What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval?

Michael S. Vitevitch
University of Kansas, Lawrence

Purpose: Graph theory and the new science of networks provide a mathematically rigorous approach to examine the development and organization of complex systems. These tools were applied to the mental lexicon to examine the organization of words in the lexicon and to explore how that structure might influence the acquisition and retrieval of phonological word-forms.
Method: Pajek, a program for large network analysis and visualization (V. Batagelj & A. Mvrar, 1998), was used to examine several characteristics of a network derived from a computerized database of the adult lexicon. Nodes in the network represented words, and a link connected two nodes if the words were phonological neighbors.
Results: The average path length and clustering coefficient suggest that the phonological network exhibits small-world characteristics. The degree distribution was fit better by an exponential rather than a power-law function. Finally, the network exhibited assortative mixing by degree. Some of these structural characteristics were also found in graphs that were formed by 2 simple stochastic processes suggesting that similar processes might influence the development of the lexicon.
Conclusions: The graph theoretic perspective may provide novel insights about the mental lexicon and lead to future studies that help us better understand language development and processing.

KEY WORDS: graph theory, mental lexicon, phonological neighborhood

In the present analysis, graph theoretic techniques were used to examine the complex cognitive system known as the mental lexicon. *Graph theory* is a branch of mathematics used by physicists, computer scientists, and other researchers interested in the "new" science of networks (Watts, 2004) to study the structure of a diverse array of complex systems in the real world. *Complex systems* are comprised of a large number of individual units that interact in relatively simple ways. Despite the simple, predictable nature of the individual components on a local level, these large systems often exhibit behaviors that may appear unpredictable at a global level.

Although graphs simply describe the structure of a system, the way in which a system is organized has important implications for the type of processing that can be carried out in that system (Strogatz, 2001; Ward, 2002). For example, Montoya and Solé (2002) used the techniques of graph theory to create a *graph* (or *network*) of various ecosystems by representing the animals in an ecosystem as *nodes* (sometimes called *vertices* or *actors*) and the predator–prey relationship between animals as *links* (sometimes called *edges* or *ties*) to examine how the extinction of a given species might affect the rest of the ecosystem. The present graph theoretic analyses examined the organization of phonological word-forms in the adult lexicon to explore how the structure of the mental lexicon might influence the process of lexical retrieval.

The current structure of a system has been influenced by (among other things) the growth and development of that system. Certain constraints on development may result in the emergence of one type of structure but not in the emergence of some other type of structure. By studying the current structure of a system, one might gain insight into the constraints that could have led to that type of structure. Although several different constraints may produce the same final structure, analyses of the current structure do serve to rule out some possible developmental mechanisms. Therefore, structural analyses employing graph theoretic techniques may help other research endeavors to focus their investigations on a more reasonably sized search space of possible developmental mechanisms. The present graph theoretic analyses examined the current structure of phonological word-forms in the adult lexicon to also gain some insight on the constraints that may influence lexical acquisition and word learning.

Some readers might be familiar with the terms *node, link,* and *network* in the context of connectionist models, or artificial neural networks. However, it is important to note that the networks discussed in the present work are not artificial neural networks. That is, the nodes in the present networks do not have activation states or rules to change activation states like the nodes in an artificial neural network. Furthermore, the links in the present networks do not have weights associated with them or a learning rule (e.g., back-propagation of error) to change the connection weights between nodes. In other words, the present networks are not artificial neural networks that model a cognitive process. Rather, the present investigation employed graph theoretical analysis techniques to examine the global structure or organization of phonological word-forms that might appear in the mental lexicon of a typically developing adult. Although increased understanding of the structure of phonological word-forms in the mental lexicon might lead to insights related to the processing of phonological word-forms during the perception or production of speech and to insights related to the mechanisms that might influence word learning, it is important to emphasize that the present graph theoretical analysis is not a model of any language process.

Several researchers have used graph theoretic techniques to examine semantic relationships among words in the mental lexicon. Similar conclusions about the structure of semantically related words were reached even though several different definitions of "semantically related" were used, including words that were free associates of each other, synonym pairs, core words from dictionary definitions, or co-occurring words in text (Albert & Barabási, 2002; Batagelj, Mrvar, & Zaveršnik, 2002; Ferrer i Cancho & Solé, 2001b, Motter, de Moura, Lai, & Dasgupta, 2002; Steyvers & Tenenbaum, 2005; Wilks &

Meara, 2002; Wilks, Meara, & Wolter, 2005). Given the arbitrary relationship between semantic and phonological representations (Saussure, 1916/1966), it is unclear whether the same network structure found in a network of semantic representations will be found in a network of phonological word-forms.

In the present graph theoretic analysis, the nodes in the network represented phonological word-forms in the English language (i.e., lexemes), and a link was placed between two nodes if they were phonological neighbors of each other (e.g., Luce & Pisoni, 1998). That is, the links in the present network were not directed and were not weighted. More complex relationships among nodes can be modeled by incorporating directed or weighted links. In a network with *directed links* (sometimes called *arcs*), connections between two nodes are not necessarily reciprocal. For example, X may buy a good from Y, but Y does not buy a good from X. See Harary, Norman, and Cartwright (1965) for an introduction to graphs with directed links. In a graph with *weighted links,* the relationship between some nodes might be stronger or weaker than the relationship between other nodes. For example, X and Y might be lifelong friends, so the link between them might be weighted with a 1. In contrast, X and Z may simply be acquaintances, so the link between them might be weighted with a .25. See Roberts (1976) for an introduction to graphs with weighted links.

Although more complex networks can be constructed, the present investigation analyzed simple network models to explore the most fundamental characteristics that might shape the mental lexicon. Some of the characteristics that shape the development of the lexicon might be common to other types of systems that have been explored with these techniques, whereas other characteristics may be unique to the lexicon. By starting out with a simple yet mathematically precise model, we can explore how well a few basic mechanisms can account for the structure found in the mental lexicon and, perhaps, shed some light on additional assumptions that may be required to account for the structure found in the mental lexicon.

Method

To represent the adult lexicon, the phonological transcriptions of approximately 20,000 words (*N* = 19,340) were examined with Pajek, a program for large network analysis and visualization (Batagelj & Mrvar, 1998). The sample of words was obtained from the 1964 *Merriam-Webster Pocket Dictionary*. Although the entries were those from the *Merriam-Webster Pocket Dictionary,* the pronunciations were derived, checked, and edited by several researchers at the Massachusetts Institute of Technology, including Dennis Klatt, Dave Shipman, Meg

Withgott, and Lori Lamel. Numerous studies have used this same sample of words to derive estimates of neighborhood density and phonotactic probability in English (e.g., Luce & Pisoni, 1998; Nusbaum, Pisoni, & Davis, 1984; Vitevitch & Luce, 2004, 2005). Although the number of words in this sample is slightly larger than the 17,000 base words that comprise the vocabulary of a well-educated adult native speaker of English (Goulden, Nation, & Read, 1990), it constitutes a reasonable approximation of the adult lexicon.

A node in the network corresponded to each of the phonological representations from the *Merriam-Webster Pocket Dictionary*. A link was placed between two nodes if the two words were phonologically similar (cf. Batagelj et al., 2002, in which a link connected two nodes if the words were semantically related). *Phonological similarity* was operationally defined by substituting, adding, or deleting a single phoneme in a given word to form a "phonological neighbor" (e.g., Greenberg & Jenkins, 1967; Landauer & Streeter, 1973; Luce & Pisoni, 1998). For example, the words *hat, cut, cap, scat,* and *_at* were considered phonologically similar to the word *cat* (*cat* has other words as neighbors, but only a few were listed for illustrative purposes). In the present network, links would connect the nodes that corresponded to those words to the word *cat* (and to each other as appropriate).

The same definition of phonological similarity has been used in numerous psycholinguistic studies. The results of these studies have shown that the number of phonological neighbors activated in the lexicon influences various language processes: (a) the *acquisition of sounds* in children (Gierut, Morrisette, & Champion, 1999); (b) the *acquisition of words* in children (Charles-Luce & Luce, 1990, 1995; Coady & Aslin, 2003; Dollaghan, 1994; Storkel, 2004; Vicente, Castro, & Walley, 2003); (c) *spoken word recognition* in young adults with no history of speech, language, or hearing impairment in English and in Spanish (Luce & Pisoni, 1998; Vitevitch, 2002b; Vitevitch & Luce, 1998, 1999; Vitevitch & Rodriguez, 2005), in older adults with no history of speech, language, or hearing impairment (e.g., Sommers, 1996), and in postlingually deafened adults who use a cochlear implant (Kaiser, Kirk, Lachs, & Pisoni, 2003), and of accented speech (Imai, Walley, & Flege, 2005); (d) *spoken word production* in children who stutter (Arnold, Conture, & Ohde, 2005), in young adults with fluent speech in English and in Spanish (Munson & Solomon, 2004; Vitevitch, 1997, 2002a; Vitevitch & Stamer, 2006), in older adults with fluent speech (Vitevitch & Sommers, 2003), and in individuals with aphasia (Gordon & Dell, 2001); and even (e) *reading* by young adults with no history of speech, language, or hearing impairment (Yates, Locker, & Simpson, 2004).

In addition to being widely used (by many researchers and over several decades), work by Luce and Large

(2001; see also Cutler, Sebastian-Galles, Soler-Vilageliu, & van Ooijen, 2000) provides some evidence for the psychological validity of the one-phoneme metric as a measure of phonological similarity. In Experiment 2 of Luce and Large (2001), participants heard a nonsense word, such as /fin/, and were asked to produce the first real word that came to mind that sounded like the nonword stimulus item. Over 70% of the responses involved a change of one phoneme in the nonword to form a real word (e.g., _mean, fun, feet_). Eighteen percent of the responses involved a two-phoneme change (either CV, C_C, or VC), and the remaining 11% of the responses consisted of various types of changes involving the addition of a single segment (or the addition of a syllable). These results suggest that the operational definition of phonological similarity used in the present analysis also captures, to a large extent, the definition of phonological similarity that speakers may have.

By employing the same psychologically valid metric that has been used in many psycholinguistic experiments in the present graph theoretic analysis, psychologically valid insight about the mental lexicon might be obtained. Analyses of complex cognitive systems comprised of nodes that are connected in a manner that is not motivated by psycholinguistic research may be interesting mathematical exercises, but they are unlikely to provide significant insight into questions of interest to language researchers.

The same techniques used by physicists and computer scientists (e.g., Albert & Barabási, 2002) were used in the current graph theoretic analysis to examine the network structure of the phonological word-forms in the mental lexicon. Different mechanisms lead to different network structures, so identification of the structure of a given network can provide some insight into the mechanisms that might have influenced the development of the observed network. In conventional graph theoretic analyses, identification of the structure of a given network is accomplished by comparing several measurements from the graph of interest to the same measurements made in a random network with the same number of nodes and the same average number of connections per node as the graph of interest. Random graphs consist of a network in which links are randomly placed between nodes. They have been widely studied and are mathematically well understood (see Erdos & Rényi, 1960, for pioneering work on random graphs) and, therefore, provide a good baseline for comparisons. These conventional graph theoretic analyses were supplemented in some cases by additional comparisons to measurements obtained from a sample of random graphs generated by the Pajek program.

The following measurements were made in the present graph theoretic analysis of the phonological word-forms in the mental lexicon: average path length (ℓ), clustering coefficient (C), degree distribution, and the extent of assortative mixing by degree in the network. The *average*

path length of a network refers to the average distance between every node in the network and every other node in the network (Watts & Strogatz, 1998). The *clustering coefficient* characterizes the extent to which nodes connected to another node are also connected to each other. A clustering coefficient of 0 implies that none of the neighbors of a node are connected to the other neighbors of that node. A clustering coefficient of 1 implies that all of the neighbors of a node are connected to each other. Values between 0 and 1 imply that a number of neighbors of a node are also neighbors of each other (Watts & Strogatz, 1998).

The number of connections per node is also referred to as the *degree* of the node, or k. The *degree distribution* refers to the proportion of nodes [$P(k)$] that have a given number of links. In a degree distribution that resembles a normal bell-shaped distribution (i.e., a Poisson or Gaussian distribution), a small number of nodes will have fewer than the average number of connections per node, and a small number of nodes will have more than the average number of connections per node, but most nodes will have the average number of connections per node. This type of degree distribution is found in a random network. In a degree distribution that resembles a power-law, many nodes have a small degree (or a few connections), and a few nodes have a large degree (or many connections). This type of degree distribution has been found in graphs of many real-world systems, including graphs depicting connections among Web pages on the Internet (e.g., Albert & Barabási, 2002); the mechanisms that lead to the development of this special type of network structure will be described later. Typically a logarithmic transformation is applied to the degree distribution to reduce variability in the data and aid in the identification of the network structure.

Assortative mixing by degree refers to the probability of a highly connected node being connected to other nodes that are also highly connected (Newman, 2002; Newman & Park, 2003). In other words, there is a positive correlation between the degree of a node and the degree of its neighbors. In a network with disassortative mixing by degree, nodes that have many connections tend to be connected to nodes with few connections, producing a negative correlation between the degree of a node and the degree of its neighbors. In a random network, where connections are placed at random, the correlation between the degree of a node and the degree of its neighbors is zero (Newman, 2002). The implications of these different types of mixing by degree will be addressed later.

Results and Discussion

After connecting the nodes in the phonological network using the similarity metric employed in previous

psycholinguistic studies, I calculated the average path length and the clustering coefficient by using the Pajek program. Note that the average path length, or distance between any two nodes in a network, can be computed only on a fully connected network. In the case of the mental lexicon, there were many "lexical hermits" ($n = 10,265$), or words that had no phonological neighbors, such as *spinach* and *obtuse.* These hermits were not connected to the large group of words that were highly connected to each other, referred to as the largest component of the network. There were also a number of words in the lexicon ($n = 2,567$) that had a few neighbors, but neither the word nor the neighbors were similar to a word in the largest component of the network. These "lexical islands" contained words like *converse, convert,* and *converge* that were connected to each other but were not connected to any of the words in the largest component. The calculations of the network characteristics were, therefore, based on the 6,508 words in the largest component of the phonological lexicon.

Average Path Length (ℓ)

The average path length (ℓ) obtained from Pajek for the phonological network was 6.05. That is, on average, approximately six links had to be traversed to connect any two nodes in the (largest component of the) network. For example, to get from the word *cat* to the word *dog,* one can traverse the links between the nodes corresponding to the words *bat, bag,* and *bog.* The value obtained from the phonological network was compared with the average path length obtained from a comparably sized and connected random network, ℓ_{ran}. Because of the well-studied nature of random graphs, the convention in graph theoretic analyses (e.g., Albert & Barabási, 2002) is to estimate the value of ℓ_{ran} using Equation 1:

$$\ell_{ran} \approx \frac{\ln(n)}{\ln<k>}, \qquad (1)$$

where n refers to the number of nodes in the network and $<k>$ refers to the mean degree. In the phonological network, the largest component contained 6,508 nodes (n), and the mean degree ($<k>$) of the nodes in the largest connected component was 9.105; therefore, $\ell_{ran} = 3.975$.

The computationally derived value of ℓ_{ran} approximates the estimate of ℓ_{ran} obtained from 100 Erdos-Rényi random networks that were constructed in Pajek and contained the same number of nodes ($n = 6,508$) and the same average degree ($<k> = 9.105$) as the phonological network. The mean value of the average path length (ℓ_{100}) from the sample of 100 random networks was 4.22 ($SD = 0.01$; 95% confidence interval = 4.20–4.24).

Although the average path length for the phonological network ($\ell = 6.05$) was somewhat larger than the derived value of ℓ_{ran} and the estimated value of ℓ_{100}, the

conventions used in graph theoretic analyses would consider these values comparable (Watts & Strogatz, 1998). To further demonstrate the comparability of these values, compare them to the average path length of a similarly sized ordered network (ℓ_{ord}). In an ordered network, each node is linked to its nearest neighboring nodes. The average path length of an ordered network can be estimated by using Equation 2:

$$\ell_{ord} \approx \frac{n}{2 < k >},\qquad(2)$$

where, again, n refers to the number of nodes in the network, and $<k>$ refers to the mean degree. Given $n = 6{,}508$ nodes, and $<k> = 9.105$, $\ell_{ord} = 357.386$. The value of the average path length obtained from the phonological network (ℓ) was much closer to and of the same order of magnitude as the value obtained for ℓ_{ran} compared with the value obtained for ℓ_{ord}, further suggesting that the average path length of the phonological network is comparable to the average path length of a similarly sized random network.

Clustering Coefficient

As is the convention in graph theoretic analyses, the clustering coefficient (C) obtained from the (largest component of the) phonological network was compared with the value of the clustering coefficient that would be obtained from a comparably sized and connected random network (C_{ran}). The value of the clustering coefficient (C) obtained from Pajek for the largest component of the phonological network was .126. That is, neighbors of a given word have a tendency to also be neighbors of each other (see Vitevitch, 2006, for work on how the clustering coefficient may influence spoken word recognition).

Because of the well-studied nature of random graphs, the convention in graph theoretic analyses (e.g., Albert & Barabási, 2002) is to estimate the value of C_{ran} by using Equation 3:

$$C_{ran} \approx \frac{< k >}{n},\qquad(3)$$

where n refers to the number of nodes in the network, and $<k>$ refers to the mean degree. Given $n = 6{,}508$ nodes, and $<k> = 9.105$, the value of C_{ran} obtained from Equation 3 is .0014, which is about 90 times smaller than the value of C obtained in the phonological network. The value of C_{ran} was also calculated from the same 100 Erdos-Rényi random networks used in the analyses of the average path length. The mean value of C from the 100 random networks (C_{100}) was .00056 ($SD = .00006$; 95% confidence interval = .00044–.00067). The value of C_{100} is similar in magnitude to the derived value, C_{ran}, but both values are smaller in magnitude than the value of C obtained from

the phonological network. In other words, $C \gg C_{ran}$ (Watts & Strogatz, 1998).

Analyses of the average path length and the clustering coefficient are often used to determine whether a given network can be classified as a small-world network. As described in Watts and Strogatz (1998; Watts, 1999), a *small-world network* has (a) an average path length that is comparable to the average path length of a random network, but (b) a clustering coefficient that is much greater than the clustering coefficient of a random network with the same number of nodes and the same average degree. Small-world networks are so called because the pattern of connections that yields Characteristics A and B makes the network easy to traverse.[1] That is, the very large system has the appearance of being relatively small. The results of the present analyses suggest that the phonological network has the characteristics of a small-world network, a structure that is shared with the semantic networks that have been previously investigated (e.g., Albert & Barabási, 2002; Batagelj et al. 2002; Ferrer i Cancho & Solé, 2001b) and with many other real-world systems. To further examine the structure of the phonological network, I performed additional analyses, including an examination of the degree distribution and the extent to which the phonological network exhibits assortative mixing by degree.

Degree Distribution

The degree distribution of a graph provides additional information about the structure of that system. To maintain consistency with traditional graph theoretic analyses, as well as with the present analyses of the average path length and the clustering coefficient, the degree distribution of the largest component of the phonological lexicon—instead of the entire lexicon—was examined.

One type of network structure that has received much attention is a *scale-free network* (Barabási & Albert, 1999). In a scale-free network, many nodes in the network have few links, but there are a few nodes with many links. That is, a scale-free network can be identified by

---

[1]The "small-world" concept may be familiar to many because of the classic work on the social "distance" between any two people in the United States (Milgram, 1967). In this study, a randomly chosen person received the name and address of a target individual and a set of instructions directing them to deliver a packet to the target individual. If the randomly chosen person knew the target individual, they could send the packet directly to the target. If the randomly chosen person did not know the target individual, they were to send the packet to someone they knew on a first name basis who was more likely to know the target individual. The process of sending the packet to someone that was more likely to know the target was repeated until the target individual finally received the packet. Milgram found that approximately six intermediate acquaintances were required to get the packet from the randomly chosen person to the target individual. This work contributed to the notions that we live in a "small world" and that there are only "six degrees of separation" between any two people on the planet (see also the work of Granovetter, 1973).

the presence of a power law function in the degree distribution. A power-law relationship appears as a straight line in a log-log plot of the degree distribution (Albert & Barabási, 2002). The degree exponent, γ, or slope of the line in the log-log plot of the degree distribution, in most scale-free networks is $2 < \gamma < 3$; however, Montoya and Solé (2002) found several examples of scale-free networks that had values for γ much less than 2 (one as low as 1.05).

An alternative method to determine whether a network is a scale-free network is to plot the *cumulative* degree distribution in a log-log scale (Newman, 2006) or the *cumulative* degree distribution in a linear-log scale (Amaral, Scala, Barthélémy, & Stanley, 2000). In a log-log scale, the cumulative degree distribution would again show a straight line if a power-law function existed, but this time with a slope of $\gamma - 1 = 1.5$ (Newman, 2006). In a linear-log plot, a power-law distribution would be evident if the line fitting the data had an upward bend to it, whereas a straight line would indicate an exponential relationship in the degree distribution instead of a power-law relationship (Amaral et al., 2000). Both the degree distribution and the cumulative degree distribution will be used to examine the structure of the phonological network.

Figure 1 shows the degree distribution for the 6,508 connected word-forms in the largest component of the phonological lexicon on a log-log scale. The best fitting power-law and exponential lines were computed over all of the data points and are also plotted in the figure. The root-mean-square error (RMSE) measure of fit shows that the observed data fit an exponential distribution ($Y = .141 * e^{-.123*X}$, RMSE = .005) better than a power-law

distribution ($Y = .672 * X^{-1.489}$, RMSE = .089). Given the poor fit of the power-law function, the degree exponent, γ, was not calculated.

Figure 2 shows the cumulative degree distribution of the phonological network in a log-log scale (Newman, 2006), with the best fitting power-law ($Y = 21.524 * X^{-2.13}$) and exponential ($Y = 2.807 * e^{-.181*X}$) lines also plotted in the figure. Figure 3 shows the cumulative degree distribution of the phonological network in a linear-log scale (Amaral et al., 2000), with the best fitting power-law and exponential lines also plotted in the figure. The RMSE measure of fit shows that the observed data again fit an exponential distribution (RMSE = .335) better than a power-law distribution (RMSE = 3.198). Given the poor fit of the power-law function, the degree exponent, γ, was not calculated. Both methods of evaluation—degree distribution and cumulative degree distribution—show that the degree distribution of the phonological network is better characterized as an exponential function than as a power-law function.

Assortative Mixing

A final analysis was performed to determine whether the phonological network showed assortative mixing or disassortative mixing by degree (Newman, 2002; Newman & Park, 2003). In a network with assortative mixing by degree, the nodes that have many connections tend to be connected to other nodes that have many connections. In other words, there is a positive correlation between the degree of a node and the degree of its neighbors. In a network with disassortative mixing by degree,

Figure 1. The degree distribution for the 6,508 word-forms in the largest fully connected component of the phonological network in a log-log plot (the solid squares). The best fitting power-law (straight line) and exponential functions (curved line) are also displayed for comparison.
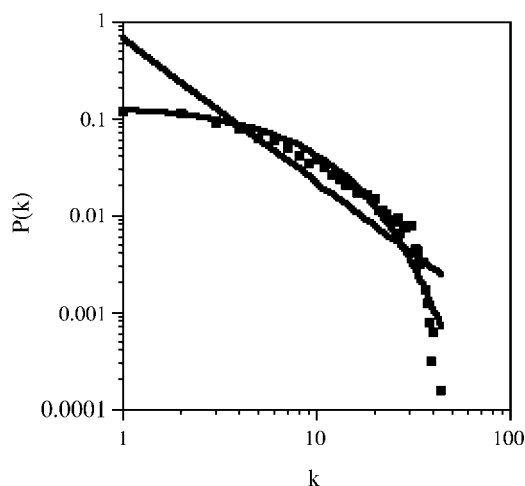
Figure 2. The cumulative degree distribution for the 6,508 word-forms in the largest fully connected component of the phonological network in a log-log plot (the solid diamonds). The best fitting power-law (straight line) and exponential (curved line) functions are also displayed for comparison.
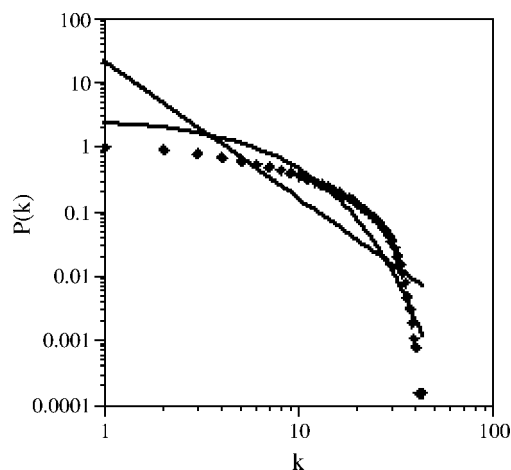
Figure 3. The cumulative degree distribution for the 6,508 word-forms in the largest fully connected component of the phonological network in a linear-log plot (the solid diamonds). The best fitting power-law (upward curving line) and exponential (straight line) functions are also displayed for comparison.
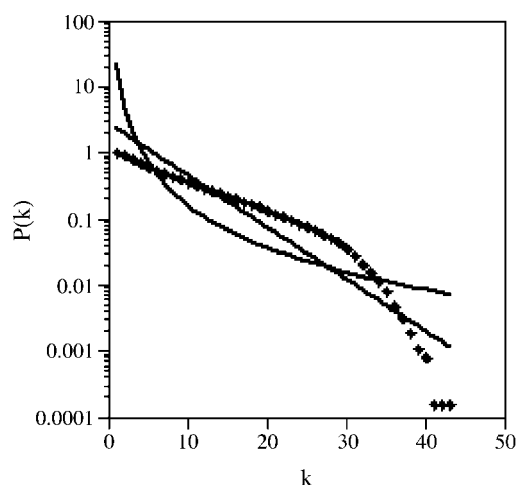
nodes that have many connections tend to be connected to other nodes with few connections—a negative correlation between the degree of a node and the degree of its neighbors. Examination of the type of mixing by degree that occurs in a network is of interest because Newman (2002) discussed different processing implications for networks with assortative and disassortative mixing by degree (a point that will be discussed later in the context of the phonological network).

The number of connections, or degree, of each of the 6,508 nodes in the fully connected component was correlated with the degree of each of its neighbors. The resulting Pearson's r (29,613) = .62, $p < .0001$, shows a positive correlation. That is, nodes with many connections tend to be connected to nodes that also have many connections, whereas nodes with few connections tend to be connected to nodes that also have few connections. Said another way, words with dense phonological neighborhoods tend to have neighbors that also have dense phonological neighborhoods, whereas words with sparse phonological neighborhoods tend to have neighbors that also have sparse phonological neighborhoods; therefore, the phonological network exhibits assortative mixing by degree.

What Advantages Does This Structure Afford the Mental Lexicon?

The analyses of the average path length and the clustering coefficient suggested that the phonological network had the characteristics of a small-world network (Watts & Strogatz, 1998). Although such networks may be very large and sparsely connected, processing on networks

with a small-world structure occurs rapidly, accurately, and robustly. Given the small-world characteristics observed in the network of phonological word-forms, it is perhaps not a coincidence that language processes are also rapid, accurate, and robust.

Additional analyses showed that the network of phonological word-forms exhibited assortative mixing by degree. Newman (2002) examined several real-world networks and discussed the implications that assortative and disassortative mixing by degree might have on "processing" in those systems. The assortative mixing observed in the structure of the phonological network may have similar implications for the process of retrieving word-forms from the mental lexicon. The following discussion, however, should not be viewed as a proposal for a model of lexical processing. Rather, it is simply an exploration of how the structure of the phonological network might affect its function (Strogatz, 2001; Ward, 2002).

To examine how the structure of the phonological word-forms might influence lexical processing, consider how a common metaphor for lexical processing—activation from acoustic–phonetic input spreading to phonologically related words—might be affected by a network exhibiting disassortative versus assortative mixing by degree. In a network with disassortative mixing by degree, the distribution of highly connected nodes throughout the network would result in activation spreading among many, if not all, of the word-forms in the mental lexicon. A large number of potential lexical candidates would then have to be rejected during each attempt at lexical retrieval, perhaps making the process of word recognition slow, resource intensive, and—especially in listening conditions that are less than ideal—very laborious.

However, in a network with assortative mixing by degree, highly connected nodes tend to be clustered together, thereby restricting the spread of activation to a more circumscribed region of the network (forming what Newman, 2002, referred to as a "reservoir" in the context of disease transmission). In this case, activation from acoustic–phonetic input would spread among a smaller set of potential candidates, even in the case of words that have many phonological neighbors (see Norris, 1994, for a *processing model* instead of a structural model that leads to a "short list" of lexical candidates). The spread of activation to a more restricted set of candidates means that not every word in the lexicon would have to be considered and rejected as a potential lexical candidate as it would in a network with disassortative mixing, ensuring rapid, seemingly automatic retrieval of the correct lexical candidate from the lexicon, even in less than ideal listening conditions. The assortative mixing found in the phonological network also suggests that models of spoken word retrieval may not have to postulate a special

ad hoc mechanism that prevents the entire lexicon from being activated whenever any acoustic-phonetic input is received, as the very structure of the phonological network may prevent this from happening.

Suppose now that the correct word-form in the lexicon does not receive enough activation to cross the threshold necessary for successful lexical retrieval. That is, the node is (temporarily) removed from the network. What are the implications for processing when a node is removed from a system that exhibits disassortative or assortative mixing by degree? In the case of the mental lexicon, how might the structure of the system influence recovery from a failed attempt at retrieving a phonological word-form? In a network with disassortative mixing by degree, the removal of one of the highly connected nodes that are distributed throughout the network is likely to result in the fracturing of the fully connected network into smaller components. If the mental lexicon were structured in this way, the reduced connectivity in the system caused by the removal of that node may make it difficult for activation to spread to another lexical candidate, resulting in catastrophic failure of the lexical retrieval process. In other words, if the correct phonological word-form is not retrieved, then nothing is retrieved.

In a network with assortative mixing by degree, the removal of one of the highly connected nodes from the cluster of highly connected nodes does little to break up the network, as the numerous "redundant" pathways found in the cluster of highly connected nodes maintain the connectivity of the network. The alternative pathways enable activation to spread to another (albeit, incorrect) candidate. Thus, if the correct word-form is not retrieved in a network with assortative mixing, catastrophic failure does not occur. Rather, the system experiences graceful degradation (McClelland, Rumelhart, & Hinton, 1986) and finds another item that best matches the input. Evidence from studies of speech perception errors (Vitevitch, 2002b) and speech production errors (e.g., Vitevitch, 1997) suggests that catastrophic failure is not likely to occur when the lexical retrieval process fails. Instead, the system experiences graceful degradation, retrieving at least partial information about a word (e.g., interlopers, or the first letter or syllable of a word that is on the "tip of the tongue"; Vitevitch & Sommers, 2003). The graceful degradation that occurs during lexical retrieval failures may in part be due to the assortative mixing by degree found in the structure of the phonological network.

Furthermore, Newman (2002) found that 5 to 10 times more nodes that were highly connected had to be removed from a network with assortative mixing than from a network with disassortative mixing to destroy the connectivity of the network. This finding suggests that the assortative mixing found in the phonological network may also contribute to the robustness of language

processing in the face of more permanent damage to the system stemming from stroke or other injury to language-related areas of the brain. Only in the case of massive damage to the system, when a large number of highly connected nodes in the system are removed, is the connectivity of the network likely to be compromised, perhaps leading to catastrophic processing failures. In the case of lexical processing, conditions of catastrophic failure, such as pure word deafness, may be the result of massive damage to the lexicon that removes a large number of highly connected nodes from the system.

Although the present graph theoretic analyses simply examined the structure that was observed among phonological word-forms in the mental lexicon, the nature of that organization has important implications for the type of processing carried out in that system (Strogatz, 2001; Ward, 2002). The present discussion highlighted how the small-world nature and the assortative mixing by degree found in the phonological network might, in part, account for the speed, accuracy, and robustness of lexical processing. The following discussion considers the mechanisms that might lead to certain network structures and explores the implications of those mechanisms for the acquisition of phonological word-forms.

What Leads to This Structure in the Mental Lexicon?

Certain mechanisms lead to the formation of certain types of network structures. Therefore, the analyses of the degree distribution of a network can be used to help identify the structure of a given network and to shed some light on the mechanisms that may have led to its formation. The mechanisms that lead to the formation of certain network structures might be viewed as mechanisms that operate on a longer time scale, akin to evolution, or as mechanisms that operate on shorter time scales, akin to development or learning. Indeed, some researchers have argued that evolution, development, and learning might differ in the time scales in which they operate, but they are all subject to a similar underlying mechanism (Dickins & Levy, 2001). Given the possibility that a similar mechanism might govern evolution, development, and learning, the following discussion will examine the mechanisms that might lead to the formation of certain network structures and will explore the implications of those mechanisms for word learning.

Much attention—even in the popular press (Barabási, 2002)—has focused on the apparent ubiquity of networks with a scale-free structure. The interest in scale-free networks might, in part, be related to the power-law relationship found in the degree distribution, as power-law distributions have also been associated with the phenomenon of self-organized criticality (Bak, Tang, & Wiesenfeld,

1988). Barabási and Albert (1999; Barabási, Albert, & Jeong, 1999) suggested that two mechanisms lead to the emergence of scale-free networks: *growth* and *preferential attachment*. Growth refers to the addition of new nodes to the network over time. Preferential attachment is a constraint that makes it more likely for new nodes being added to the system to connect to nodes that are already highly connected. Barabási and Albert (1999; Barabási et al., 1999) found that both growth and preferential attachment were necessary to create the power-law degree distribution in scale-free networks (however, see Ferrer i Cancho & Solé, 2001a, for evidence that a process that optimizes the distance between nodes and the number of links per node may also lead to a power-law degree distribution).

In the case of phonological word forms in the mental lexicon, there is some evidence for growth and preferential attachment. First, consider that the mental lexicon grows over time. Although learning new word-forms is something that is typically associated with and primarily studied in children (e.g., Storkel, 2001, 2003), it is not controversial to state that adults also learn new words. In other words, the mental lexicon continues to grow over time.

Now consider preferential attachment, or the tendency for a new node to attach to a node that is connected to many rather than to few nodes in the system. In the case of the mental lexicon, a novel word-form that is phonologically similar to many known words (i.e., with a dense phonological neighborhood) should be acquired more easily than a novel word-form that is phonologically similar to few known words (i.e., with a sparse phonological neighborhood). Indeed, Storkel (2001, 2003; see also Beckman & Edwards, 2000; Gathercole, Hitch, Service, & Martin, 1997; Storkel & Morrisette, 2002) found that preschool-age children learned novel words that had common sound sequences (which are correlated with word-forms with dense phonological neighborhoods; Vitevitch, Luce, Pisoni, & Auer, 1999) more rapidly than novel words that had rare sound sequences (i.e., sparse neighborhoods). Similarly, Storkel, Armbrüster, and Hogan (2006) found that college-age adults learned novel words with dense neighborhoods more rapidly than novel words that had sparse neighborhoods, further suggesting that a mechanism like preferential attachment may influence the growth of the mental lexicon.

Furthermore, work by Page (2000; see also Grossberg, 1972) illustrates how a localist neural network with a competitive learning algorithm (i.e., not the type of network examined in the present investigation) might exhibit behavior that resembles growth via preferential attachment. In a localist neural network, an individual node represents a given concept or, in the case of the phonological lexicon, a word-form. When a novel word-form is presented to the localist network, several uncommitted nodes become partially activated by the input and compete with each other to become the node that will be committed to representing that input pattern in the future. Each node will adapt its weights to better match the input pattern. Eventually, one node will match the input pattern better than the other competing nodes and will become committed to representing that word. The "losing" nodes remain uncommitted (i.e., they do not represent a known word), but because of the previous competition, their weights are in an excellent position to represent a new input pattern that is similar to the previously learned input pattern. Such a mechanism not only accounts for the advantage found in word learning for similar sounding words (i.e., words with dense neighborhoods) over unique words (i.e., words with sparse neighborhoods) but also illustrates how a connectionist model could exhibit behavior that resembles preferential attachment.

Given the evidence for growth and preferential attachment in the mental lexicon, it is somewhat surprising that a power-law relationship was not observed in the analysis of the degree distribution of phonologically similar words in the mental lexicon (cf. Albert & Barabási, 2002; Batagelj et al., 2002; Ferrer i Cancho & Solé, 2001b; Motter et al., 2002; Steyvers & Tenenbaum, 2005). Instead, an exponential function provided a better fit to the degree distribution of the phonological network.

Also note that Newman (2002) found that the Barabási and Albert (1999) scale-free network had neither assortative nor disassortative mixing by degree; rather, the correlation of the degrees of connected nodes, perhaps counterintuitively, was 0. Recall that assortative mixing by degree was observed in the phonological network. The exponential degree distribution and the presence of assortative mixing by degree suggest that it is unlikely that the phonological network is a scale-free network like those examined by Barabási and Albert (1999). What mechanisms might lead to the network characteristics observed in the phonological network: a degree distribution that does not follow a power-law and assortative mixing by degree?

In addition to showing that growth and preferential attachment lead to a scale-free network with a power-law degree distribution, Barabási and Albert (1999) also demonstrated that a system that does not grow or that grows without preferential attachment is not likely to exhibit a power-law degree distribution. Recall, however, that the mental lexicon does grow over time and does seem to be influenced by a mechanism that resembles preferential attachment. Perhaps additional constraints on the formation of the network—and by implication, on the acquisition of novel words—led to the characteristics observed in the phonological network.

Amaral et al. (2000) found that if there is a cost associated with the attachment of a new node (i.e., the node may be able to accommodate only a fixed number of links), then a power-law degree distribution, like that in the scale-free model proposed by Barabási and Albert (1999), is not likely to be observed. In the case of phonological word-forms, restrictions on word length, on the sounds used in that language (i.e., phonemic inventory), and on the sequencing of those sounds in a word (i.e., phonotactic constraints) may limit the number of new nodes that can link to an already existing node. The costs associated with adding a new link in the phonological lexicon may, in part, account for the failure to find a degree distribution that follows a power law (see also Krapivsky, Redner, & Leyvraz, 2000, for the influence of nonlinear preferential attachment on the degree distribution). Although additional constraints on preferential attachment may produce degree distributions that do not follow a power law (e.g., an exponential degree distribution), it is not clear whether they will also lead to assortative mixing by degree or to the presence of lexical hermits—nodes that were not connected to any other node in the system—as was observed in the phonological network.

One type of growing network that does exhibit all of these characteristics—exponential degree distribution, assortative mixing by degree, and the presence of isolated nodes and islands—is the randomly grown network examined by Callaway, Hopcroft, Kleinberg, Newman, and Strogatz (2001). These characteristics emerged in the growing network examined by Callaway et al. because of two stochastic processes: (a) new connections are placed between randomly chosen pairs of nodes, and (b) new nodes that are added to the system are not required to attach to a preexisting node. Do such processes also influence the acquisition of phonological word-forms?

Consider the first stochastic process described by Callaway et al. (2001), in which new connections are placed between randomly chosen pairs of nodes in the network. In the phonological network, there is some probability that a new node might link to an already existing node in the network. In terms of the mental lexicon, a new word might be phonologically similar to an already known word. Callaway et al. also noted that older nodes, or those items that had been in the network for some time, tended to have a higher degree, or more connections than more recently added nodes. That is, it is also likely that the already existing node will have a high degree, or in terms of the mental lexicon, it is likely that the already known word will have a dense phonological neighborhood. Consistent with this prediction, Storkel (2004) found a positive correlation between age of acquisition and neighborhood density, such that words learned early in life tended to sound similar to many words (i.e., they had a higher degree) than a word learned later in life. The resemblance

between the network model examined by Callaway et al. and the mental lexicon regarding the relationship between age of acquisition and neighborhood density is quite striking.

It is important to note, however, that the relationship observed by Storkel (2004) is correlational in nature and, therefore, could be the result of several possibilities. For example—as implied by the first stochastic process in the network model examined by Callaway et al. (2001)—the first words that are learned may lay a lexical foundation that makes it more likely that some words (i.e., those that are similar to known words) but not others (i.e., those that are not similar to known words) will be acquired in the future. Another possibility—as the results of laboratory-based word-learning experiments by Storkel and others might suggest (Storkel 2001, 2003; Storkel et al. 2006; see also Beckman & Edwards, 2000; Gathercole et al., 1997)—is that words with dense neighborhoods are easier to learn and are therefore acquired earlier in life than words with sparse neighborhoods. Alternatively, a third variable may be responsible for the apparent relationship between age of acquisition and neighborhood density. In the case of the mental lexicon, the frequency with which a word occurs in the language is correlated with both age of acquisition and neighborhood density. Finally, a combination of some or all of these possibilities might be at work. The exact nature of the relationship between age of acquisition and neighborhood density is an interesting question that cannot be definitively answered by the present analyses and must, therefore, be left for future research to address.

Consider further the first stochastic process described by Callaway et al. (2001): New connections are placed between randomly chosen pairs of nodes in the network. This stochastic process also implies that there is some probability that a link may be placed between two preexisting nodes in the network. That is, two known words that were not previously identified as being phonologically similar may, at a subsequent point in time, become phonologically similar. How can two words that were not previously phonologically similar become phonologically similar at a later point in time? Perhaps the lexical restructuring hypothesis proposed by Metsala and Walley (1998) might account for the later emergence of links between nodes in the lexicon. Metsala and Walley suggested that continued vocabulary growth leads to the internal restructuring of lexical representations. That is, lexical representations gradually become more detailed throughout early and middle childhood. Such a process may lead to changes in phonological similarity between previously unrelated items and may account for the subsequent placement of a link between two preexisting nodes.

To illustrate how the lexical restructuring hypothesis might account for the subsequent placement of a

link between two preexisting nodes, imagine that the lexicon of a child at an early point in time consisted only of the words *cat* and *dog*. With such a small vocabulary, very abstract representations of those word-forms that lack much detail might be sufficient to distinguish between those two word-forms. Abstract representations for *cat* and *dog* might consist of the first segment in each word and some general information indicating that (perhaps, a certain number of) additional speech sounds follow; suppose something like /k–/ and /d–/.

If a new node with the abstract representation /d–/, corresponding to the recently learned word *doll,* was added to the network, one would no longer be able to distinguish between the word-forms for *dog* and *doll*. More detail would need to be added to the phonological representations for *dog* and *doll* (perhaps /d-[stop]/ and /d-[glide]/) in order to distinguish between those similar words. The subsequent addition of this more detailed information might result in a link being placed between these two preexisting yet similar nodes at some later point in time.

Although there is some evidence to support the lexical restructuring hypothesis (e.g., Edwards, Beckman, & Munson, 2004; Munson, Swenson, & Manthei, 2005; Storkel, 2002; Walley, 1993), it has been articulated only in very general terms and is not without its critics (e.g., Swingley, 2003; Swingley & Aslin, 2002). Admittedly, there are still some open questions regarding the lexical restructuring hypothesis, including whether the addition of more detailed information to lexical representations may result in two words that were previously considered similar to each other to no longer be phonological neighbors. This possibility might correspond to the removal of a link between two nodes in a network. However, such a mechanism is not present in the model proposed by Callaway et al. (2001). Despite the lack of specific details in the lexical restructuring hypothesis, such a mechanism makes it at least plausible that the simple stochastic process proposed by Callaway et al. not only contributes to the structure observed in the randomly grown network that they examined, but may also contribute to the structure observed in the phonological network.

Consider now the second stochastic process described by Callaway et al. (2001): New nodes added to the system are not required to attach to a preexisting node in the network. Accidental gaps and various restrictions (e.g., on word length, on the phonemic inventory, or on phonotactic sequencing) have resulted in numerous word-forms that are not phonologically related to any other extant word. The process of adding new nodes to the system without requiring them to attach to a preexisting node in the network can account for the presence of these lexical hermits in the phonological network. This would not be true in the scale-free network model proposed by Barabási and Albert (1999), in which a new node has to

also connect to a preexisting node. In that model, there is no way to account for the presence of lexical hermits as observed in the phonological network. Thus, this simple stochastic process not only contributes to the structure observed in the randomly grown network examined by Callaway et al., but may also contribute to the structure observed in the phonological network.

It is quite striking that these two simple stochastic processes can account for several nontrivial results related to the acquisition and subsequent organization of phonological word-forms in the mental lexicon. The present analyses certainly do not prove that (only) these two stochastic processes caused the structure observed in the phonological network, nor do they account for all of the patterns that were observed, but they do suggest that a long list of complicated and detailed constraints that capture the microscopic details of language may not be necessary to produce the structure observed in the phonological network. Future simulations and graph theoretic analyses will further investigate which mechanisms led to the structure that was observed in the phonological network examined here.

General Discussion

Although graph theoretic concepts have been used for some time to study social interactions and social networks (e.g., Kochen, 1989; Wasserman & Faust, 1994; Wellman & Wortley, 1990), their use has been conspicuously absent from studies of cognitive processing (see Sporns, Chialvo, Kaiser, & Hilgetag, 2004, for the application of graph theory to neuroscience). The present work (see also Steyvers & Tenenbaum, 2005) suggests that these techniques can also be used to increase our understanding of complex cognitive systems, such as the mental lexicon. Speech-language pathology, psycholinguistics, and cognitive science more generally might obtain great insights by viewing complex cognitive systems from this alternative perspective.

Alternative perspectives, such as connectionist modeling, have, in the past, reshaped and advanced our understanding of various psychological processes—including the influence of evolution, development, and learning on those processes—in significant ways (e.g., Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996). Looking at the mental lexicon in graph theoretic terms may have similar consequences, as it places the lexicon, and other complex psychological systems, in a broader context and allows us to see that cognitive systems may be governed by the same underlying principles—such as those that led to small-world structures—that govern other complex systems found in the world. If complex cognitive systems are subject to

the same constraints as other real-world systems, then ad hoc specific mechanisms may not be needed to account for processes like word learning and word retrieval (e.g., Markson & Bloom, 1997).

Callaway et al. (2001) stated, "We do not claim that our model is an accurate reflection of any particular real-world system, but we find that studying a model that exhibits network growth in the absence of other complicating features leads to several useful insights" (p. 1). The present analyses employed simple models, like those explored by Callaway et al. These networks with uncomplicated features indeed provided several useful insights regarding lexical access and acquisition. Specifically, a few simple stochastic processes may lead to a structure that significantly influences the learning and retrieval of word-forms in the mental lexicon.

Although the quantitative fit between the phonological network and the graph theoretic models was not perfect (e.g., the faster-than-exponential decay in Figure 3), the qualitative similarity of the phonological network, a real-world cognitive system, to the model examined by Callaway et al. (2001) may offer a unique opportunity for each field to learn from and contribute to the development of the other. Cross-disciplinary analyses of the mental lexicon might lead to the discovery of various parameters that influence the development of network structures in many real-world systems. The study of the structure observed in the mental lexicon might also lead to the development of new techniques to better classify different types of networks. Indeed, several experts in network analysis recently discussed 10 topics that future research should focus on ("Virtual Round Table," 2004). First among those research topics was the following question: Are there formal ways of classifying the structure of different growing models? The present analyses have revealed a number of puzzles that might stimulate such cross-disciplinary investigation.

One puzzle that might capture the attention of researchers relates to the amount of assortative mixing by degree that was observed in the phonological network. Recall that the analysis of assortative mixing by degree in the phonological network found a correlation of .62. The value of the correlation coefficient obtained in the present analyses, however, is much greater than that reported by Newman (2002) for other real-world networks (a maximum of .363 for a coauthorship network of physicists) and is greater than the value predicted by the randomly grown network examined by Callaway et al. (2001; less than .4). It is not clear that any of the network models that were examined in the present analysis can account in any way for this observation.

What could lead to a greater amount of assortative mixing by degree? Perhaps somewhat counterintuitively, such a situation might emerge in the phonological network because of numerous restrictions or constraints on word formation. With some constraints on what constitutes a legal word in a given language, one might imagine that maximally dissimilar word-forms would populate the lexicon to limit confusability among the items and to facilitate the transmission of information. With many more constraints, however, the lexicon that emerges might instead resemble a group of individuals wearing a uniform of some sort. That is, the options for what shirt and what pants to wear (or what segments can co-occur) are so limited that everyone ends up looking very similar, like students in a Catholic school, employees in a fast food restaurant, soldiers in the military, or words that are morphologically related to each other.

In addition to leading to greater assortative mixing by degree, additional constraints on word formation, such as those imposed by morphology, might also underlie another anomaly that was observed in the phonological network, namely, the existence of a rather large island of related word-forms. Recall that the phonological network consisted of many words in a large interconnected component, many lexical hermits, and a number of islands containing several words that were related to each other but not to anything else. Although the processes proposed by Callaway et al. (2001) predict the existence of hermits and of islands of various sizes, additional analyses (that were not reported here) suggest that it is not very likely that an island containing over 50 nodes, as was observed in the phonological network, should exist in a network of this size and connectivity. (Because approximately two thirds of the words in this group contain the morphologically relevant sequence of segments /shɪn/, such as *faction, fiction,* and *fission,* I facetiously refer to this group of words as the island of the "shunned.") Further research is required to determine how to best represent additional constraints on word formation, like those imposed by morphology, in a network model, and to determine whether analogous constraints also influence other real-world systems.

Although there is much to be gained from such opportunities for interdisciplinary research, future research focusing on the mental lexicon may also benefit from the application of graph theoretical techniques. Advances in our understanding of language processing and language-related disorders might be made by examining the lexicon of individual children, instead of an "average" lexicon as was done in the present analyses, to determine whether language acquisition is proceeding along a typical trajectory. One might also employ graph theoretic approaches to help identify certain words in the vocabulary of an individual child that might facilitate the diffusion of sound change throughout the lexicon (e.g., Gierut, 2001).

Future graph theoretic analyses of the lexicon could also use more complex graphs to examine other structural

characteristics that might influence language processing. For example, a network with weighted links might be used to examine how overlapping words like *cap* and *captain* might influence lexical retrieval (however, see Newman, Sawusch, & Luce, 2005). Alternatively, a graph that allows multiplexity (two or more links of different types might exist between nodes; see Koehly & Pattison, 2005) could be used to examine how semantic information might interact with phonological information during lexical processing. The present analysis shows some of the potential that graph theory and the new science of networks (Watts, 2004) hold for understanding cognitive processing. In short, graph theory offers a new and useful set of mathematically rigorous tools to increase our understanding of language-related processes and other complex cognitive systems.

Acknowledgments

References

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

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences, 97,* 11149–11152.

Arnold, H. S., Conture, E. G., & Ohde, R. N. (2005). Phonological neighborhood density in the picture naming of young children who stutter: Preliminary study. *Journal of Fluency Disorders, 30,* 125–148.

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A, 38,* 364–374.

Barabási, A. L. (2002). *Linked: The new science of networks.* Cambridge, MA: Perseus.

Barabási, A. L., & Albert, R. (1999, October 15). Emergence of scaling in random networks. *Science, 286,* 509–512.

Barabási, A. L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A, 272,* 173–187.

Batagelj, V., & Mrvar, A. (1998). Pajek: A program for large network analysis. *Connections, 21,* 47–57.

Batagelj, V., Mrvar, A., & Zaveršnik, M. (2002). Network analysis of texts. In E. Tomaž & J. Gros (Eds.), *Proceedings of the 5th International Multi-Conference Information Society—Language Technologies* (pp. 143–148). Ljubljana, Slovenia: Multi-Conference Information Society.

Beckman, M. E., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development, 71,* 240–249.

Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E. J., & Strogatz, S. H. (2001). Are randomly grown graphs really random? *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 64*(4, Pt. 1).

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicons. *Journal of Child Language, 17,* 205–215.

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighborhoods in young children's receptive vocabularies. *Journal of Child Language, 22,* 727–735.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighborhoods in the developing lexicon. *Journal of Child Language, 30,* 441–469.

Cutler, A., Sebastian-Galles, N., Soler-Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28,* 746–755.

Dickins, T. E., & Levy, J. P. (2001). Evolution, development and learning: A nested hierarchy? In R. M. French & J. P. Sougné (Eds.), *Connectionist models of learning, development and evolution: Proceedings of the Sixth Neural Computation and Psychology Workshop* (pp. 263–270). London: Springer-Verlag.

Dollaghan, C. A. (1994). Children's phonological neighborhoods: Half empty or half full? *Journal of Child Language, 21,* 257–271.

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47,* 421–436.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Erdos, P., & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 5,* 17–61.

Ferrer i Cancho, R., & Solé, R. V. (2001a, November). *Optimization in complex networks* (Working Paper No. 01-11-068). Santa Fe, NM: Santa Fe Institute.

Ferrer i Cancho, R., & Solé, R. V. (2001b). The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences, 268,* 2261–2266.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33,* 966–979.

Gierut, J. A. (2001). A model of lexical diffusion in phonological acquisition. *Clinical Linguistics & Phonetics, 15,* 19–22.

Gierut, J. A., Morrisette, M. L., & Champion, A. H. (1999). Lexical constraints in phonological acquisition. *Journal of Child Language, 26,* 261–294.

Gordon, J. K., & Dell, G. S. (2001). Phonological neighborhood effects: Evidence from aphasia and connectionist modeling. *Brain and Language, 79,* 21–23.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11,* 341–363.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78,* 1360–1380.

Greenberg, J. H., & Jenkins, J. J. (1967). Studies in the psychological correlates of the sound system of American English. In L. A. Jacobivits & M. S. Miron (Eds.), *Readings in the psychology of language* (pp. 186–200). Englewood Cliffs, NJ: Prentice-Hall.

Grossberg, S. (1972). Neural expectation: Cerebellar and retinal analogy of cells fired by learnable or unlearnable pattern classes. *Kybernetik, 10,* 49–57.

Harary, F., Norman, R. Z., & Cartwright, D. (1965). *Structural models: An introduction to the theory of directed graphs*. New York: Wiley.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America, 117,* 896–907.

Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of Speech, Language, and Hearing Research, 46,* 390–404.

Kochen, M. (1989). *The small world: A volume of recent research advances commemorating Ithiel de Sola Pool, Stanley Milgram, Theodore Newcomb*. Norwood, NJ: Ablex Publishing.

Koehly, L. M., & Pattison, P. (2005). Random graph models for social networks: Multiple relations or multiple raters. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 162–191). New York: Cambridge University Press.

Krapivsky, P. L., Redner, S., & Leyvraz, F. (2000). Connectivity of growing random networks. *Physical Review Letters, 85,* 4629–4632.

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

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

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

Markson, L., & Bloom, P. (1997, February 27). Evidence against a dedicated system for word learning in children. *Nature, 385,* 813–815.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 1. Foundations* (pp. 3–44). Cambridge, MA: MIT Press.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning literacy* (Vol. 4, pp. 89–120). Mahwah, NJ: Erlbaum.

Milgram, S. (1967). The small-world problem. *Psychology Today, 2,* 60–67.

Montoya, J. M., & Solé, R. V. (2002). Small world patterns in food webs. *Journal of Theoretical Biology, 214,* 405–412.

Motter, A. E., de Moura, A. P. S., Lai, Y. C., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 65*(6, Pt. 2), 065102. Retrieved February 18, 2003, from http://prola.aps.org/pdf/PRE/v65/i6/e065102.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47,* 1048–1058.

Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children: Evidence from repetition tasks. *Journal of Speech, Language, and Hearing Research, 48,* 108–124.

Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters, 89,* 208701. Retrieved December 18, 2006, from http://arxiv.org/abs/cond-mat/0205405.

Newman, M. E. J. (2006). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46,* 323–351.

Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 68*(3, Pt. 2), 036122. Retrieved December 18, 2006, from http://prola.aps.org/pdf/PRE/v68/i3/e036122.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (2005). Do postonset segments define a lexical neighborhood? *Memory & Cognition, 33,* 941–960.

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

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In *Research on Speech Perception Progress Report No. 10* (pp. 357–376). Bloomington: Indiana University.

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23,* 443–512.

Roberts, F. S. (1976). *Discrete mathematical models*. Englewood Cliffs, NJ: Prentice-Hall.

Saussure, F. de. (1966). *Course in general linguistics* (Wade Baskin, Trans.). New York: McGraw-Hill. (Original work published 1916)

Sommers, M. S. (1996). The structural organization of the mental lexicon and its contribution to age-related declines in spoken word recognition. *Psychology and Aging, 11,* 333–341.

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences, 8,* 418–425.

Steyvers, M., & Tenenbaum, J. (2005). The large scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41–78.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44,* 1321–1337.

Storkel, H. L. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language, 29,* 251–274.

Storkel, H. L. (2003). Learning new words: II. Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research, 46,* 1312–1323.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics, 25,* 201–221.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49,* 1175–1192.

Storkel, H. L., & Morrisette, M. L. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech, and Hearing Services in Schools, 33,* 24–37.

Strogatz, S. H. (2001, March 8). Exploring complex networks. *Nature, 410,* 268–276.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46,* 265–294.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13,* 480–484.

Vicente, S., Castro, S. L., & Walley, A. (2003). A developmental analysis of similarity neighborhoods for European Portuguese. *Journal of Portuguese Linguistics, 2,* 115–133.

Virtual round table on ten leading questions for network research [Editorial]. (2004). *European Physical Journal, B*, *38,* 143–145.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech, 40,* 211–228.

Vitevitch, M. S. (2002a). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 735–747.

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

Vitevitch, M. S. (2006, November–December). *The clustering coefficient of phonological neighborhoods influences spoken word recognition.* Poster presented at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9,* 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374–408.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers, 36,* 481–487.

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

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language, 68,* 306–311.

Vitevitch, M. S., & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders, 3,* 64–73.

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31,* 491–504.

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes, 21,* 760–770.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13,* 286–350.

Ward, L. M. (2002). *Dynamical cognitive science.* Cambridge, MA: MIT Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.

Watts, D. J. (1999). *Small worlds: The dynamics of networks between order and randomness*. Princeton, NJ: Princeton University Press.

Watts, D. J. (2004). The "new" science of networks. *Annual Review of Sociology, 30,* 243–270.

Watts, D. J., & Strogatz, S. H. (1998, June 4). Collective dynamics of "small-world" networks. *Nature, 393,* 440–442.

Wellman, B., & Wortley, S. (1990). Different strokes from different folks: Community ties and social support. *American Journal of Sociology, 96,* 558–588.

Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18,* 303–324.

Wilks, C., Meara, P., & Wolter, B. (2005). A further note on simulating word association behaviour in a second language. *Second Language Research, 21,* 359–372.

Yates, M., Locker, L., Jr., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review, 11,* 452–457.

Contact author: Michael S. Vitevitch, Spoken Language Laboratory, Department of Psychology, 1415 Jayhawk Boulevard, University of Kansas, Lawrence, KS 66045. E-mail: mvitevit@ku.edu.