Path-Length and the Misperception of Speech: Insights from Network Science and Psycholinguistics

Michael S. Vitevitch, Rutherford Goldstein, and Elizabeth Johnson

Abstract. Using the analytical methods of network science we examined what could be retrieved from the lexicon when a spoken word is misperceived. To simulate misperceptions in the laboratory, we used a variant of the semantic associates task—the phonological associate task—in which participants heard an English word and responded with the first word that came to mind that sounded like the word they heard, to examine what people actually do retrieve from the lexicon when a spoken word is misperceived. Most responses were 1 link away from the stimulus word in the lexical network. Distant neighbors (words >1 link) were provided more often as responses when the stimulus word had low rather than high degree. Finally, even very distant neighbors tended to be connected to the stimulus word by a path in the lexical network. These findings have implications for the processing of spoken words, and highlight the valuable insights that can be obtained by combining the analytic tools of network science with the experimental tasks of psycholinguistics.

1 Introduction

Network analysis has been used to examine semantic (De Deyne and Storms 2008; Hills et al. 2009; Kenett et al. 2011) and phonological (Arbesman et al. 2010; Carlson et al. 2011; Sonderegger 2011) relationships among words in a variety of languages. Although analyses of the structure of networks formed by the semantic and phonological relationships among words have provided unique insights into these languages, it is important to also examine how that observed structure influences language processing (Borge-Holthoefer and Arenas 2010). In the present chapter

Michael S. Vitevitch · Rutherford Goldstein · Elizabeth Johnson
University of Kansas, Lawrence, KS 66045 USA
e-mail: {mvitevit,rgolds}@ku.edu,libbyjohnson54@gmail.com

A. Mehler et al. (eds.), *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, Understanding Complex Systems, DOI: 10.1007/978-3-662-47238-5_2

we used the analytic tools from network analyses and data from a psycholinguistic experiment to explore (1) the structure that exists in the network formed when words serve as vertices (or nodes) and edges (or links) connect words that sound similar to each other (i.e., they are phonologically related) (Vitevitch 2008), and (2) how that structure might influence what is perceived when listeners misperceive a spoken word.

Analysis of speech production errors, such as slips of the tongue, malapropisms, and tip-of-the-tongue experiences, has played an important role in increasing our understanding of the processes involved in speech production (Brown and McNeill 1966; Fay and Cutler 1977; Fromkin 1971). Curiously, however, there has been considerably less research examining speech perception errors, such as mondegreens and slips of the ear (Bond 1999). Instead, most research on speech perception and spoken word recognition has used laboratory-based tasks to examine how certain characteristics of words—such as the frequency with which the word occurs in the language—influence the speed and accuracy that a word can be successfully recognized, with little attention paid to the errors that occurred. The dearth of research on speech perception errors is unfortunate because analyses of such errors have the potential to inform and constrain models of speech perception and spoken word recognition just as similar analyses of speech production errors have informed and constrained models of speech production

There are several models of speech perception and spoken word recognition that have existed for several decades (Luce et al. 2000; Luce and Pisoni 1998; Marslen-Wilson 1987; McClelland and Elman 1986; Norris 1994). However, (to our knowledge) none of these models has been used to predict what will be perceived when a spoken word is misperceived. Given the basic assumptions of these models—multiple word-forms that resemble the acoustic-phonetic input are activated and then compete with each other for recognition—what was perceived when a misperception occurred might have appeared so obvious as to not require any further comment: one of the other partially-activated competitors will be perceived if the word that was actually spoken is not, for some reason, correctly perceived. This simple assumption raises an interesting question, however: what do the partially activated lexical competitors look like?

Of the studies that have examined speech perception errors, most have examined collections of actual perception errors, so-called slips of the ear, as in Bond (1999). Corpus analyses have much ecological validity and have increased our understanding of the spoken word recognition system, but concerns are often raised about the reliability of such data due to the possible influence of perceptual biases in the initial collection of the errors.

In the present study, rather than analyze a corpus of perceptual errors to examine the partially activated lexical competitors that might be erroneously perceived in a slip of the ear, we instead used techniques from network science and a laboratory-based psycholinguistic task. The techniques of network science enabled us to determine the range of lexical competitors that could be perceived as activation diffuses through a network-like representation of the mental lexicon, like the model described in Vitevitch et al. (2011). That is, on average, how many candidates might one have to select from when a misperception occurs (and, to a lesser extent, what do those candidates look like)? The laboratory-based psycholinguistic task allowed us to examine in several ways (and, admittedly, in a somewhat contrived way) what people might actually perceive when a misperception occurs. Both of these approaches provided us with information and insight that could not be examined using the typical method of analyzing a corpus of extent perceptual errors.

2 Network Analysis: What Can Be Perceived When Speech Is Misperceived?

Previous work on slips of the ear suggests that when a misperception of speech occurs, what is perceived is phonologically rather than semantically similar to what was uttered (Bond 1999). But how phonologically similar are the spoken and misperceived words? A commonly used metric to compare the similarity of two strings of characters—in this case, the phonemes in two words—is Levenshtein distance (Levenshtein 1966) (see also Coltheart et al. (1977), Greenberg and Jenkins (1964), Landauer and Streeter (1973), and Luce and Pisoni (1998)). That is, two words are considered phonological neighbors if you can add, delete, or substitute a phoneme in one word to form the other word (i.e., a Levenshtein distance of 1).

This same definition of phonological similarity was used in Vitevitch (2008) to create a network of approximately 20,000 English word-forms. Nodes in the network represented phonological word-forms, and links connected phonological neighbors. As an example, the node for the word *cat* /kæt/ would have a link connecting it to the nodes representing the words *hat* /hæt/, *cut* /kʌt/, *cap* /kæp/, *at* / _æt /, *scat* /skæt/, etc. (the underlined phonemes indicate the location of the change relative to /kæt/). A small portion of this network is illustrated in Fig. 1 showing the word speech, the neighbors of speech, and the neighbors of those neighbors (i.e., the 2-hop neighbors of speech).

The analysis of the phonological network in Vitevitch (2008) showed that 53% (10,265 of 19,340 words) of the words in the network were isolates, or did not have a phonological neighbor; these words, like spinach and obtuse, have been referred to as lexical hermits in Vitevitch (2008) and elsewhere. Furthermore, 13% (2,567 of 19,340 words) of the words in the network formed small components, referred to as lexical islands (Vitevitch 2008), which contained words that were connected to each other, but not to the rest of the network. The remaining 34% of the words in the lexicon (6,508 of 19,340 words) formed what is known as a giant component, or a group of nodes that are connected to each other in some way, but not connected to the other (smaller) components, or to the isolates

In the present analysis, we used a *Hop Plot* to examine the shortest distance (or shortest path length) that exists between two nodes in the network. This distribution of the distances in the network allows us to show the number of nodes (i.e., words) that can be reached by traversing *d* links in the network. The findings in

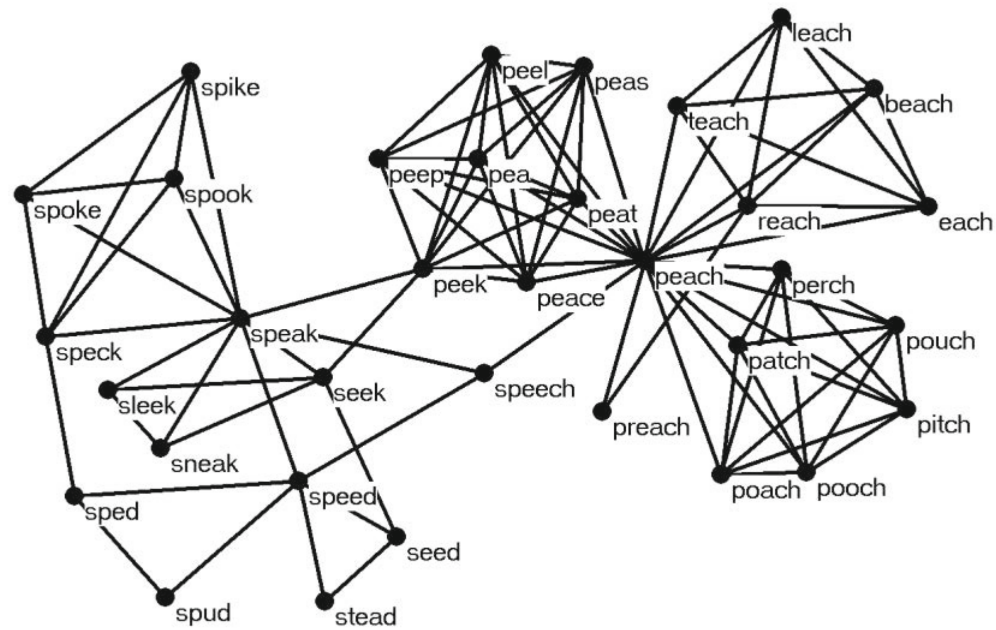


Fig. 1 A small portion of the English phonological network analyzed in Vitevitch (2008). Nodes represent words in the lexicon, and links connect words that are phonologically similar (i.e., they differ by a single phoneme).

Bond (1999) suggest that when listeners misperceive a word, they "hear" something that is phonologically similar to the word that was uttered. We, therefore, used the Hop Plot to determine, on average, how many lexical competitors at different levels of "phonological similarity" (as defined by d, or the number of links traversed) are available to a listener when a misperception of a word occurs. This analysis gives us a better understanding of the possible range of lexical competitors—in terms of number of lexical competitors, and the extent to which they are phonologically similar to the uttered word—that might be erroneously perceived.

Because isolates and nodes in the smaller components (i.e., lexical islands) are, by definition, unreachable, this analysis, as is the convention in network science, only considered the words in the giant component. Our focus on the words in the giant component is reasonable not only for computational reasons (i.e., the distance to unreachable nodes is undefined), but from a theoretical perspective as well. Consider that the distribution of words in a word co-occurrence network fall into a core and a periphery (Dorogovtsev and Mendes 2001). The size of the core or "kernel lexicon" remained relatively invariant as language evolved, and is comparable in size to the giant component we examined in the present study. The existence of lexical islands and lexical hermits in the phonological network (i.e., words in the periphery) raises interesting questions about how such items might be retrieved from the lexicon, but those questions are beyond the scope of the present investigation.

Distance between nodes was assessed in terms of the smallest number of links between the two selected nodes. Recall that in the context of the phonological network, a link corresponds to a single phoneme change (i.e., an addition, deletion, or substitution) between adjacent nodes.

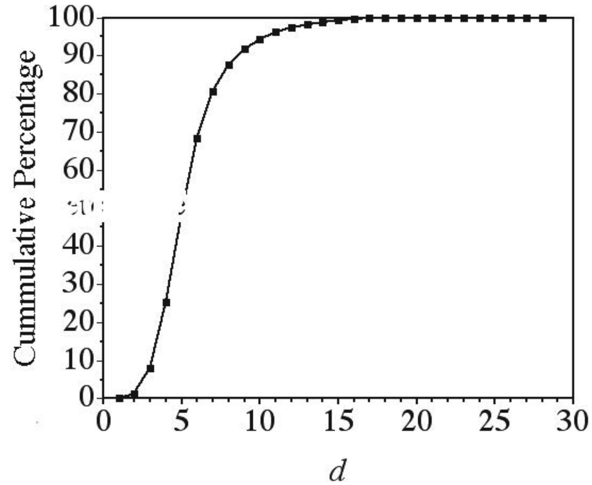
Figure 2 shows the Hop Plot for the 6,508 words in the giant component of the English phonological network examined in Vitevitch (2008). The x-axis is the number of links in the shortest path connecting two nodes. The y-axis is the cumulative percentage of node pairs that are at most d links from each other. Thus, a distance of 1 indicates the percentage of node pairs (i, j) that are reachable by 1-hop, or a distance of 1 link. The longest shortest-path between two nodes in the giant component consisted of 29 links, and exists between the words *connect* and *rehearsal*. The path from the word *connect* to *rehearsal* included the following words (each differing from immediately adjacent words by a single phoneme): *connect*, *collect*, *elect*, *affect*, *effect*, *infect*, *insect*, *inset*, *insert*, *inert*, *inurn*, *epergne*, *spurn*, *spin*, *sin*, *sieve*, *live*, *liver*, *lever*, *leva*, *leaven*, *heaven*, *haven*, *raven*, *riven*, *rivet*, *revert*, *reverse*, *rehearse*, *rehearsal*.

The Hop Plot shows that, on average, .14% of the words in the giant component (or 9.1 of 6,508 words) were reachable by going 1-link away from a given word. Thus, if one were to randomly select a word in the lexicon and change 1 phoneme in that word, one would have, on average, fewer than a dozen competitors (i.e., 9.1 words). If activation were to diffuse through the network to a distance of 2-hops away from a given word, one would activate, on average, 1.33% of the words in the giant component (86.6 of 6,508 words), increasing by an order of magnitude the number of competitors. The number of "phonologically similar" competitors continues to increase dramatically as the distance between words increases. At a distance of 3-hops, 7.9% of the words in the giant component were reachable (or 514.1 words of the 6,508 words in the giant component), and at a distance of 4-hops, 25.2% of the words in the giant component were reachable (or 1,640 words of the 6,508 words in the giant component).

The rapidly increasing number of nodes that can be reached as distance slowly increases visually illustrates one aspect of the small-world phenomenon (Albert et al. 1999; Kleinberg 2000; Watts and Strogatz 1998): despite the large number of items in the system, a large system like the phonological network can nevertheless be traversed quickly. However, the same small-world characteristic that contributes to efficient navigation in a network—being able to reach a large number of nodes very quickly—may lead to detrimental effects when trying to quickly and correctly perceive a spoken word (Luce and Pisoni 1998), or when trying to recover from the misperception of a spoken word. Restricting processing in some way to candidates that are 1 or 2 hops away from a given node may keep the number of competitors to a reasonable number, and may facilitate recovery when misperceptions do occur.

The "double-edged" nature of the path-length between two nodes should not be surprising given the "double-edged" nature of other network characteristics. For example, nodes with many connections, or a high degree (a.k.a. hubs), contribute

Fig. 2 The Hop Plot for the 6,508 words in the giant component of the English phonological network analyzed in Vitevitch (2008). The *x*-axis is *d*, the distance (i.e., number of links) of the shortest path connecting two nodes. The *y*-axis is the cumulative percentage of node pairs that are at most *d* links from each other.



to the stability of scale-free networks when the system is randomly attacked, but can also be the "Achilles heel" of the system when attacks target the hubs (Albert et al. 2000). Also compare the effects of clustering coefficient on speech perception (Chan and Vitevitch 2009) and production (Chan and Vitevitch 2010) to the effects of clustering coefficient on certain memory processes (Vitevitch et al. 2012) for another example of the "double-edged" nature of certain network characteristics.

3 Psycholinguistic Experiment: What Is Perceived When Speech Is Misperceived?

To further examine what might be perceived when a spoken word is misperceived, we used a variant of a well-known psycholinguistic task, the semantic associate task (Nelson et al. 1998), as a laboratory analogue of what happens when one experiences a slip of the ear. In a naturally occurring slip of the ear, a listener hears a word, but does not perceive the word that was uttered. Instead the listener perceives a word that is phonologically similar to the uttered word (Bond 1999).

In our variant of the semantic associate task—the phonological associate task—we presented an English word over a pair of headphones to participants, and asked them to respond with the first word that came to mind that *sounded like* the word they heard (see also Luce and Large (2001)). Note that each participant was allowed to define what "sounded like" meant. By leaving open the meaning of "sounded like," we were able to explore the parameters that listeners may use (implicitly) to define phonological similarity. The responses would also allow us to examine other characteristics of the words that listeners might perceive when they misperceive a word.

Admittedly, this task is contrived, and lacks the ecological validity found in the analysis of naturally occurring slips of the ear. However, the laboratory context of this task enables us to carefully control certain variables, and manipulate others to better explore what might be perceived when a spoken word is misperceived. Such precise control over relevant variables is not possible in analyses of naturally occurring slips of the ear, where one is at the mercy of the material reported in the corpus. For example, an analysis of slips of the ear showed that the words in the corpus that had been misperceived tended to have a higher degree (i.e., more phonological neighbors) than words in general (Vitevitch 2002b). Although this finding was consistent with predictions derived from models of spoken word recognition regarding the difficulty of recognizing spoken words with many phonological neighbors (Luce and Pisoni 1998), the prevalence of naturally occurring slips of ear in words with high degree limits our ability to understand what happens when one misperceives a word with low degree (i.e., few phonological neighbors). Given the prevalence of words with low degree in the network it is important to examine these words as well. Using a laboratory-based psycholinguistic task as in the present experiment, therefore enabled us to examine both types of words, those with high and those with low degree (i.e., many and few phonological neighbors), thereby giving us a more complete understanding of misperceived words than would be possible from an analysis of a corpus of slips of the ear. Finally, our use of the present task gave us the opportunity to obtain usable responses for every stimulus word rather than limiting our analysis to the smaller number of errors that might be obtained in a perceptual identification task, for example.

In psycholinguistic experiments in which a variable—like degree, or the number of phonological neighbors of a word—is manipulated, a prediction regarding performance as a function of that variable is often advanced. In the present case we are using a psycholinguistic task in a more exploratory manner—we wished to simply better understand what might be perceived when a spoken word is misperceived—so we will not advance any specific hypotheses regarding performance. Our inclusion of words with both high and low degree (i.e., many and few phonological neighbors) allows us to explore this question more completely, despite the typical distribution of such items in corpora of naturally occurring slips of the ear (i.e., predominantly words with high degree), and the typical distribution of such items in the lexicon itself (i.e., predominantly words with low degree).

3.1 Method

The same participants, materials, and procedure used in Experiment 2 of Vitevitch et al. (2014) were used in the present investigation. The responses from that experiment were analyzed in a different way in the present investigation. For the convenience of the reader we provide some details regarding the participants, materials, and procedure.

Fourteen native English-speaking students enrolled at the University of Kansas gave their written consent to participate in the present experiment. None of the participants reported a history of speech or hearing disorders.

The materials consisted of 100 English monosyllabic words containing three phonemes in a consonant-vowel-consonant syllable structure. A male native speaker of American English (the first author) produced all of the stimuli by speaking at a normal speaking rate and loudness in an IAC sound attenuated booth into a high-quality microphone. The pronunciation of each word was verified for correctness.

The words differed in degree/neighborhood density, but were similar with regards to a number of other lexical characteristics that are known to influence language processing. *Degree/Neighborhood density* refers to the number of words that sound similar to the stimulus word based on the addition, deletion or substitution of a single phoneme in that word (Luce and Pisoni 1998). A word like *cat*, which has many neighbors (e.g., *at*, *bat*, *mat*, *rat*, *scat*, *pat*, *sat*, *vat*, *cab*, *cad*, *calf*, *cash*, *cap*, *can*, *cot*, *kit*, *cut*, *coat*), has high degree and (in Psycholinguistic terms) is said to have a *dense phonological neighborhood*, whereas a word, like *dog*, that has few neighbors (e.g., *dig*, *dug*, *dot*, *fog*) has low degree and (in Psycholinguistic terms) is said to have a sparse phonological neighborhood (N.B., each word has additional neighbors, but only a few were listed for illustrative purposes). Half of the stimuli had high degree/dense phonological neighborhoods (*mean* = 27.7 neighbors, *sd* = 1.6), and the remaining stimuli had low degree/sparse phonological neighborhoods (*mean* = 14.9 neighbors, *sd* = 1.5; F (1, 98) = 1648.62, $p < .0001$).

Although the stimuli differed in degree/neighborhood density, they were comparable with regards to the following characteristics. *Subjective familiarity ratings* of the words, measured on a seven-point scale, were obtained from Nusbaum et al. (1984). Words with high degree/dense neighborhoods had a mean familiarity value of 6.87 (*sd* = .22) and words with low degree/sparse neighborhoods had a mean familiarity value of 6.82 (*sd* = .28, F (1, 98) = 1.50, p = .22). The mean familiarity value for the words in the two groups indicates that all of the words were highly familiar.

The mean *frequency of occurrence* in the language ($\log_{10}$ of the raw values from Kučera and Francis (1967)) was 1.03 (*sd* = .58) for the words with high degree/-dense neighborhoods, and 1.00 (sd = .58) for the words with low degree/sparse neighborhoods (F (1, 98) = .08, p = .77).

Neighborhood frequency is the mean word frequency of the neighbors of the target word. Words with high degree/dense neighborhoods had a mean log neighborhood frequency value of 2.03 (*sd* = .24), and words with low degree/sparse neighborhoods had a mean log neighborhood frequency value of 1.94 (*sd* = .25; F (1, 98) = 2.99, p = .09).

Phonotactic probability refers to how often a certain segment occurs in a certain position in a word (*positional segment frequency*) and to how often two adjacent segments occur next to each other in a certain position (*biphone frequency*; as in Vitevitch and Luce (2005)). The mean positional segment frequency for words with high degree/dense neighborhoods was .147 (*sd* = .02) and for words with low degree/sparse neighborhoods was .140 (*sd* = .02, F (1, 98) = 2.11, p = .15). The mean biphone frequency for words with high degree/dense neighborhoods was .007 (*sd* = .003) and for words with low degree/sparse neighborhoods was .007 (*sd* = .003, F

(1, 98) = .009, $p = .93$). These values were obtained from a web-based phonotactic probability calculator (Vitevitch and Luce 2004).

Each participant was seated in front of a computer that controlled the presentation of stimuli and the collection of responses. In each trial, the word "READY" appeared on the computer screen for 500 ms. Participants then heard one of the randomly selected stimulus words through a set of headphones at a comfortable listening level. Each stimulus was presented only once. Participants were asked to type in the first English word that came to mind that "sounded like" the word that they heard over the headphones. The participants could use as much time as they needed to respond. Participants were able to see their responses on the computer screen when they were typing and could make corrections to their responses before they hit the RETURN key, which initiated the next trial. Although different effects might be found when a closed-response-set rather than an open-response-set is used, there does not appear to be any difference in performance depending on whether responses are spoken versus typed in tasks like that used in the present experiment (Clopper et al. 2006).

3.2 Results

Misspelled words and typographical errors in the responses were corrected to form English words according to the following criteria: (1) transposition of adjacent letters in the word was corrected, and (2) the addition of a single letter in the word was removed if the letter was within one key of the target letter on the keyboard. Of the 1400 responses, 4.56% were misspellings or typographical errors that could not be resolved into English words according to the criteria above, were semantically but *not* phonologically related to the stimulus, or were repetitions of the stimulus word. These responses could not be analyzed, leaving 1336 responses for examination.

Of the responses that we could analyze, 1125 (84.21% of the 1336 responses) were 1 link away from the stimulus word. That is, the responses differed from the stimulus word by one phoneme. We found 181 responses (13.54% of the 1336 responses) that were 2 links away from the stimulus word (i.e., differing from the stimulus word by two phonemes), 28 responses (2.1% of the 1336 responses) that were 3 links away from the stimulus word, 1 response (.07% of the 1336 responses) that was 6 links away from the stimulus word, and 1 response (.07% of the 1336 responses) that was 8 links away from the stimulus word. Thus, when asked to produce a word that "sounded like" a given word, listeners overwhelmingly selected a word that was a short path-length in the network of phonological word-forms away from the stimulus word, and only occasionally selected words at longer path-lengths from the stimulus, giving us additional insight into the criteria that typical users of language (rather than trained language scientists) employ to define phonological similarity.

We further examined in several ways the 1125 responses that differed from the stimulus word by 1 link as a function of degree/neighborhood density of the stimulus word. Our first analysis of these words examined how many different words were given in response to a stimulus word. That is, when presented with *cat*, did everyone give *hat* as the response, or was there some variety in the words that "sounded like" the word *cat*?

The 14 participants gave a mean of 9.16 different words (sd = 1.74) in response to words with a high degree/dense phonological neighborhood, and 8.66 different words (sd = 2.03) in response to words with a low degree/sparse phonological neighborhood. The analysis of the path-length between words indicated that there was remarkable consensus among participants regarding what "sounded like" the stimulus word: a word that was 1 link away from the stimulus word. Despite that agreement, the present analysis suggests that participants did not converge on the same path in the lexical network. Participants instead indicated that a variety of words in the phonological neighborhood (regardless of whether it was a dense or a sparse neighborhood) "sounded like" the stimulus word.

It is striking that the number of different words that participants indicated "sounded like" the stimulus word approximates the value of 9.1 obtained in the Hop Plot in Fig. 2 for the average number of words that could be reached by 1 hop in the network of phonological word-forms. Future research could explore whether the recurrence of this value is simply a coincidence, or is indicative of some sort of cognitive constraint on language processing, such as the well-known constraint in short-term memory of 7 plus or minus 2 chunks (Miller 1956). One way to distinguish between these two possibilities is to increase the number of respondents in this task. If, with additional participants, we obtain even more variety in the number of different words that "sounded like" the stimulus word, then we can rule out the possibility that the value of 9 is indicative of some sort of cognitive constraint on language processing. If that value is again observed, then additional investigation of some sort of cognitive constraint may be warranted.

A second analysis examined the percentage of responses that differed from the stimulus word by 1 link *as a function of degree/neighborhood density of the stimulus word*. For the stimulus words with high degree/dense phonological neighborhoods we found that 84.86% (sd = 12.13) of the responses given to these words were 1 link away from the stimulus word (meaning that 15.14% of the responses were more than 1 link away from the stimulus word), and for stimulus words with low degree/sparse phonological neighborhoods 76.43% (sd = 16.78) of the responses given to these words were 1 link away from the stimulus word (meaning that 23.57% of the responses were more than 1 link away from the stimulus word). This difference was statistically significant (t (98) = 2.88, $p < .01$), and on the one hand is not surprising. That is, words with low degree/sparse phonological neighborhoods have few words that "sounded like" the stimulus word that are 1 link away, so activation may diffuse across longer paths (i.e., two or more links) to activate a word that "sounded like" the stimulus.

On the other hand, however, the finding that a smaller proportion of stimulus words with low degree/sparse phonological neighborhoods had responses that were 1 link away from the stimulus word is peculiar, and raises additional questions. For example, consider this result in conjunction with the previous finding regarding the number of different words that participants indicated "sounded like" the stimulus word. If words with low degree/sparse phonological neighborhoods have few options to choose amongst for words that "sounded like" the stimulus word, then why was there variability in the number of different words that participants indicated "sounded like" the stimulus word? That is, why did participants give 2-hop neighbors as responses instead of simply producing the same 1-hop neighbors again and again (and therefore producing a smaller number of different types of words that "sounded like" the stimulus word for the words with low degree/sparse phonological neighborhoods)? This returns us to the provocative hypothesis that there may be some sort of cognitive constraint on language processing: during spoken word recognition a fixed number of candidates may be evaluated by the word recognition system. In the case of words with high degree/dense phonological neighborhoods, that fixed number of candidates may be reached (or exceeded) by 1-hop neighbors. Whereas in the case of words with low degree/sparse phonological neighborhoods, that fixed number of candidates may be reached only by considering more distant phonological neighbors (i.e., words more than 1-hop away). Additional analyses and psycholinguistic experiments may be warranted to examine further this speculative hypothesis.

Our next analysis examined the 211 responses that were more than 1 link away from the stimulus word. Given the insight provided by the Hop Plot, we again turned to the tools of network science, and examined the phonological network analyzed in Vitevitch (2008) to determine if a connected path of words existed between the stimulus word and the more distant responses. To illustrate (see Fig. 1), imagine spud was the stimulus, and the response was beach; one can get from *spud* to *beach* by going through the words *speed-speech-peach*, a path length of 4 links.

In 205 of the 211 cases (97.16%) there existed a path of words between the stimulus and the response. The 6 (2.84%) exceptions to this were (stimulus word $\rightarrow$ response): *lag* $\rightarrow$ *stagnant*, *niche* $\rightarrow$ *kitchen*, *poach* $\rightarrow$ *potion*, *poach* $\rightarrow$ *approach*, *noose* $\rightarrow$ *caboose*, and *bib* $\rightarrow$ *bibliography*. Note that the network analyzed in Vitevitch (2008) contained fewer than 20,000 words. If a larger network were analyzed—one that approached the higher estimates of vocabulary size offered by some (e.g., 216,000 words (Diller 1978))—it is possible that a path might be found between the stimulus and the response in these 6 cases as well.

Despite these 6 exceptions (less than .5% of the 1400 responses) the result of this analysis suggests that words that "sounded like" each other—even distantly related words—tend to connect to each other along a path of real words in the lexicon. The existence of lexical intermediaries observed in the present analysis raises some concerns about measures of word-form similarity that ignore such items, such as the Orthographic Levenshtein Distance-20 (OLD-20 (Yarkoni et al. 2008)), and the Phonological Levenshtein Distance-20 (PLD-20 (Suárez et al. 2011)). In OLD-20/PLD-20, Levenshtein distance is computed between a target word and all other words in the lexicon. OLD-20/PLD-20 is then the mean edit distance of the 20 closest neighbors. The computations of OLD-20/PLD-20 do not consider whether real-word intermediaries exist or not; the measure only considers the number of letter/phoneme changes (respectively) that are required to turn one word into another. However, the present findings show that distant phonological neighbors tend to be connected to a word via a path of real words, raising questions about the psychological validity of metrics such as OLD-20 and PLD-20 that do not consider the absence (or existence) of lexical intermediaries.

4 Conclusion

In the present chapter we used analytical tools from network science and experimental methods from psycholinguistics to examine a question about language processing that is less often examined: What is perceived when a spoken word is misperceived? A Hop Plot was used to assess the proportion of nodes that can be reached (on average) at a given distance, thereby providing us with information about the number of "phonologically similar" competitors one might expect to consider as activation diffuses across the network. This analysis revealed that a relatively small proportion of the network (.14% or 9.1 of 6,508 words) could be reached via 1 link. However, the proportion of words that could be reached increased dramatically as the number of links traversed increased.

With the information provided by our network analysis about how many candidates one might choose amongst when one misperceives a spoken word, we turned to the question of what those candidates actually look like, and examined that question with the phonological associate task, in which participants heard a word and responded with the first word that came to mind that "sounded like" that word. Although this task is admittedly somewhat artificial, it does mimic certain important aspects of the processes that are used "in the wild" to recover from the misperception of spoken words. Furthermore, the ability to carefully select certain words to use as stimuli enabled us to examine certain variables while controlling for other variables, which is something that cannot be done easily when analyzing a corpus of speech perception errors. Moreover, our ability to manipulate the variable of degree/neighborhood density allowed us to examine what happens when misperceptions occur in words with low degree/sparse neighborhoods; this is not possible in analyses of extent speech perception errors because such words rarely appear in such corpora (Vitevitch 2002b).

Several interesting results were observed in the phonological associate task: (1) most responses were 1 link away from the stimulus word, (2) responses that were more distant (>1 link away from the stimulus word) tended to occur more for words with low degree/sparse neighborhoods than for words with high degree/dense phonological neighborhoods, and (3) responses that were more distant tended to be connected to the stimulus word by a path of real words in the lexicon.

The observation that most responses were 1 link away from the stimulus word provides important insight into the criteria that listeners use to indicate that two words "sounded like" each other. Other logical and linguistically motivated possibilities exist, including responding with a longer word that contained the stimulus word (e.g., *cat* → *catalog*), or appending various morphemes to the stimulus word (e.g., *dog* → *doggedly*), but such alternative responses were quite rare in the present study.

The observation that distant responses (>1 link away from the stimulus word) tended to occur more for words with low degree/sparse neighborhoods than for words with high degree/dense phonological neighborhoods is also interesting, especially in light of the first observation. If most responses are 1 link away, one might expect that participants would have more consistency amongst themselves in identifying words that "sounded like" the stimulus words with low degree/sparse neighborhoods. That is, most participants should have provided the same word as a response to a given stimulus word instead of the wide variety of responses that was observed for each stimulus word. The fact that listeners instead went beyond the 1-hop neighbors even though there were still words to choose from—recall the mean number of neighbors for the stimuli with low degree/sparse phonological neighborhoods was 14.9 neighbors—is interesting, and opens up several new avenues for future research, including the hypothesis that a fixed number of candidates might be evaluated during spoken word recognition.

Another interesting avenue for future research is to examine the amount of time it takes to recover from the misperception of a spoken word. Unfortunately we did not measure the time to respond in the present study. Had we done so we could have compared the response times of the items that were 1-2 hops away from the stimulus to the response times of the items that were more than 2 hops away from the stimulus. Future experiments that compare a free-response condition in the phonological associate task to a condition with an imposed time-pressure to respond could shed light on the mechanisms that may be employed to recover from the misperception of a spoken word (De Deyne et al. 2012).

The present results also highlight the existence of lexical intermediates and the potential importance they may play in certain language-related processes. In the responses that were 2 or more hops away from the stimulus word, 97.16% of the responses had a path of extent words connecting the response to the stimulus. Recent work using a game called word-morph—in which participants were given a word, and asked to form a disparate word by changing one letter at a time—demonstrates that participants can exploit their knowledge of the paths between words to efficiently traverse large distances in a lexical network (Iyengar et al. 2012). For example, when asked to "morph" the word *car* into the word *shy* participants might have changed *car* into *cat-pat-pet-set-see-she* and finally into *shy*. Once participants in this task identified certain "landmark" words in the lexicon, the task of navigating from one word to another became trivial, enabling the participants to solve subsequent word-morph puzzles very quickly; solving times dropped from 10-18 minutes in the first 10 games, to about 2 minutes after playing 15 games, to about 30 seconds after playing 28 games. The results of the present study suggest that lexical intermediaries may also play a role in the misperception of a spoken word.

Another recent study further highlights the importance of intermediate lexical items (Geer and Luce 2012). In an auditory shadowing task and a lexical decision task distant neighbors (i.e., words 2 links away from the target) inhibited lexical intermediaries (i.e., words 1 link away from the target), thereby reducing the amount of inhibition that the target word receives from those intermediaries. Referring to Fig. 1, if *speech* is the target word, the word *spud* would inhibit the word *speed*, the word *beach* would inhibit *peach*, etc., thereby reducing the amount of inhibition that *speech* receives from *speed*, *peach*, etc. Said another way, the words that inhibit a target word are themselves inhibited by other words. Thus, the number of distant neighbors can influence retrieval of a target word by moderating the influence that near neighbors have on the target word (Geer and Luce 2012). The findings from the present study together with the findings from the word-morph game and the findings in Geer and Luce (2012) indicate that additional research on the role of lexical intermediaries on processing is warranted.

More broadly speaking, the present chapter illustrates how network science can be used to investigate questions related to complex *cognitive* systems, in addition to questions related to complex social, biological, or technological systems, areas typically analyzed by network scientists (Albert and Barabási 2002). Combining the power of laboratory-based experiments that are often used in the psychological sciences with the analytical tools and system-wide view of network science—as in the present chapter—holds much promise for advancing the psychological sciences into new areas of inquiry and for resolving ongoing debates. This approach has already increased our understanding of the brain (Sporns 2010), as well as the cognitive processes involved in human navigation (Iyengar et al. 2012), semantic memory (Hills et al. 2009; Marslen-Wilson 1987), and human collective behavior (Goldstone et al. 2008).

In the context of spoken language processing, the tools of network science have enabled us to measure the global as well as the local structure of words stored in the mental lexicon. Previous attempts to examine the structure of the lexicon have only focused on one level. Consider the work of Zipf (Zipf 1935), which found (among other things) a power-law relationship between the frequency with which a word occurs and its rank order. Consider other analyses (Baayen 1991; Baayen 2001; Frauenfelder et al. 1993; Landauer and Streeter 1973), which examined how certain lexical characteristics, such as word-frequency or phoneme frequencies, were related to other lexical characteristics, such as neighborhood density. Consider further the work on neighborhood spread (Vitevitch 2007), onset density (Vitevitch 2002a), and phonotactic probability (Vitevitch and Luce 2005). We see these and many other studies as attempts to measure some aspect of the structural relations among words in the lexicon with the statistical tools that were available at the time. Each of these attempts captured some aspect of that lexical structure, but only at one level of the system. Network science offers a more complete set of methodological tools that can be used to examine multiple levels of a system.

More important, network science offers a theoretical perspective that integrates the observations made at each level of the system. Previous observations of the structure of the lexicon were not only limited to one level of the system, but were often viewed as disparate findings instead of being cumulative, complementary, or somehow connected. That is, each of these previous findings provided yet another entry to the already long list of lexical variables that were known to influence processing in some way (Cutler 1981), instead of contributing to a cohesive description of the lexical system. We believe that the methods and theory of network science offer psychological scientists a unique and powerful framework to develop comprehensive models of cognitive processes and representations that can then be subjected to empirical tests. The present chapter serves as an example of how to combine the analytic tools of network science with the experimental tasks of psychology to examine (and raise) new questions about cognitive processing and representation.

References

Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Review of Modern Physics 74, 47–97 (2002)

Albert, R., Jeong, H., Barabási, A.L.: Diameter of the World-Wide-Web. Nature 401, 130–131 (1999)

Albert, R., Jeong, H., Barabási, A.L.: The Internet's Achilles' Heel: Error and attack tolerance of complex networks. Nature 406, 200 (2000)

Arbesman, S., Strogatz, S.H., Vitevitch, M.S.: The Structure of Phonological Networks Across Multiple Languages. International Journal of Bifurcation and Chaos 20, 679–685 (2010)

Baayen, R.H.: A stochastic process for word frequency distributions. In: Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (1991)

Baayen, R.H.: Word Frequency Distributions. Kluwer, Dordrecht (2001)

Bond, Z.S.: Slips of the Ear: Errors in the Perception of Casual Conversation. Academic Press (1999)

Borge-Holthoefer, J., Arenas, A.: Semantic Networks: Structure and Dynamics. Entropy 12, 1264–1302 (2010)

Brown, R., McNeill, D.: The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior 5, 325–337 (1966)

Carlson, M.T., Bane, M., Sonderegger, M.: Global Properties of the Phonological Networks in Child and Child-Directed Speech. In: Danis, N., Mesh, K., Sung, H. (eds.) Proceedings of the 35th Annual Boston University Conference on Language Development. BUCLD, vol. 35, pp. 97–109 (2011)

Chan, K.Y., Vitevitch, M.S.: The Influence of the Phonological Neighborhood Clustering-Coefficient on SpokenWord Recognition. Journal of Experimental Psychology: Human Perception & Performance 35, 1934–1949 (2009)

Chan, K.Y., Vitevitch, M.S.: Network structure influences speech production. Cognitive Science 34, 685–697 (2010)

Clopper, C.G., Pisoni, D.B., Tierney, A.T.: Effects of Open-Set and Closed-Set Task Demands on Spoken Word Recognition. Journal of the American Academy of Audiology 17, 331–349 (2006)

Coltheart, M., Davelaar, E., Jonasson, J.T., Besner, D.: Access to the internal lexicon. In: Dornic, S. (ed.) Attention and Performance VI, pp. 535–556. Academic Press, New York (1977)

Cutler, A.: Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? Cognition 10, 65–70 (1981)

De Deyne, S., Navarro, D.J., Perfors, A., Storms, G.: Strong structure in weak semantic similarity: A graph based account. In: Proceedings of the 34th Annual Conference of the Cognitive Science Society, Sapporo, Japan (2012)

De Deyne, S., Storms, G.: Word association: Network and Semantic properties. Behavior Research Methods 40, 213–231 (2008)

Diller, K.C.: The language teaching controversy. Newbury House, Rowley (1978)

Dorogovtsev, S.N., Mendes, J.F.F.: Language as an evolving word web. Proceedings of the Royal Society B: Biological Sciences 268, 2603–2606 (2001)

Fay, D., Cutler, A.: Malapropisms and the structure of the mental lexicon. Linguistic Inquiry 8, 505–520 (1977)

Frauenfelder, U.H., Baayen, R.H., Hellwig, F.M., Schreuder, R.: Neighborhood density and frequency across languages and modalities 32, 781–804 (1993)

Fromkin, V.A.: The Non-Anomalous Nature of Anomalous Utterances. Language 47, 27–52 (1971)

Geer, M., Luce, P.A.: Neighbors of Neighbors in SpokenWord Processing (2012) (unpublished manuscript submitted for publication)

Goldstone, R.L., Roberts, M.E., Gureckis, T.M.: Emergent processes in group behavior. Current Directions in Psychological Science 17, 10–15 (2008)

Greenberg, J.H., Jenkins, J.J.: Studies in the psychological correlates of the sound system of American English. Word 20, 157–177 (1964)

Hills, T.T., Maouene, M., Maouene, J., Sheya, A., Smith, L.: Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? Psychological Science 20, 729–739 (2009)

Iyengar, S.R.S., Madhavan, C.E.V., Zweig, K.A., Natarajan, A.: Understanding Human Navigation Using Network Analysis. Topics in Cognitive Science 4, 121–134 (2012)

Kenett, Y.N., Kenett, D.Y., Ben-Jacob, E., Faust, M.: Global and local features of semantic networks: Evidence from the Hebrew mental lexicon. PLoS One 6, e23912 (2011)

Kleinberg, J.M.: Navigation in a small world. Nature 406, 845 (2000)

Kuçera, H., Francis, W.N.: Computational Analysis of Present Day American English. Brown University Press, Providence (1967)

Landauer, T.K., Streeter, L.A.: Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. Journal of Verbal Learning & Verbal Behavior 12, 119–131 (1973)

Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady 10, 707 (1966)

Luce, P.A., Goldinger, S.D., Auer Jr., E.T., Vitevitch, M.S.: Phonetic priming, neighborhood activation, and PARSYN. Perception and Psychophysics 62, 615–625 (2000)

Luce, P.A., Large, N.R.: Phonotactics, density and entropy in spoken word recognition. Language and Cognitive Processes 16, 565–581 (2001)

Luce, P.A., Pisoni, D.B.: Recognizing spoken words: The neighborhood activation model. Ear and Hearing 19, 1–36 (1998)

Marslen-Wilson, W.D.: Functional parallelism in spoken word-recognition. Cognition 25, 71–102 (1987)

McClelland, J.L., Elman, J.L.: The TRACE model of speech perception. Cognitive Psychology 18, 1–86 (1986)

Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63, 81–97 (1956)

Nelson, D.L., McEvoy, C.L., Schreiber, T.A.: The University of South Florida word association, rhyme, and word fragment norms (1998), `http://www.usf.edu/FreeAssociation`

Norris, D.: Shortlist: A connectionist model of continuous speech recognition. Cognition 52, 189–234 (1994)

Nusbaum, H.C., Pisoni, D.B., Davis, C.K.: Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. Research on Speech Perception Progress Report 10, 357–376 (1984)

Sonderegger, M.: Applications of graph theory to an English rhyming corpus. Computer Speech and Language 25, 655–678 (2011)

Sporns, O.: Networks of the brain. MIT Press (2010)

Suárez, L., Tan, S.H., Yap, M.J., Goh, W.D.: Observing neighborhood effects without neighbors. Psychonomic Bulletin and Review 18, 605–611 (2011)

Vitevitch, M.S.: Influence of onset density on spoken-word recognition. Journal of Experimental Psychology: Human Perception and Performance 28, 270–278 (2002a)

Vitevitch, M.S.: Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. Language and Speech 45, 407–434 (2002b)

Vitevitch, M.S.: The spread of the phonological neighborhood influences spoken word recognition. Memory & Cognition 35, 166–175 (2007)

Vitevitch, M.S.: What can graph theory tell us about word learning and lexical retrieval? Journal of Speech Language Hearing Research 51, 408–422 (2008)

Vitevitch, M.S., Chan, K.Y., Goldstein, R.: Insights into failed lexical retrieval from network science. Cognitive Psychology 68, 1–32 (2014)

Vitevitch, M.S., Chan, K.Y., Roodenrys, S.: Complex network structure influences processing in long-term and short-term memory. Journal of Memory & Language 67, 30–44 (2012)

Vitevitch, M.S., Ercal, G., Adagarla, B.: Simulating retrieval from a highly clustered network: Implications for spoken word recognition. Frontiers in Psychology 2, 369 (2011), doi:10.3389/fpsyg.2011.00369

Vitevitch, M.S., Luce, P.A.: A web-based interface to calculate phonotactic probability for words and nonwords in English. Behavior Research Methods, Instruments, and Computers 36 (2004)

Vitevitch, M.S., Luce, P.A.: Increases in phonotactic probability facilitate spoken nonword repetition. Journal of Memory & Language 52, 193–204 (2005)

Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. Nature 393, 409–410 (1998)

Yarkoni, T., Balota, D., Yap, M.J.: Moving beyond Coltheart’s N: A new measure of orthographic similarity. Psychonomic Bulletin & Review 15, 971–979 (2008)

Zipf, G.K.: The Psychobiology of Language. Houghton-Mifflin (1935)