed feedforward spread of activation, such as the Independent Network model of Caramazza (1997) and the WEAVER++ model of Levelt et al. (1999; c.f., Dell, 1990). However, a different class of models advocates *interactive activation* during naming (e.g., Dell, 1986; 1990; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Dell & Gordon, 2003; Foygel & Dell, 2000; Harley, 1993; Rapp & Goldrick, 2000; Schwartz, Dell, Martin, Gahl, & Sobel, 2006; Stemberger, 1985), where information flow from later stages feeds back to influence the outcome of earlier stages of processing. We propose that because of interactive activation, high frequency counterparts exert both a beneficial and a deleterious impact on homophonic targets, depending on the stage of retrieval. We suggest that the literature is replete with contradictory results because of the operation of these two counterposing effects, and that the conditions that legislate their relative strengths have not been controlled. The goal of this study is to provide evidence for both a beneficial and a detrimental impact of homophony in naming (i.e., the *dual nature account of homophone*

naming), as well as to reveal factors that influence the relative strength of the two effects in determining homophony's net impact.

The model of naming we adopt bears large similarity to the two-step interactive theory of lexical access (Foygel & Dell, 2000; Schwartz et al., 2006). In the current framework, semantics maps to an intermediate lexical (i.e., word node) layer (hereafter, Stage-1 retrieval), which in turn maps to phonological constituents (hereafter, Stage-2 retrieval); and, information flow is interactive. Similar to many other models of naming (e.g., Bloem & La Heij, 2003; Caramazza, 1997; Cutting & Ferreira, 1999; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Levelt et al., 1999), we assume Stage-1 selection is competitive. Following Foygel and Dell (2000; also Dell et al., 2004; Schwartz et al., 2006), we assume the two stages of retrieval can be separately damaged (i.e., *lesioned*). Within such a framework, we propose that a high frequency counterpart exerts two influences when a low frequency homophonic mate is to be named: a deleterious effect on retrieving the correct word node during Stage-1 retrieval, but a facilitative effect on retrieving correct constituent phonology during Stage-2 retrieval. The rationale in postulating a negative effect of homophony is an expectation that, because of interactive activation and identical phonology with the target, the relatively high frequency homophonic counterpart can interfere with selection of the target during Stage-1 retrieval. However, these same conditions are also expected to confer a benefit during Stage-2 retrieval, instantiated as facilitated selection of the constituent phonemes of a target because of reverberated activation from the high frequency counterpart's word node.

To evaluate this hypothesis, the strategy was to measure homophony's effect when either stage of retrieval was selectively damaged, the expectation being that dysfunction at a stage would facilitate detection of the negative and positive effects of homophony. Participants with chronic aphasia from left-hemisphere stroke were recruited to comprise two groups: in the *Stage-1 group*, the neuropsychological profile of participants was consistent with a naming impairment in selecting words from semantics; the profile of participants in the *Stage-2 group* was consistent with dysfunction in retrieving phonology. The impact of homophony on naming was measured using the experimental design of Jescheniak and Levelt (1994). While controlling for a large number of psycholinguistic variables, three picture sets were used to elicit nouns belonging to the following conditions: (1) *homophone targets*, i.e., low frequency homophones with high frequency counterparts; (2) *LF controls*, i.e., words matched in frequency to the depicted meanings of the homophones; (3) *HF controls*, i.e., words matched to the average summed frequency of the homophone targets and their counterparts.

The expectation is that in the Stage-1 group, dysfunction in Stage-1 retrieval will exaggerate vulnerability to the negative impact of homophony, diminishing homophony's net benefit relative to the Stage-2 group whose Stage-1 processing is generally intact. Thus, we expected that the advantage for the homophone targets over LF controls would be greater in the Stage-2 group compared to the Stage-1 group; and, that the decrement in accuracy for the homophone targets compared to the HF controls would be greater in the Stage-1 group compared to the Stage-2 group. Both patterns would point to less frequency inheritance in the Stage-1 group. To evaluate these predictions, we inspected each of two individual two-

way interactions of participant group by word type, where word type included (1) homophones versus LF controls; and (2) homophones versus HF controls. The behavioral study is followed by results from a series of simulations designed to implement and evaluate the dual nature account of homophone naming in a computational framework.

Method

Participants

Participants were recruited from the Moss Rehabilitation Research Institute Participant Registry. All participants were right-handed native English speakers with chronic aphasia secondary to left-hemisphere stroke.² They gave informed consent under a protocol approved by the Institutional Review Board of Albert Einstein Healthcare Network (Philadelphia, PA). Participants were paid \$15 for each session of participation.

The goal of recruitment for this study was to develop two groups of participants with naming impairments that were maximally distinct in implicating either Stage-1 or Stage-2 dysfunction. This required ongoing enrollment and testing of naming ability and other language-based functions (relevant tests are listed in Tables 1 and 2). Through this process two participants were enrolled but subsequently excluded because they could not be classified unambiguously into either participant group: their predominant naming error category was omissions not unlike the Stage-1 group (Table 1, top panel), but in contrast to the Stage-1 group they produced few semantic errors. However, their inclusion in the Stage-2 group was disqualified because of little evidence of Stage-2 impairment (i.e., they had good repetition and produced very few phonological naming errors). A third participant was excluded subsequent to participation because his performance was at floor in the main task (i.e., only produced one correct response).

Results will be presented from eleven participants (4 female): five in the Stage-1 group, and six in the Stage-2 group, characterized in Tables 1 and 2. Z-scores were computed relative to a reference sample of 107 individuals with wide-ranging aphasia (unselected for type) from left hemisphere stroke, whose behavioral data was available in the Moss Aphasia Psycholinguistics Project (Mirman, Strauss, Brecher, Walker & Sobel et al., 2010; www.mappd.org/) and for whom there was also CT or MRI confirmation of left hemisphere stroke.

Table 1 lists performance on the 175-item Philadelphia Naming Test (PNT; Roach, Schwartz, Martin, Grewal, & Brecher, 1996) including proportions out of total responses of semantic errors (i.e., semantically related word substitutions, e.g., dog→cat), phonological errors (i.e., word or nonword error outcomes that are phonologically related but semantically unrelated to a target, e.g., dog→log), and omissions (i.e., failure to provide a naming attempt such as personal comments, no responses, or descriptions). Table 2 lists months post onset, Western Aphasia Battery (Kertesz, 1982) aphasia quotient, aphasia classification, word repetition ability, and performance on measures of word comprehension and nonverbal semantics. The word comprehension tests include the semantic subtest of the PNT Picture

²The clinical scan of one participant (P1) showed evidence of bilateral damage.

Naming Verification Test (i.e., participant decides whether the word matches the picture; semantically related foils are possible; Mirman et al., 2010) and Synonymy Triplets (i.e., pick two of three words that are synonyms; Martin, Schwartz, & Kohen, 2006). Nonverbal semantics was assessed by the picture version of the Camel and Cactus Test (Does camel go best with cactus, tree, sunflower, rose?; Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000).

Table 1 shows that the participants in the Stage-1 group produced a very high proportion of semantic errors in naming compared to the reference sample (average $z = 1.84$) and the Stage-2 group. Semantic errors are the traditional error type associated with difficulty in retrieving words from semantics. These participants also produced a high proportion of omissions, accounting for a majority of their errors (group average = 55%). Though omissions can originate from disrupted access to output phonology as seen in instances of *tip-of-the-tongue* in aphasia (e.g., Badecker, Miozzo, & Zanittini, 1995; Beeson, Holland, & Murray, 1997; Goodglass, Kaplan, & Weintraub, 1976), we interpret the high rate of omissions in this group as indicative of Stage-1 rather than Stage-2 naming impairment for a number of reasons. Firstly, it has been noted that omissions tend to accompany semantic errors in large case series investigations of naming impairments in chronic aphasia (Dell et al., 1997; Lambon Ralph, Sage, & Roberts, 2002; Schwartz et al., 2006) and that they localize similarly (i.e., anterior temporal lobe) in lesion-symptom mapping (Chen, Graziano, Middleton, & Mirman, 2013). These associations provide evidence for a common mechanism underlying the two errors types. Furthermore, Dell et al. (2004) provided evidence in a computational investigation that was consistent with the notion that omission errors in aphasia are lexically-generated, resonating with our characterization here. On the flipside, there was little evidence the omissions in the Stage-1 group originated from disrupted access to output phonology because the participants in the Stage-1 group showed only minor or nonexistent phonological output problems: the proportion of phonologically related naming errors was low (average 6%) and less than the mean of the reference sample (average $z = -0.50$); and, word repetition was high (average 94% accuracy), exceeding the mean of the reference sample (average $z = 0.52$).

Commensurate with other studies suggesting that Stage-1 impairment is associated with semantic deficits (Hanley & Nickels, 2009; Nozari & Dell, 2013), the participants in the Stage-1 group also tended to show impaired performance on tests of word comprehension and nonverbal semantic comprehension (Table 2). When semantic errors in comprehension accompany semantic errors in naming, this can suggest that deficient semantic representations and/or disruption in the access to such representations underlies both problems (Gainotti, Miceli, Caltagirone, Silveri, & Masullo, 1981; Hart & Gordon, 1990; Hillis, Rapp, Romani, & Caramazza, 1990; Jefferies & Lambon Ralph, 2006; Rapp & Goldrick, 2000). For our purposes, what is most important is that semantically-driven retrieval of words is impaired, be it from deficient input from semantics, weak connections from semantics to words, or a combination of both.

Compared to the Stage-1 group and the reference sample, the Stage-2 group demonstrated good word comprehension (average $z = 0.58$ and 0.51 for PNV and Synonym Triplets, respectively) and nonverbal semantic comprehension (average $z = 0.33$). On the PNT,

semantic errors and omissions were relatively rare and lower than the reference sample (average $z = -0.48$ and -0.39 for semantic errors and omissions, respectively) and the Stage-1 group. Thus, overall, participants in the Stage-2 group may be described as demonstrating good word comprehension and core semantics with little evidence of Stage-1 naming impairment.

Phonological errors were relatively high (average z -score = 1.35) and accounted for the majority of naming errors (average 67%) in the Stage-2 group. Of note, the category of phonological errors included both nonwords and words (i.e., *formals*) that were phonologically related to the target. Such nonword errors are rather uncontroversially understood to originate from faulty phonological processes. However, in addition to a phonological locus, formals can also arise as errors in word selection, e.g., due to shared phonology with a target in an interactive framework (Gagnon, Schwartz, Martin, Dell, & Saffran, 1997; Martin & Saffran, 1992). Yet, there is little evidence the formals produced by the Stage-2 group were word substitution errors: formals only accounted for an average of 31% of their phonological errors; and, this is no higher than various estimates taken from error corpora of the likelihood of a phonological error resulting in a word by chance alone (e.g., 33% for onset exchange errors in Garrett, 1976; >36% for exchanges, anticipations and perseverations in Dell and Reich, 1981). Relative to the reference sample, participants in the Stage-2 group tended to show impaired word repetition (average $z = -0.54$), providing additional evidence for a phonological disruption in naming.

Materials and Design

All picture stimuli in the main experiment depicted common objects, thus requiring nouns in the naming task. The pictures were an assortment of black-and-white as well as color pictures taken from Mirman, Strauss, Dixon, and Magnuson (2010), Magnuson, Dixon, Tanenhaus, and Aslin, (2007), a colorized version (Rossion & Pourtois, 2004) of the Snodgrass and Vanderwart picture corpus (1980), and various internet sources. There were three conditions of word type: (1) 31 homophone targets, (2) 26 LF control words, and (3) 26 HF control words.³

Word frequency was taken from the lemma-based frequency counts of the online Celex database (<http://celex.mpi.nl/>; Baayen, Piepenbrock, & van Rijn, 1993). It was possible to obtain a frequency estimate directly from Celex for the depicted meaning of a homophone when its counterparts were heterographic (e.g., deer/dear) and/or the different meanings came from different grammatical categories (e.g., watch—N, watch—V). However, for homophones whose counterparts were nouns and homographic (e.g., *elephant's* trunk vs. *car* trunk or *tree* trunk), the frequency of the depicted meaning had to be estimated. To do this, we drew on norms by Twilley et al. (1994) who used an associate production task to estimate the relative frequency of the different meanings of a large set of homographs. To illustrate, for the homophone “trunk”, the proportion of associates corresponding to its

³In the course of stimuli construction, we aimed to maximize power by maximizing the number of items in the conditions while controlling for numerous lexical variables between conditions (Table 3). This process resulted in a different number of items per condition. However, the planned statistical method of mixed logit regression is robust when applied to unbalanced designs (Janssen, 2012).

depicted meaning in our naming task (i.e., *elephant* trunk) was .13, where as it was .66 for *car* trunk and .12 for *tree* trunk (the remainder of associates for this item were classified as unclear by Twilley et al.). Elephant trunk assumed 14% of the classifiable noun meanings (i.e., $.13/ (.13 + .66 + .12) = .14$). Using the noun frequency count from Celex for trunk (i.e., 26 per million), we estimated the frequency of elephant trunk to be $.14 \times 26 = 3.64$ per million.

Average log-transformed frequency for the LF controls, HF controls, homophone targets and average log-transformed summed frequency of all homophone meanings is displayed in Table 3. Two-tailed pairwise t-tests applied to these estimates revealed frequency of the LF controls and depicted homophone targets did not differ; nor did that of the HF controls and all homophone meanings (all $ps > .05$). Frequency of the HF controls and all homophone meanings both exceeded that of the LF controls and homophone targets (all $ps < .001$).

We controlled in a groupwise manner important psycholinguistic variables known to impact naming including phonological neighborhood density (e.g., Middleton & Schwartz, 2010; Vitevitch, 1997, 2002), familiarity, number of phonemes, phonotactic probability, imageability, and name agreement. Phonological neighborhood density (i.e., how many words are similar to a target except for one phoneme), familiarity, and number of phonemes were taken from the Washington University Speech and Hearing Neighborhood Database (directed by Mitch Sommers, Ph.D.; <http://neighborhoodsearch.wustl.edu/neighborhood/Home.asp>) a web-based interface to the Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984). Phonotactic probability, or the average frequency of the sound sequences in the words in each condition, was taken from the Irvine Phonotactic Online Dictionary (Vaden, Halpin, & Hickok, 2009).

To obtain imageability values for our items, we were unable to rely on published norms because there are no comprehensive norms that distinguish between the meanings of homographic homophones. Instead, we had 18 college-aged participants provide imageability ratings to all stimuli names following a modified procedure of Cortese and Fugett (2004). Particularly important for disambiguating the homophones, all words were preceded by a disambiguating phrase (e.g., shoot the duck). We controlled for the sensibility/familiarity of the phrases between conditions by equating the frequency of the phrases as measured in the internet search engine Google. In the imageability rating task, participants were instructed to rate the imageability of the underlined word, rather than the whole phrase.

We also assessed name agreement for our materials. Instead of using a naming task, which may have privileged homophones over the low frequency control words due to potential benefits from frequency inheritance, we used a comprehension task to measure name agreement. Seventeen college-aged controls completed the task, which involved judging on a 7-point scale the degree to which the noun fit the picture.⁴

⁴The name agreement task was originally used to guide construction of the materials. As a result, values were absent for six of the LF control words. In the ANOVA and pairwise analyses of this variable, these values were replaced with the mean for the LF condition. Parallel analyses excluding these items altogether produced the same results. The mean for the LF controls displayed in Table 3 is the mean based on the replacement method.

Table 3 reports the means and standard errors for the control variables of phonological neighborhood density, number of phonemes, phonotactic probability, familiarity, imageability and name agreement. There were no reliable differences between any of the conditions for any of the control variables by two-tailed pairwise t-tests (all $ps > .05$).⁵

Procedure

All pictures were presented in a random order for naming on a desktop or laptop computer. Participants were instructed to name the pictures as best they could, and were allowed to continue to try to retrieve the word as long as they felt comfortable. The task took about 10 minutes to complete.

In order to increase the number of observations per participant, after four or more months we invited each participant to complete the task again in a second session. Though all of the Stage-2 participants were interested and able to complete a second session, four of the Stage-1 participants were not: two were lost to follow-up (S4, S5) and one had suffered a second neurological incident (S2). For those who completed the task in a second session, the average delay between sessions was 13 months (range: 4 months—23months). The analyses were conducted on all available data for each participant, and we accounted for potential differences in performance across sessions for individual participants by including session as a predictor in each analysis (see below).

Data Coding

Participants' verbal responses were digitally recorded, transcribed, and checked by two research assistants trained in IPA. A research assistant naïve to the hypotheses applied the standard coding scheme of the Philadelphia Naming Test (Roach et al., 1996; see Schwartz et al., 2006 for more details) to the first complete response (non-fragment) produced within 30 seconds. A second naïve rater checked the coded responses, and any disagreements in coding were resolved in discussions with a third naïve rater. A response was coded as accurate if it was produced fully with no phonological errors. African American vernacular variations in pronunciation as well as singulars produced in place of a plural or vice versa (e.g., socks for the target sock) were also accepted as correct. In a departure from the standard coding scheme of the PNT, which employs lenient scoring for motor-speech apraxia, responses by participants with apraxia that involved a single phonological error were still coded as errors. The main error categories included semantic errors, phonological errors, and omissions (defined in the Participants section).

⁵Because of the difficulty inherent in estimating values associated with specific meanings of the homophones, a final variable that was problematic to control was semantic neighborhood density. Mirman (2011) found that greater numbers of 'near' semantic neighbors (i.e., high feature overlap with a target) deleteriously affected naming accuracy in aphasia. To the degree our homophone targets tend to have many near neighbors, and our Stage-1 participants are more vulnerable than the Stage-2 participants to the negative effects of this semantic level variable, this could produce the predicted pattern of results. Thus, we conducted an analysis of naming accuracy following Mirman (2011), who constructed two groups of items (drawn from the PNT) that had many near versus few near semantic neighbors while controlling for a host of other variables. The decrease in accuracy in the many near neighbors condition compared to the few near neighbors condition was similar in the two groups (4% in the Stage-2 group and 5% in the Stage-1 group), providing no indication that the Stage-1 group was more sensitive to semantic neighborhood density than the Stage-2 group.

Data Analysis

Accuracy was analyzed using a mixed logit regression approach (Jaeger, 2008; Quene & Van den Bergh, 2008), where the logit (log odds) of the categorical dependent variable (i.e., correct/error) was modeled as a function of fixed factors and random effects. The regressions were conducted using the lme4 package in R version 2.15.3 (R Development Core Team, 2012). Session was also included as a fixed effect in all models to capture any variance that may have resulted from some participants completing the task twice.

All models described below included random intercepts for participants and items to capture the correlation among observations that can arise from multiple participants giving responses to the same set of items (i.e., crossed random effects; see Quene & Van den Bergh, 2008). Across all mixed model analyses, random slopes for key design variables entered as fixed effects were included if they improved model fit by chi-square deviance in model log likelihoods (Baayen, Davidson, & Bates, 2008).⁶ This only occurred once, in the “omnibus” interaction model of accuracy, described shortly. See Table 4 for model coefficients and associated statistics.

The accuracy analysis started with first establishing the equivalent of an “omnibus” interaction of word type and participant group using a model comparison procedure, where the change in model fit was evaluated with a chi-square deviance in model log likelihoods. The full model included the following fixed factors: two-level factor of session (session 1/ session 2); three-level factor of word type (homophone targets/LF controls/HF controls); two-level factor of participant group (Stage-1/Stage-2); interaction of word type and participant group. In the reduced model, the interaction term was omitted. In the full model, by setting the homophone targets/Stage-2 group cell of the design as the reference, coefficients for two lower-order interactions were estimated: (1) the relative difference in accuracy between homophones and LF controls as a function of participant group (lower order interaction 1, Table 4); (2) the relative difference in accuracy between homophones and HF controls as a function of participant group (lower order interaction 2, Table 4).

Following this analysis, we conducted the equivalent of two planned pairwise tests on accuracy using the mixed logit approach to verify the absence and presence of frequency inheritance in the Stage-1 and Stage-2 group, respectively. For the contrast applied to the Stage-1 group data, a mixed logit model including the fixed effects of word type and session were fit to data from the homophone and HF control conditions; in the Stage-2 group, a mixed logit model including the fixed effects of word type and session were fit to data from the homophone and LF control conditions (see Appendix S1 in online supplemental materials for additional details of the model fitting procedure for the pairwise models of accuracy). Parallel pairwise analyses were conducted on the prominent error type in each group: for the Stage-1 group, omissions (presence/absence) in the homophone and HF

⁶A recent paper (Barr, Levy, Scheepers, & Tily, 2013) advocates always including random slopes for fixed effects that are key design variables (i.e., *maximal random effects*) regardless of the quality of the model or whether the additional model complexity is justified by log likelihood. Across all analyses in the behavioral and computational sections, maximal random effects did not impact the statistical significance ($\alpha = .05$) of key design variables except in one case: In the pairwise model of accuracy in the Stage-2 group, inclusion of a random slope for word type by-participants inflated the p-value for the fixed effect of word type from $p=0.03$ to $p=0.07$. Several indices militated against inclusion of maximal random effects in this case. Appendix S1 in the online supplemental material outlines the model fitting procedure for the two pairwise models of accuracy.

control conditions were modeled as a function of word type and session; in the Stage-2 group, phonological errors (presence/absence) in the homophone and LF control conditions were modeled as a function of word type and session.

Results and Discussion

Figure 1 depicts boxplots of average proportion accurate responding as a function of word type and participant group. See Table 4 for mixed logit regression model results, including coefficients and associated statistics for the predictors included in each model. The interaction of word type and participant group in the omnibus analysis was significant, indicated by a reliable improvement in the fit of the full versus the reduced model [$X^2(2) = 8.71, p = .013$]. Both lower order interactions were significant (p 's $< .05$). The first lower order interaction indicates that compared to the Stage-2 group, in the Stage-1 group, homophones were more similar to the LF controls. The interpretation of the second lower order interaction is that compared to the Stage-2 group, in the Stage-1 group, homophones were at a greater disadvantage relative to the HF controls. Note, follow-up analyses showed that estimates for the two lower order interactions changed negligibly when the log frequency values for items (using either the frequency of the depicted meaning of the homophone or cumulative frequency) was added to the full model as a continuous covariate. This indicates that the lower order interactions were not produced spuriously (e.g., from non-linear frequency effects and/or distributional differences in frequency values in the word type conditions).

The pairwise model on accuracy in the Stage-1 group indicated that the log odds of a correct response for the HF controls exceeded that of the homophones ($p = .04$). A more intuitive expression of this difference is in terms of the observed odds ratio of 1.86, which indicates that there was an 86% increase in the odds of a correct response with HF controls versus homophones for the Stage-1 group. In the Stage-2 group, the log odds of a correct response for homophones exceeded that of the LF controls ($p = .03$). An observed odds ratio of 1.58 corresponds to a 58% increase in the odds of a correct response for homophones versus LF controls in the Stage-2 group. Overall, the performance in the two groups was markedly different with respects to homophony: the Stage-2 group showed frequency inheritance, with homophone accuracy patterning after the HF controls and reliably exceeding that of the LF controls; the Stage-1 group showed no frequency inheritance, with homophone accuracy patterning after the LF controls, and significantly less than the HF controls.

Figure 2 shows the proportion of responses that comprised the major error types (including semantic errors, phonological errors, and omissions) as a function of word type and participant group. The "other" category was composed primarily of unrelated word errors and a few picture-part errors (i.e., where they named part of the depicted object). See Table 4 for results of the pairwise mixed logit analyses on the errors. The Stage-1 group failed to accrue benefits from homophony because of heightened omissions in the homophones condition (31%) compared to the HF control condition (20%; $p = .03$). In the Stage-2 group, the advantage in the homophone condition was due to fewer phonological errors (7%) compared to the LF condition (15%; $p < .001$). Note, a plausible alternative interpretation of this finding is that the LF words were harder than the homophones to articulate and errors in

phonetic/articulatory procedures were driving the difference in rate of phonological errors. Phonetic errors masquerading as phonological errors are most likely to be single phoneme errors in the two participants with speech apraxia (e.g., errors originating from erroneous selection of a phonetic feature or poor coordination of the articulators). However, counter to this alternative explanation, the proportion of single phoneme errors for the participants with apraxia was higher in the homophones condition (75%) compared to the LF condition (66%). Across the Stage-2 group, single phoneme errors were more likely in the homophones condition (70%) than the LF condition (63%), suggesting that the LF words were not more difficult to articulate.

That the beneficial impact of homophony in the Stage-2 group manifested as a reduction in phonological errors aligns well with prior interpretations of frequency inheritance effects, the notion being that a high frequency counterpart exerts its beneficial influence on phonological retrieval (e.g., Biedermann & Nickels, 2008a; 2008b; Jescheniak & Levelt, 1994). The detrimental effect of homophony manifested in increased omissions in the Stage-1 group. This is consistent with the proposal that a highly activated counterpart at the word level disrupts Stage-1 retrieval when considered in tandem with evidence that lexically-driven access is a major contributor to omission errors in aphasia (Chen et al., 2013). However, the question remains whether the dissociation in accuracy as a function of participant group and word type can be taken as evidence for both a beneficial and a detrimental impact of homophony. For instance, the results could reflect that the Stage-1 group is simply insensitive to the benefits of frequency inheritance, and not that homophony is negatively impacting performance in this group. A series of computational explorations and additional analyses, reported next, speak to this issue.

Computational Simulations

Model Structure

To implement a computational version of the dual nature account of homophony, we adapted the standard interactive two-step model of lexical retrieval from Dell and colleagues (Dell et al., 2004; Foygel & Dell, 2000; Schwartz et al., 2006), depicted in Figure 3. The model has three layers: a semantic layer, a word layer, and a phoneme layer. Units in the semantic layer represent semantic features of concepts or objects (10 semantic units per lexical item; e.g., the semantic features for the concept “cat”: mammal, furry, small, etc.). Units in the word layer gloss as lemmas (note, we revisit this assumption in the General Discussion). Units in the phoneme layer correspond to individual phonemes coded with respect to syllable position. The semantic layer links to the word layer (hereafter, *s-weights*), and the word layer links to the phoneme layer (hereafter, *p-weights*). All connections are bidirectional. To simulate the naming impairment in our two participant groups, the connection strength of the *s-weights* and *p-weights* was manipulated, with greater deviations below an “intact” value of 0.1 corresponding to increasing impairment in Stage-1 or Stage-2 retrieval, respectively.

We started with the two basic lexical neighborhood structures used in Schwartz et al., (2006), a “mixed” and a “nonmixed” neighborhood. The lexicon of the mixed neighborhood included: the target “cat”; a semantically related but phonologically unrelated neighbor (i.e.,

“dog”, semantic neighbor); a semantically and phonologically related (mixed) neighbor (“rat”, mixed neighbor); a semantically unrelated but phonologically related neighbor (“hat”, formal neighbor) and two semantically and phonologically unrelated neighbors (one of which was “log”, unrelated word). The nonmixed neighborhood was identical except the mixed neighbor was changed to a formal neighbor. Following Schwartz et al. (2006) the results of any simulation were determined by averaging across the mixed and nonmixed structures, with the results weighted at 20% versus 80% for the mixed and nonmixed neighborhoods, respectively. Hereafter, the term *standard structure* should be taken to refer to the combination (and aggregated results) of the mixed and nonmixed structures.

To implement the word type manipulation from the experimental design, we represented frequency in the model by changing the strength of the target’s p-weights (i.e., the p-weights linking from the word node CAT to /k/ /æ/ /t/) relative to the model’s p-weight strength (i.e., strength of all p-weights in the model, which was manipulated across the various simulations). To represent the LF control condition, the target’s p-weights were set to model p-weight strength multiplied by 0.7 (reducing the target’s p-weights relative to all other p-weights in the model). In the HF control condition, the target’s p-weights equaled all other p-weights in the model. For the homophone condition, to represent the high frequency homophonic counterpart we first added a word node that was phonologically identical to the target “cat” (i.e., same onset, vowel, coda), and set its p-weight strength to 1.2 times model p-weight strength. Note, frequency of the HF counterpart was set higher than frequency of the target in the HF control condition. This is because there was a strong trend ($p = .13$) for average cumulative homophone frequency to exceed average HF control frequency in the experimental materials. The HF counterpart was semantically nonoverlapping with the target and the other lexical neighbors (i.e., had 10 unique semantic feature units). The target’s p-weights in the homophone condition were set to model p-weight strength multiplied by 0.7, thus equating the frequency characteristics of the target in the homophone condition and the LF control condition. Hereafter, reference to manipulation of s-weights and p-weights denotes changes in the strength of the s- or p-weights of all units in the model except when explicitly noted.

Processing

We used the standard activation rule in the interactive two-step model of lexical retrieval, where the activation of all units was updated following the equation in (1):

$$A_{j,t-1} = A_{j,t-1}(1-q) + \sum w_{ij} A_{i,t-1} + noise \quad (1)$$

In (1), $A_{j,t}$ was the activation of unit j at time step t ; q was the decay constant; w_{ij} was the connection strength from the source unit i to the receiving unit j ; each unit’s activation level was perturbed by normally distributed noise, specifically the sum of two components, intrinsic noise ($SD = 0.01$) and activation noise ($SD = 0.16 A_{j,t-1}$).

The usual processing characteristics of the standard interactive two-step model (e.g., Foygel & Dell, 2000) were adopted. Processing consisted of two stages marked by “jolts” of activation to specific units. Stage-1 retrieval began with a jolt of ten units of activation to

each of the target's semantic features. Activation was allowed to spread through the network for eight time steps, after which the most highly activated word unit was selected by applying a 100-unit jolt of activation. This constituted completion of Stage-1 retrieval and initiation of Stage-2 retrieval, wherein activation was again allowed to spread through the network for an additional eight time steps. At the conclusion of the eighth time step, the most highly active onset, vowel, and coda among the phonemes were selected and associated with the corresponding slots in a CVC phonological frame to determine model output. Each output was categorized into one of the following response types: correct response, semantic error (e.g., dog for target "cat"), formal error (e.g., hat), mixed error (e.g., rat), unrelated word error (e.g., log), and nonword error (e.g., dat).

We adopted these processing characteristics fully except for one additional feature, described shortly. A limitation of earlier versions of the interactive two-step model of lexical retrieval (c.f., Dell et al., 2004) was that it did not account for omission errors in naming. Omission errors are present in all the classic subtypes of aphasia, and for the participants in the Stage-1 group, omission errors accounted for a majority of their naming errors in the main task as well as in the PNT. Recent evidence suggests semantically-driven lexical access is a prominent cause of omission naming errors in aphasia (Chen et al., 2013). This motivated incorporation of a lexical-threshold rule using normalized activation to enable omissions to arise in our model during Stage-1 retrieval, adapted from a version of the two-step interactive model that was designed to account for omission errors (Dell et al., 2004). Normalization is a common assumption in many computational models (e.g., McNellis & Blumstein, 2001; Verguts & Fias, 2004), though it can be implemented more neurally via lateral inhibition in the word layer (Chen & Mirman, 2012; Usher & McClelland, 2001). In the current model, after the eighth time step during Stage-1 retrieval, a word unit was selected and Stage-2 retrieval was initiated only if that unit's normalized activation as given in (2) achieved or surpassed the threshold. Otherwise, the model generated no response (i.e., omission error).

$$N_{j,8} = \frac{L_{j,8}}{\sum_{i=1}^{all} L_{i,8}} \quad (2)$$

Across simulations, the decay rate was kept constant at 0.6, as was the lexical-threshold, set to 0.3. Each naming model was run through 1000 simulated trials.

Model Results and Predictions

Figure 4 shows simulated proportion of correct target productions in the homophone condition, LF control condition, and HF control condition at varying levels of s-weight strength with intact p-weights (left panel), and varying levels of p-weight strength with intact s-weights (right panel). Firstly, Figure 4 (both panels) shows that with intact processing at both stages (i.e., s-weights and p-weights both set to 0.1), the model demonstrated the standard frequency effect on naming accuracy (e.g., Kittredge et al., 2008) in that the target in the HF control condition was associated with greater accuracy than in the LF control condition. Also of note, the model was capable of showing frequency inheritance

(i.e., target accuracy in the homophone condition > LF control condition), such as when s- and p-weights were both intact (Figure 4, both panels).

The model's ability to capture the standard effect of word frequency on target accuracy as well as frequency inheritance forms a solid foundation for exploring the dual nature account of homophony. In that vein, the first important observation was the relationship revealed by the model between Stage-1 and Stage-2 impairment and homophony's impact on target accuracy. In Figure 4 (left panel), simulated target accuracy in the homophone condition was at a disadvantage compared to the LF control condition for a large range of s-weights. In contrast, there was a consistent advantage for target accuracy in the homophone condition compared to the LF control condition for almost the entire range of p-weights (right panel).

The intuition for how this dissociation was generated by the model is as follows: a homophone target and its counterpart are distinguishable only by semantics because their phonological representation is exactly the same. When semantic information is impoverished (implemented here as weak s-weights), the difference in activation between the homophone target and its HF counterpart is reduced, decreasing the likelihood the target lexical node will surpass the activation threshold. Though semantic input likely favors the target (even with low s-weight strength), the HF counterpart remains a problem during Stage-1 retrieval because of interactive activation and shared phonology with the target. In contrast, when phonological information is impoverished (weak p-weights), intact semantic information is sufficient to faithfully distinguish the target from its counterpart, allowing an observation of frequency inheritance to emerge.⁷

In line with this explanation, inspection of the model showed that the disadvantage of target accuracy in the homophone versus the LF control condition at various levels of s-weight strength was due to elevated omission errors. Further exploration of the homophone and LF control conditions revealed an important relationship, setting the stage for a testable prediction in the behavioral data—with intact p-weights, as s-weight strength decreased, the rate of omission errors on the target in the homophone condition relative to the LF control condition increased (see Figure 5, left panel). Calculating omission error rates in the homophones condition relative to the LF control condition was important because it allowed us to diagnose the impact of the presence of a high frequency homophone counterpart specifically, while holding all other processing characteristics constant.

We used multiple regression to evaluate whether this pattern was evident in the behavioral data. The dependent variable was defined as the difference in proportion of omissions for each participant in the homophones condition versus the LF control condition (termed *relative omission rate*). The predictors in the regression were s- and p-weights for all participants, based on each individual's pattern of performance on the Philadelphia Naming Test (see Dell et al., 2004, for details of how s- and p-weights are estimated in the two-step model). Note, since it was not possible to hold p-weights at an intact level (as was done in the simulation), we elected to statistically control p-weights by including them as a predictor

⁷Note, it is not clear why at the lowest s-weights, the model shows a facilitative effect of homophony relative to the LF controls. It may be an interesting topic for future research to evaluate this aspect of the model, specifically that when semantic input is virtually absent, in the presence of intact phonological information, having a homophonic counterpart can be beneficial.

in the model. In the regression, s-weights were a significant predictor ($\beta = -0.74$, $p = .03$; $R^2 = 0.53$) of relative omission rate but not p-weights ($\beta = -0.01$, $p = .96$); and, the direction of the coefficient indicates that as s-weights decreased, relative omission rate increased (see Figure 5, right panel). This result confirms the first novel prediction of the model and provides evidence for our proposed detrimental impact of homophony: increasing degradation in s-weights led to an increasingly exaggerated (and negative) effect of homophony during Stage-1 retrieval, manifesting in omission errors. Because the omission error rates were calculated relative to the LF control condition, this analysis points to the unique impact of high frequency homophonic counterparts on Stage-1 retrieval.

The model made a second testable prediction having to do with the frequency of the HF homophonic counterpart and the dominant error type associated with each stage of retrieval (see Figure 6). In these simulations, the frequency of the target was held constant (i.e., p-weights for target “cat” were set to the model’s p-weight strength multiplied by 0.7), and we varied the frequency of the HF homophonic counterpart by setting the counterpart’s p-weights to model p-weight strength multiplied by 0.8, 1.0, or 1.2. The model produced an interaction of the frequency manipulation and error type (see Figure 6): as shown in the left panel, across the range of s-weights (with intact p-weights), omission errors increased with increasing counterpart frequency; in the right panel, across the range of p-weights (with intact s-weights) phonological errors decreased with increasing counterpart frequency. The intuition for this result is that as the counterpart becomes more frequent relative to the target, it is more highly activated during Stage-1 retrieval (via interactive activation). Hence, higher counterpart frequency is associated with more interference during Stage-1 retrieval, leading to an increase in omissions. However, higher counterpart frequency also translates into more reverberated activation through shared phonemes, thus reducing phonological error rates.

To evaluate whether this pattern was reflected in the behavioral data, we applied mixed logit regression to a dependent variable defined at the trial level as follows: 1=omission/0=other for the Stage-1 group, and 1=phonological error/0=other for the Stage-2 group. For each item in the homophone condition, we calculated the degree to which the frequency of the HF counterpart “dominated” the combined frequency of the HF counterpart and LF homophone target using equation (3):

$$\text{frequency dominance} = \frac{\log \text{HF counterpart frequency}}{\log \text{LF homophone frequency} + \log \text{HF counterpart frequency}} \quad (3)$$

In the mixed logit model, random intercepts were entered for participants and items, along with fixed effects of dominance, session, participant group (the Stage-2 group was set as the reference), and the interaction of dominance and participant group. The interaction of dominance and participant group was reliable (coefficient = 5.64 standard error = 2.56, $p = .03$). To understand the nature of this interaction, for each homophone target in the Stage-1 group we plotted the proportion of omissions (out of all responses in the Stage-1 group), and the proportion of phonological errors (out of all responses in the Stage-2 group), as a function of each item’s calculated frequency dominance (Figure 7). Figure 7 shows that whereas phonological error proportions were somewhat negatively related to dominance,

omission error proportions were positively related to dominance. We note that in contrast to the simulations, phonological error proportions appear to only trend downwards with increasing frequency dominance, showing a weaker effect of dominance than in the simulation. However, the effect of dominance on phonological errors in the participant data may be less strong because the participants in the Stage-2 group had variable s-weights. In contrast, in the simulation, s-weights were intact. Importantly, the interaction of dominance with the two error types largely confirms the second novel prediction generated by the model. More generally, these results are consistent with the view that homophony exerts dissociable effects on the two stages of naming.

General Discussion

In the present study, participants with chronic aphasia with a primary naming impairment in the retrieval of words from semantics (Stage-1 group), or in phonology from words (Stage-2 group), showed a marked dissociation in naming homophones with high frequency counterparts. The Stage-1 group showed no frequency inheritance, with accuracy on the homophone targets patterning after the LF controls and showing a disadvantage compared to the HF controls. In contrast, the Stage-2 group showed full frequency inheritance, with accuracy on the homophone targets patterning after the HF controls and showing an advantage compared to the LF controls. We take these contrasting results as evidence for the dual nature account of homophony. In the Stage-2 group, the beneficial impact of homophony exerted an effect by mitigating phonological errors. This pattern is consistent with claims that high frequency counterparts facilitate the phonological retrieval of homophonic mates (e.g., Jescheniak & Levelt, 1994; Biedermann & Nickels, 2008a; 2008b). The negative impact of homophony, shown by the Stage-1 group, appeared to diminish accuracy on homophone targets by elevating omission error rates.

We modeled the impact of a HF counterpart on a LF homophonic mate during the two main stages of retrieval during naming with a framework similar to that of Dell and colleagues (Foygel & Dell, 2000; Schwartz et al., 2006) but that included the additional assumption of competitive selection; and, invoked a lexical threshold rule to account for omission errors (Dell et al., 2004). When p-weights were lesioned (with intact s-weights), the model showed a consistent advantage for homophones over the LF controls, reflecting the Stage-2 group behavioral data. Conversely, when s-weights were lesioned (with intact p-weights), the model showed a disadvantage for homophone targets compared to the HF controls, similar to the Stage-1 group performance. The s-weight lesioned model also revealed a disadvantage for homophone targets compared to the LF controls (except at very mild or very extreme levels of s-weight dysfunction) because of heightened omission errors. Further model exploration lead to a testable prediction which was confirmed: in a posthoc analysis across all participants, s-weights were negatively related to an increase in omission error rates in the homophones condition relative to the LF control condition. This shows that the negative effect of homophony is to perturb word selection, which exerts greater detriment as the mapping from semantics to words is increasingly compromised. More generally, this is strong evidence for interactive activation: because the different meanings of homophones are semantically nonoverlapping, a high frequency counterpart could only become a

competitor during Stage-1 mapping via interactive activation through shared phonology with the homophone target.

Additional simulations revealed that increasing frequency dominance of the high frequency counterpart relative to the homophone target exerted distinct effects on the two major error types associated with each stage of retrieval. Consistent with this pattern, a posthoc analysis on the behavioral data showed a trend downwards in phonological errors (Stage-2 group) with increasing dominance, but an increase in omission errors with increasing dominance (Stage-1 group). Overall, the results from both posthoc analyses validated the model, and provided support for the dual-nature account of homophony.

Models of Homophone Production

We adapted the two-step interactive model of lexical access to simulate the dual nature account of homophony and generate predictions because it is an explicit, well-understood framework for modeling naming accuracy and errors. It was also particularly well suited to our goal of investigating distinct effects of homophony on the two stages of retrieval. It is important to note that a number of other interactive models may be compatible with this framework with minimal modification (e.g., Cutting & Ferreira, 1999; Dell, 1990; Harley, 1993; Rapp & Goldrick, 2000). Our results are more obviously inconsistent with models that assume information flow is strictly feedforward, including the WEAVER++ model (Levitt et al., 1999) and the Independent Network model (Caramazza, 1997). The WEAVER++ model straightforwardly accounts for our finding of frequency inheritance in the Stage-2 group because it assumes that homophonic counterparts access the same lexeme. However, this model would need to relax its feedforward assumption to account for the detrimental impact of homophony on naming. Our results are also inconsistent with the Independent Network model as formulated in Caramazza (1997), but would be consistent with an interactive version, similar to that described by Caramazza et al., (2001) and Caramazza and Miozzo (1998). The present findings suggest such a change would be well motivated. In fact, an interactive Independent Network model would be very similar to the architecture developed here. One obvious difference, however, is that we assume modality-neutral lexical representations (i.e., lemmas) mediate between semantics and phonological encoding. This means that the same lexical representation for a word is accessed whether the task is comprehension, oral picture naming or written picture naming. In the Independent Network model, only modality-specific lexical representations mediate and lexico-grammatical properties of a word are coded redundantly in each modality, i.e., in separate orthographic and phonological form representations. In other words, a different lexical representation is accessed whether a picture name is retrieved for written production or oral production. We assume modality-neutral representations are intermediary because such an assumption is made in the two-step interactive theory that we adapted for exploration of the dual nature of homophones. However, we admit our data simply do not speak to the issue of the modality-specificity of intermediary lexical representations.

Frequency Inheritance Revisited

The dual nature account has potential to shed light on the contradictory results found in prior work on frequency inheritance in homophone production. One implication of our account is

that when Stage-1 retrieval is exclusively semantically-driven (i.e., such as during picture naming), the detrimental impact of homophony should have a greater chance overwhelming homophony's benefits than when external lexical input can be used to facilitate Stage-1 retrieval (i.e., in word naming, or word translation). Interestingly, the literature appears to reflect this. The two paradigms used in the original four demonstrations of frequency inheritance involved written word naming (Dell, 1990; Experiment 1) and translation (Jescheniak & Levelt, 1994, Experiment 6; Jescheniak et al., 2003, Experiments 1 & 2). Such tasks arguably impose less demand on Stage-1 retrieval than naming. In word naming, Stage-1 retrieval could be facilitated by input to the target word node from orthography. Likewise, when translating words, evidence suggests bilinguals can rely at least partially on direct L1 to L2 word mappings (e.g., Chen & Leung, 1989; Kroll & Curley, 1988), which could supplement or circumvent Stage-1 retrieval. In contrast, the majority of the failures to replicate the frequency inheritance effect using controlled experimentation has involved naming from pictures or definitions (Anton-Mendez et al., 2012; Bonin & Fayol, 2002; Caramazza et al., 2001, Experiments 1a & 2a; Cuetos et al. 2010, Experiments 1a, 2a, 3a; Miozzo et al., 2004, Experiments 1, 2, 3). Commensurate with the dual nature account, when the requirement of semantically-driven Stage-1 mapping is reduced, the negative impact of homophony will be weaker, and homophony's benefits will prevail in determining its net impact on performance.

Our account is also consistent with the one other study (to our knowledge) that has investigated homophone production in aphasia using the current design (Miozzo et al. 2004). In that study, participant AW produced a preponderance of omission errors (i.e., by our classification, which included no responses and circumlocutions) and semantic errors in the experimental tasks and neuropsychological testing. Phonological naming errors were very rare and AW was reported to have good repetition. Because AW performed well on comprehension tasks, Miozzo et al. attributed her naming impairment to unavailability of phonological information. However, given that semantic errors and omissions dominated her naming performance, and there was little evidence of marked phonological output disturbances (i.e., few phonological naming errors, good repetition), AW's naming impairment is most consistent with a Stage-1 retrieval deficit in our theoretical framework. Interestingly, AW's performance in the experimental tasks mirrored that of the Stage-1 group pattern, showing no benefits on homophone targets compared to LF controls across three experiments.

Because the literature on homophone production has mainly focused on possible benefits to accuracy from frequency inheritance, our stance regarding the proposed benefits of homophony in word production is arguably less controversial than our proposal concerning its negative influences. However, in domains of lexical processing that do not involve production such as word recognition, the negative impact of homophony is better known. For example, many studies have shown that reaction times in lexical decision tasks are slower for low frequency targets with high frequency heterographic homophone mates, compared to nonhomophones (e.g., in French, Ferrand & Granger, 2003; in English, Pexman, Lupker, & Jared, 2001; Pexman, Lupker, & Reggin, 2002; Rubenstein, Lewis, & Rubenstein, 1971). The preferred interpretation is as follows: phonological activation arises when an orthographic form is presented, which feeds back to orthographic/lexical

representations in the course of recognition. Hence, due to identical phonology with the target and a counterpart's frequency characteristics, a high frequency homophonic counterpart becomes a strong lexical competitor with the target, making lexical decision difficult. The parallels to our account of the negative side of homophony—lexical competition created via interactive activation through shared phonology—are striking.

An interesting complication in work on homophone effects in lexical decision is that the number of homophone mates appears to impact whether a target will show an advantage or disadvantage compared to control words. Specifically, Hino, Kusunose, Lupker, and Jared (2012) found in Japanese that targets with a single homophonic mate (typical in English) were associated with a processing disadvantage, whereas homophone targets with multiple mates showed a processing advantage. Interestingly, available evidence suggests there may be parallel effects in naming: in contrast to the detrimental impact of homophony, studies have shown that Stage-1 retrieval is more accurate for targets with more versus fewer phonologically similar but nonidentical neighbors (Kittredge et al., 2008; Middleton & Schwartz, 2010).

These different effects of formally related neighbors may point to a principle proposed by Chen and Mirman (2012) that captures a range of neighborhood effects across domains of lexical processing: “strongly active neighbors [have] a net inhibitory effect, and weakly active neighbors [have] a net facilitative effect.” With regards to the diverse effects of formally related neighbors on Stage-1 retrieval in naming, the explanation could go something as follows: homophone counterparts in the current study were strongly active neighbors because in English, homophone targets tend to have only one or a few counterparts (a finite pool of activation is distributed among a small number of neighbors); homophone counterparts have identical phonology with a target; and, (in our study) the counterparts generally had higher frequency than the targets. In contrast, phonologically nonidentical but related neighbors are weak neighbors because there are more of them, and phonological overlap with a target is less than absolute. Notably, increasing numbers of these neighbors translates to greater Stage-1 facilitation consistent with the notion that a finite pool of activation shared amongst many versus few neighbors ensures no single neighbor is strongly active. This enables the net benefits of multiple weakly related neighbors on Stage-1 retrieval to emerge via reverberatory feedback between shared phonemes and the formally related neighbors. However, we feel it is important to emphasize the speculative nature of the discussion. Further research is required to determine how these factors (i.e., similarity of formal neighbors, number of neighbors and their frequency) play out in determining lexical competition in speech production as well as how competition should be implemented (e.g., as the normalization activation scheme implemented here, or possibly as direct lateral inhibition between lexical units as in Chen and Mirman, 2012).

A more general implication of this discussion is that homophony may not be special. Rather, homophony may be best characterized as a type of extreme formal relatedness, where its effects are dictated by processing dynamics that arise from the joint action of variables such as similarity of neighbors to targets and to each other, cohort size, and frequency. Only by delineating such processing dynamics will we fully understand when and why formally-related neighbors manifest as friends or foes.

Acknowledgments

This work was supported by NIH research grant (RO1-DC000191) awarded to Myrna Schwartz, and NIH training grant (T32-HD007425). A great many thanks are in order to Rachel Jacobson, Adelyn Brecher, Anne Mecklenburg and Jennifer Gallagher for data collection and processing. Many thanks to Bonnie Nozari and Dan Mirman for statistical assistance, as well as Katherine Rawson for normative data collection.

References

- Anton-Mendez I, Schuetze CT, Champion MK, Gollan TH. What the tip-of-the-tongue (TOT) says about homophone frequency inheritance. *Memory and Cognition*. 2012; 40(5):802–811. [PubMed: 22351522]
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59(4):390–412.
- Baayen, RH.; Piepenbrock, R.; van Rijn, A. The CELEX lexical database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania; 1993.
- Badecker W, Miozzo M, Zanuttini R. The 2-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*. 1995; 57(2):193–216. [PubMed: 8556841]
- Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*. 2013; 68:255–278.
- Beeson PM, Holland AL, Murray LL. Naming famous people: An examination of tip-of-the-tongue phenomena in aphasia and Alzheimer's disease. *Aphasiology*. 1997; 11(4–5):323–336.
- Biedermann B, Blanken G, Nickels L. The representation of homophones: Evidence from remediation. *Aphasiology*. 2002; 16(10–11):1115–1136.
- Biedermann B, Coltheart M, Nickels L, Saunders S. Effects of homophony on reading aloud: Implications for models of speech production. *Language and Cognitive Processes*. 2009; 24(6):804–842.
- Biedermann B, Nickels L. The representation of homophones: More evidence from the remediation of anomia. *Cortex*. 2008a; 44(3):276–293. [PubMed: 18387557]
- Biedermann B, Nickels L. Homographic and heterographic homophones in speech production: Does orthography matter? *Cortex*. 2008b; 44(6):683–697. [PubMed: 18472038]
- Bloem I, La Heij W. Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*. 2003; 48(3):468–488.
- Bonin P, Fayol M. Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology*. 2002; 14(3):289–313.
- Bozeat S, Lambon-Ralph MA, Patterson K, Garrard P, Hodges JR. Nonverbal semantic impairment in semantic dementia. *Neuropsychologia*. 2000; 38:1207–15. [PubMed: 10865096]
- Burke DM, Locantore JK, Austin AA, Chae B. Cherry pit primes Brad Pitt: Homophone priming effects on young and older adults' production of proper names. *Psychological Science*. 2004; 15(3):164–170. [PubMed: 15016287]
- Caramazza A. How many levels of processing are there in lexical access? *Cognitive Neuropsychology*. 1997; 14(1):177–208.
- Caramazza A, Bi YC, Costa A, Miozzo M. What determines the speed of lexical access: Homophone or specific-word frequency? A reply to Jescheniak et al. (2003). *Journal of Experimental Psychology: Learning Memory and Cognition*. 2004; 30(1):278–282.
- Caramazza A, Costa A, Miozzo M, Bi YC. The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2001; 27(6):1430–1450.
- Caramazza A, Miozzo M. The relation between syntactic and phonological knowledge in lexical access: Evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*. 1997; 64(3):309–343. [PubMed: 9426505]

- Caramazza A, Miozzo M. More is not always better: A response to Roelofs, Meyer, and Levelt. *Cognition*. 1998; 69(2):231–241. [PubMed: 9894406]
- Chen Q, Graziano K, Middleton E, Mirman D. Lesion-symptom mapping errors of omission in speech production. 2013 Manuscript submitted for publication.
- Chen Q, Mirman D. Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*. 2012; 119(2):417–430. [PubMed: 22352357]
- Cortese MJ, Fugett A. Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods Instruments and Computers*. 2004; 36(3):384–387.
- Cuetos F, Bonin P, Ramon Alameda J, Caramazza A. The specific-word frequency effect in speech production: Evidence from Spanish and French. *Quarterly Journal of Experimental Psychology*. 2010; 63(4):750–771.
- Cutting JC, Ferreira VS. Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning Memory and Cognition*. 1999; 25(2):318–344.
- Dell GS. A spreading-activation theory of retrieval in sentence production. *Psychological Review*. 1986; 93(3):283–321. [PubMed: 3749399]
- Dell GS. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*. 1990; 5(4):313–349.
- Dell, GS.; Gordon, JK. Neighbors in the lexicon: Friends or foes. In: Schiller, NO.; Meyer, AS., editors. *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin: Mouton de Gruyter; 2003.
- Dell GS, Lawler EN, Harris HD, Gordon JK. Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*. 2004; 21(2–4):125–145. [PubMed: 21038196]
- Dell GS, Reich PA. Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*. 1981; 20(6):611–629.
- Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. *Psychological Review*. 1997; 104(4):801–838. [PubMed: 9337631]
- Ferrand L, Grainger J. Homophone interference effects in visual word recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*. 2003; 56(3):403–419.
- Ferreira VS, Griffin ZM. Phonological influences on lexical (mis)selection. *Psychological Science*. 2003; 14(1):86–90. [PubMed: 12564760]
- Foygel D, Dell GS. Models of impaired lexical access in speech production. *Journal of Memory and Language*. 2000; 43(2):182–216.
- Gagnon DA, Schwartz MF, Martin N, Dell GS, Saffran EM. The origins of formal paraphasias in aphasics' picture naming. *Brain and Language*. 1997; 59(3):450–472. [PubMed: 9299072]
- Gahl S. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*. 2008; 84(3):474–496.
- Gainotti G, Miceli G, Caltagirone C, Silveri MC, Masullo C. The relationship between type of naming error and semantic-lexical discrimination in aphasic patients. *Cortex*. 1981; 17(3):401–409. [PubMed: 7333113]
- Garrett, MF. Syntactic processes in sentence production. In: Wales, RJ.; Walker, E., editors. *New approaches to language mechanisms*. Amsterdam: North-Holland; 1976.
- Goodglass H, Kaplan E, Weintraub S, Ackerman N. Tip-of-tongue phenomenon in aphasia. *Cortex*. 1976; 12(2):145–153. [PubMed: 954449]
- Hanley JR, Nickels L. Are the same phoneme and lexical layers used in speech production and comprehension? A case-series test of Foygel and Dell's (2000) model of aphasic speech production. *Cortex*. 2009; 45:784–790. [PubMed: 19103444]
- Harley TA. Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*. 1993; 8(3):291–309.
- Harley TA. Will one stage and no feedback suffice in lexicalization? *Behavioral and Brain Sciences*. 1999; 22(1):45.
- Hart J, Gordon B. Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Annals of Neurology*. 1990; 27(3):226–231. [PubMed: 2327733]

- Hillis AE, Rapp B, Romani C, Caramazza A. Selective impairment of semantics in lexical processing. *Cognitive Neuropsychology*. 1990; 7(3):191–243.
- Hino Y, Kusunose Y, Lupker SJ, Jared D. The processing advantage and disadvantage for homophones in lexical decision tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2013; 39(2):529–551.
- Howard D, Nickels L, Coltheart M, Cole-Virtue J. Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*. 2006; 100(3):464–482. [PubMed: 16413014]
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008; 59(4):434–446. [PubMed: 19884961]
- Janssen DP. Twice random, once mixed: Applying mixed models to simultaneously analyze random effects of language and participants. *Behavior Research Methods*. 2012; 44(1):232–247. [PubMed: 21858733]
- Jefferies E, Ralph MAL. Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain*. 2006; 129(8):2132–2147. [PubMed: 16815878]
- Jescheniak JD, Levelt WJM. Word-frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning Memory and Cognition*. 1994; 20(4):824–843.
- Jescheniak JD, Meyer AS, Levelt WJM. Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2003; 29(3):432–438.
- Kempen G, Huijbers P. The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*. 1983; 14(2):185–209.
- Kertesz, A. Western aphasia battery. New York: Grune & Stratton; 1982.
- Kittredge AK, Dell GS, Verkuilen J, Schwartz MF. Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*. 2008; 25(4):463–492. [PubMed: 18704797]
- Kroll, JF.; Curley, J. Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. In: Gruneberg, M.; Morris, P.; Sykes, R., editors. *Practical aspects of memory*. Vol. 2. London: John Wiley & Sons; 1988. p. 389–395.
- Lambon Ralph MA, Sage K, Roberts J. Classical anomia: A neuropsychological perspective on speech production. *Neuropsychologia*. 2000; 38(2):186–202. [PubMed: 10660229]
- Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 1999; 22(1):1–75. [PubMed: 11301520]
- Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. *Cognitive Science*. 2007; 31(1):133–156. [PubMed: 21635290]
- Martin N, Saffran EM. A computational account of deep dysphasia: Evidence from a single case-study. *Brain and Language*. 1992; 43(2):240–274. [PubMed: 1393522]
- Martin N, Schwartz MF, Kohen FP. Assessment of the ability to process semantic and phonological aspects of words in aphasia: A multi-measurement approach. *Aphasiology*. 2006; 20(2–4):154–166.
- McNellis MG, Blumstein SE. Self-organizing dynamics of lexical access in normals and aphasics. *Journal of Cognitive Neuroscience*. 2001; 13:151–170. [PubMed: 11244542]
- Middleton EL, Schwartz MF. Density pervades: An analysis of phonological neighborhood density effects in aphasic speakers with different types of naming impairment. *Cognitive Neuropsychology*. 2010; 27(5):401–427. [PubMed: 21718214]
- Miozzo M, Caramazza A. The representation of homophones: Evidence from the distractor-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31(6):1360–1371.
- Miozzo M, Jacobs ML, Singer NJW. The representation of homophones: Evidence from anomia. *Cognitive Neuropsychology*. 2004; 21(8):840–866. [PubMed: 21038236]
- Mirman D. Effects of near and distant semantic neighbors on word production. *Cognitive, Affective and Behavioral Neuroscience*. 2011; 11:32–43.

- Mirman D, Strauss TJ, Brecher A, Walker GM, Sobel P, Dell GS, Schwartz MF. A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive functions. *Cognitive Neuropsychology*. 2010; 27(6):495–504. [PubMed: 21714742]
- Mirman D, Strauss TJ, Dixon JA, Magnuson JS. Effect of representational distance between meanings on recognition of ambiguous spoken words. *Cognitive Science*. 2010; 34(1):161–173. [PubMed: 20354577]
- Nozari N, Dell GS. How damaged brains repeat words: A computational approach. *Brain & Language*. 2013; 126:327–337. [PubMed: 23933472]
- Nusbaum, HC.; Pisoni, DB.; Davis, CK. Research on Speech Perception, Progress Rep. No. 10. Indiana University, Psychology Department, Speech Research Laboratory; 1984. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words.
- Pexman PM, Lupker SJ, Jared D. Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2001; 27(1):139–156.
- Pexman PM, Lupker SJ, Reggin LD. Phonological effects in visual word recognition: Investigating the impact of feedback activation. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2002; 28(3):572–584.
- Quene H, Van den Bergh H. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*. 2008; 59(4):413–425.
- R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2012. URL <http://www.r-project.org/>
- Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. *Psychological Review*. 2000; 107(3):460–499. [PubMed: 10941277]
- Roach A, Schwartz MF, Martin N, Grewal RS, Brecher A. The Philadelphia naming test: scoring and rationale. *Clinical Aphasiology*. 1996; 24:121–33.
- Rossion B, Pourtois G. Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*. 2004; 33(2):217–236. [PubMed: 15109163]
- Rubenstein H, Lewis SS, Rubenstein M. Evidence for phonemic recording in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*. 1971; 10(6):645–657.
- Schwartz MF, Dell GS, Martin N, Gahl S, Sobel P. A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language*. 2006; 54(2): 228–264.
- Shatzman KB, Schiller NO. The word frequency effect in picture naming: Contrasting two hypotheses using homonym pictures. *Brain and Language*. 2004; 90(1–3):160–169. [PubMed: 15172534]
- Snodgrass JG, Vanderwart M. Standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*. 1980; 6(2):174–215. [PubMed: 7373248]
- Sommers, MS. [Accessed October, 2011] Washington University in St. Louis Speech and Hearing Lab Neighborhood Database: A web-based implementation of the 20,000-word Hoosier Mental Lexicon. URL: <http://neighborhoodsearch.wustl.edu/neighborhood/Home.asp>
- Stemberger, JP. An interactive activation model of language production. In: Ellis, AW., editor. *Progress in the psychology of language*. Vol. 1. Hillsdale, NJ: Erlbaum; 1985. p. 143–186.
- Twilley LC, Dixon P, Taylor D, Clark K. University of Alberta norms of relative meaning frequency for 566 homographs. *Memory and Cognition*. 1994; 22(1):111–126. [PubMed: 8035680]
- Usher M, McClelland JL. The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*. 2001; 108:550–592. [PubMed: 11488378]
- Vaden, KI.; Halpin, HR.; Hickok, GS. [Accessed October, 2011] Irvine Phonotactic Online Dictionary, Version 2.0. 2009. URL: <http://www.iphod.com>
- Verguts T, Fias W. Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*. 2004; 16:1493–1504. [PubMed: 15601514]
- Vigliocco G, Hartsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*. 2002; 128(3):442–472. [PubMed: 12002697]
- Vitevitch MS. The neighborhood characteristics of malapropisms. *Language and Speech*. 1997; 40(3): 211–228. [PubMed: 9509578]

Vitevitch MS. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning Memory and Cognition*. 2002; 28(4):735–747.

Figure 1. Boxplots corresponding to accuracy performance as a function of participant group and word type.

Figure 2.

Proportion of responses (corrects and major error categories) as a function of participant group and word type.

Figure 3.
The two-step interactive model of word retrieval. See text for description.

Figure 4.

Simulated proportion of correct responses as a function of word type condition and varying strengths of s-weights (with intact p-weights; left panel) and p-weights (with intact s-weights; right panel).

Figure 5.

Relationship of s-weight strength on relative omission error rates, with simulated results (left panel) and observed results (right panel). Error band in the behavioral graph corresponds to 95% confidence interval around the linear fit line.

Figure 6.
 Simulated effects of homophone counterpart frequency on proportion of omission errors (left panel) and phonological errors (right panel).

Observed Data
 Omission Proportions (Stage-1 Group) and Phonological Error Proportions (Stage-2 Group) for Individual Items by Counterpart Frequency Dominance

Figure 7. Relationship of frequency dominance of the homophone counterpart to omission errors in the Stage-1 group and phonological errors in the Stage-2 group, with 95% confidence intervals around linear fit lines. Markers represent average error proportion out of all responses per item.

Table 1
Philadelphia Naming Test Performance (Including Error Proportions) for Individual Participants and by Group

Stage-1 Group							
Part	Correct	S ^a	z	O ^b	z	P ^c	Oth ^e
S1	.55	.10	1.16	.33	1.06	.01	−0.95
S2	.48	.13	1.90	.32	1.03	.03	−0.70
S3	.41	.13	2.05	.35	1.18	.06	−0.44
S4	.27	.09	1.01	.34	1.15	.10	−0.14
S5	.17	.17	3.10	.26	0.73	.08	−0.29
Ave.	.38	.12	1.84	.32	1.03	.06	−0.50

Stage-2 Group							
Part	Correct	S ^a	z	O ^b	z	P ^c	Oth ^e
P1	.72	.01	−1.08	.02	−0.53	.22	0.93
P2	.67	.05	−0.03	.03	−0.50	.22	0.93
P3	.30	.05	−0.03	.15	0.16	.42	2.71
P4	.50	.02	−0.78	.02	−0.53	.42	2.71
P5	.70	.04	−0.33	.04	−0.44	.15	0.37
P6	.75	.03	−0.63	.03	−0.50	.16	0.42
Ave.	.61	.03	−0.48	.05	−0.39	.26	1.35

Note. Part = participant; Philadelphia Naming Test: Roach et al. (1996); z = z scores calculated from the mean and standard deviation of a large diverse group of individuals (N = 107) diagnosed with aphasia resulting from left-hemisphere stroke.

^a Semantically related errors (potentially also phonologically related);

^b Omissions, including circumlocutions and no response errors;

^c Phonologically related (but semantically unrelated) word or nonword error;

^d Perseveration of an earlier response;

^e Other category, composed primarily of picture part responses, unrelated word errors, phonological distortions of semantically related or unrelated word errors, and fragments.

Table 2

Neuropsychological Traits for Individual Participants and by Group

Stage-1 Group												
Part	Mo ^a	Aq ^b	Typ ^c	Rep ^d	z	CaC ^e	z	PNVT ^f	z	Syn ^g	z	Ap ^h
S1	54	79	A	100	0.94	55	-1.38	88	-0.11	57	-1.31	none
S2	10	60	A	95	0.61	61	-0.97	75	-1.02	60	-1.13	none
S3	59	64	W	90	0.29	59	-1.10	78	-0.81	67	-0.71	none
S4	24	44	W	92	0.42	38	-2.53	56	-2.35	37	-2.49	none
S5	54	54	W	91	0.35	53	-1.51	54	-2.49	33	-2.73	none
Ave.	40	60		94	0.52	53	-1.50	70	-1.35	51	-1.67	

Stage-2 Group												
Part	Mo ^a	Aq ^b	Typ ^c	Rep ^d	z	CaC ^e	z	PNVT ^f	z	Syn ^g	z	Ap ^h
P1	34	89	A	74	-0.77	78	0.19	98	0.59	93	0.83	none
P2	55	68	B	79	-0.44	81	0.39	100	0.73	87	0.47	mod to sev
P3	33	56	C	87	0.09	92	1.14	99	0.66	93	0.83	none
P4	24	65	C	65	-1.36	80	0.32	98	0.59	80	0.06	mild
P5	87	71	C	94	0.55	80	0.32	98	0.59	80	0.06	none
P6	11	77	C	66	-1.29	70	-0.36	94	0.31	93	0.83	none
Ave.	41	71		78	-0.54	80	0.33	98	0.58	88	0.51	

Note. Part = participant; z = z scores calculated from the mean and standard deviation of a large diverse group of individuals (N = 107) diagnosed with aphasia resulting from left-hemisphere stroke.

^aMonths post onset;

^bWestern Aphasia Battery quotient (Kertesz, 1982);

^cAphasia type, where A = anomic, W = Wernicke's, B = Broca's, C = Conduction;

^dWord repetition test (in percentages);

^eCamels and Cactus Test (Bozeat et al., 2000), test of nonverbal semantic comprehension (in percentages);

^fPNVT = Semantic subtest of the picture-word verification test (Mirman et al., 2010), a measure of word comprehension (in percentages);

^gSynonym selection task (Martin et al., 2006), a measure of word comprehension (nouns and verbs);

h Speech motor apraxia severity.

Table 3

Lexical Characteristics of Materials with Means (Standard Errors) per Condition

	<u>Homophones</u>		<u>Low Frequency Controls</u>	<u>High Frequency Controls</u>
	<u>Depicted Meaning</u>	<u>All Meanings</u>		
Log Word Frequency (Per Million)	1.18 (0.09)	2.17 (0.16)	1.20 (0.10)	1.92 (0.10)
Phonological Neighborhood Density	21.55 (1.75)		18.62 (1.46)	18.88 (1.72)
Number of Phonemes	3.13 (0.16)		3.15 (0.12)	3.19 (0.14)
Phonotactic Probability	3540 (386)		3157 (270)	3242 (304)
Familiarity [†]	6.98 (0.01)		6.97 (0.01)	6.77 (0.18)
Imageability [†]	6.11 (0.07)		6.20 (0.08)	6.29 (0.08)
Name Agreement [†]	6.45 (0.10)		6.55 (0.06)	6.52 (0.10)

Note.[†] Seven-point scale.

Table 4

Mixed Logit Model Coefficients and Associated Test Statistics

Full Model on Accuracy						
Fixed Effects		β	SE	Z	p	
Stage-1 Group	Intercept	1.52	0.33	4.65	<.001	
	Session	0.49	0.14	3.49	<.001	
	LF controls ^a	−0.51 ^a	0.27	−1.88	.06	
	HF controls ^a	−0.11 ^a	0.28	−0.38	.71	
	Group	−1.94	0.45	−4.28	<.001	
	Lower Order Interaction ^b (LF controls in the Stage-1 group)					
		0.78 ^b	0.35	2.24	.03	
Stage-2 Group	Lower Order Interaction ^b (HF controls in the Stage-1 group)					
			1.03 ^b	0.36	2.85	.004
	Random Effects					
	Participants	σ^2				
		0.40				
	Items	0.49				
Group (By-items)		0.36				
Pairwise Models on Accuracy						
Fixed Effects		β	SE	Z	p	
Stage-1 Group	Intercept	−0.50	0.50	−1.02	.31	
	Session	0.84	0.33	2.54	.01	
HF controls ^a		0.94	0.45	2.08	.04	
Random Effects		σ^2				
		0.74				
Participants		1.99				
Items		β				
Stage-2 Group	Fixed Effects		SE	Z	p	
	Intercept	1.41	0.19	7.57	<.001	
	Session	0.47	0.19	2.40	.02	
LF controls ^a		−0.47	0.22	−2.19	.03	
Random Effects		σ^2				
		0.02				

Full Model on Accuracy				
Fixed Effects				
	β	SE	Z	p
Items	0.13			
Pairwise Model on Omission Errors				
Fixed Effects				
	β	SE	Z	p
Intercept	-0.74	0.33	-2.21	.03
Session	-0.41	0.34	-1.22	.22
HF controls ^a	-0.69	0.32	-2.13	.03
Random Effects				
	s^2			
Participants	0.31			
Items	0.58			
Pairwise Model on Phonological Errors				
Fixed Effects				
	β	SE	Z	p
Intercept	-2.55	0.26	-9.78	<.001
Session	-0.43	0.26	-1.66	.10
LF controls ^a	0.97	0.27	3.61	<.001
Random Effects				
	s^2			
Participants	0.06			
Items	0.00			

Note. Excluding the intercepts, β = model estimation of the change in log odds from the reference category for each fixed effect; SE = standard error of the estimate; Z = Wald Z test statistic; s^2 = Random effect variance.

^aReference is homophone targets condition;

^bReference is homophone targets condition in the Stage-2 group.