American Psychological Association

Journal of Experimental Psychology: General

ISSN: 0096-3445

2020, Vol. 149, No. 12, 2376–2394
http://dx.doi.org/10.1037/xge0000876

An Investigation of Network Growth Principles in the Phonological Language Network

Cynthia S. Q. Siew
National University of Singapore

Michael S. Vitevitch
University of Kansas

This article investigated how network growth algorithms—preferential attachment, preferential acquisition, and lure of the associates—relate to the acquisition of words in the phonological language network, where edges are placed between words that are phonologically similar to each other. Through an archival analysis of age-of-acquisition norms from English and Dutch and word learning experiments, we examined how new words were added to the phonological network. Across both approaches, we found converging evidence that an *inverse* variant of preferential attachment—where new nodes were instead more likely to attach to existing nodes with *few* connections—influenced the growth of the phonological network. We suggest that the inverse preferential attachment principle reflects the constraints of adding new phonological representations to an existing language network with already many phonologically similar representations, possibly reflecting the pressures associated with the processing costs of retrieving lexical representations that have many phonologically similar competitors. These results contribute toward our understanding of how the phonological language network grows over time and could have implications for the learning outcomes of individuals with language disorders.

Keywords: language networks, word learning, language development, network growth, phonology

The learning of new words occurs throughout life. It is estimated that an average adult will accumulate 30,000 to 80,000 words (Aitchison, 2012) within his or her lifetime. By no means does this imply that learning new words is a trivial cognitive phenomenon. Word learning is a complex process that involves different stages and is influenced by many cognitive as well as linguistic factors (Gaskell & Ellis, 2009; Gupta & Tisdale, 2009; Storkel & Lee, 2011).

Past research has shed much light on the cognitive and linguistic factors that influence how children and adults learn new words. For example, children with a small number of words in their vocabulary or who have small working memory capacities were poorer at learning novel words (Hansson, Forsberg, Löfqvist, Mäki-Torkko, & Sahlén, 2004; Nash & Donaldson, 2005). The nature of the child's learning environment also plays a crucial role in facilitating vocabulary acquisition. For instance, young children learn new words more quickly when their parents name objects found in the learning environment (Tan & Schafer, 2005). Finally, the lexical characteristics of words, such as the frequency with which sounds in a word occur in the language and the number of known words that sound like the word being learned, have also been shown to influence different aspects of word learning (Storkel, Armbrüster, & Hogan, 2006; Storkel & Lee, 2011).

In most previous work there is an implicit assumption that words are learned independently of each other, despite obvious semantic and phonological relationships that exist among words. Previous studies did not explicitly consider how the relationships that exist among words affect word learning, but more recent work suggests that these relationships can be exploited to predict which words children will learn next (Beckage & Colunga, 2013; Beckage, Smith, & Hills, 2011; Bilson, Yoshida, Tran, Woods, & Hills, 2015; Stella, Beckage, & Brede, 2017).

One way to represent the interrelationships that exist among words is to construct a network of words using the principles found in network science. Network science is an emerging interdisciplinary field where networks are used to study complex systems in biology, telecommunications, cognition, and social domains (for a review, see Barabási, 2009). In the networks used to represent these complex systems, nodes represent entities (e.g., people) and connections represent relationships between pairs of entities (e.g., friendship links).

Networks can also be used to examine the semantic or phonological relationships among words in the mental lexicon, that part of long-term memory that contains information about the words we know in a given language (Steyvers & Tenenbaum, 2005; Vitevitch, 2008). In both types of lexical networks, nodes represent words. In the semantic network, connections are placed between pairs of words that are semantically related to each other (i.e., they are similar in meaning), whereas in the phonological network, connections are placed between pairs of words that are phonolog-

This article was published Online First June 25, 2020.

Cynthia S. Q. Siew, Department of Psychology, National University of Singapore; Michael S. Vitevitch, Department of Psychology, University of Kansas.

All data and scripts associated with the present article are available on the Open Science Framework (https://osf.io/68p2z/).

Correspondence concerning this article should be addressed to Cynthia S. Q. Siew, Department of Psychology, National University of Singapore, 9 Arts Link, Block AS4 #02-23, Singapore 117572. E-mail: cynthia@nus.edu.sg

ically similar to each other (i.e., they sound similar to each other). These connections are also known as edges in the network science terminology, and the number of connections of edges incident on a node corresponds to its degree. In a phonological network, the degree of a word corresponds to its number of phonological neighbors.

The application of network science to study cognition has greatly increased our understanding of memory- and language-related processes (for a review, see Siew, Wulff, Beckage, & Kenett, 2019). A central tenet in network science is the idea that the structure of the complex system governs the processes and functions that operate within that system (Strogatz, 2001). In the context of the cognitive sciences, network analysis can be applied to quantify and measure the structure of cognitive systems and permit investigations into how cognitive processes such as lexical retrieval and acquisition function in connection with the underlying structure of the mental lexicon. From a number of such studies we have learned that the structure of the network of phonological word forms has measurable influences on speech perception (Chan & Vitevitch, 2009), speech production (Chan & Vitevitch, 2010), word learning (Goldstein & Vitevitch, 2014), and short-term memory (STM) processes (Vitevitch, Chan, & Roodenrys, 2012). Note, however, that these studies have implicitly assumed that the lexical network is static; that is, the number of nodes and connections existing within the network are fixed. In reality, a person's lexical network continues to grow as we encounter new words and add them to our vocabularies. Indeed, a growing number of papers has demonstrated how network science methods can be used to investigate changes in lexical networks during typical (Hills, Maouene, Maouene, Sheya, & Smith, 2009; Sizemore, Karuza, Giusti, & Bassett, 2018) and atypical (Beckage et al., 2011) language acquisition, as well as across the life span (Dubossarsky, De Deyne, & Hills, 2017). However, these studies have almost exclusively focused on the development of semantic networks, where edges between words reflect free associations, shared features, or word co-occurrence statistics.

Hills et al. (2009) examined how several network growth algorithms—preferential attachment (PATT), preferential acquisition (PACQ), and lure of the associates (LA)—might influence the acquisition of new words and development of the semantic network. The network growth algorithms examined by Hills et al. (2009) reflect different ways in which the structure of the learner's mental lexicon and of his or her learning environment can be exploited to facilitate the acquisition of a new word. As shown by Hills et al. (2009) and many others (Beckage & Colunga, 2013; Peters & Borovsky, 2019; Sailor, 2013; Stella et al., 2017), the semantic relationships that exist among words in the mental lexicon have significant impact on the acquisition of new words.

Although previous corpus analyses of normative vocabularies have investigated how PATT and PACQ growth algorithms influence the acquisition of semantic representations (e.g., Beckage & Colunga, 2013; Sailor, 2013), it is not known how or if these network growth algorithms influence the acquisition of new *phonological* representations. As pointed out by Hills and Siew (2018), different network growth models predict learning on networks with different edges. For instance, growth models based on network distinctiveness predicts acquisition in a semantic network constructed from the semantic features of words (Engelthaler & Hills, 2017; Sizemore et al., 2018), whereas growth models based on PACQ predicts learning in semantic networks constructed from free associations and co-occurrences in child-directed language (Hills, Maouene, Riordan, & Smith, 2010). Collectively, these studies reveal that different kinds of "connective tissue" in language, operationalized as the different types of edges that could be constructed between words in a language network, matter in different ways during the process of language acquisition. However, gaps in our understanding of how these network growth models influence the acquisition of *phonological* word forms remain as these previous studies did not examine these network growth models with respect to the phonological similarity relationships between words. Hence, the goal of the present work is to examine if network growth models, specifically, PATT, PACQ, and LA, influence the acquisition of new phonological representations.

Network Growth Algorithms

In this section, each of the three network growth algorithms (lure of associates, PATT, PACQ) proposed by Hills et al. (2009) will be described in greater detail, as specifically related to the acquisition of new phonological word forms. Figure 1 provides a summary and examples to illustrate each of the three network growth algorithms. The LA network growth algorithm predicts that new words that are phonologically similar to *several* known words are learned better than new words that are phonologically similar to few known words (Figure 1a). In the word learning literature, past work has manipulated the phonological similarity of a nonword (a made-up word that does not exist in English) to other known words. An example of a nonword that is phonologically similar to many known English words is "fash" (phonological neighbors with 18 English words in the adult phonological network, such as *fish*, *fat*, *ash*, *bash, cache, fad,* and *flash*), whereas an example of a nonword that is phonologically similar to few known English words is "vuch" (phonological neighbors with 7 English words in the adult phonological network, such as *dutch*, *touch*, *vouch*, and *much*). In network science terms this would correspond to the number of phonological connections the node representing the novel word would form with known words (Gaskell & Dumay, 2003; Stamer & Vitevitch, 2012; Storkel, 2001, 2003; Storkel et al., 2006). Thus, these previous studies examined the LA network growth algorithm even though they did not call it that. These studies found that nonwords that were phonologically similar to many known words were more likely to be acquired—being phonologically similar to many known words helps the novel word to be better integrated within the lexicon during the integration phase of word learning. In contrast, to the best of our knowledge, there are no studies that have directly investigated the PATT or PACQ network growth algorithms with respect to the learning of new phonological representations.

The PATT network growth algorithm predicts that a novel word that is similar to a known word that itself is phonologically similar to *many* other known words will be better learned than a novel word that is similar to a known word that itself is phonologically similar to few other known words (Figure 1b). Following Hills et al. (2009)'s definition, PATT growth values indicate the average number of phonological neighbors of known words that the novel word would be phonologically similar to. Words with high PATT values are predicted to be more likely learned compared to words with low PATT values. An example of a nonword with a high

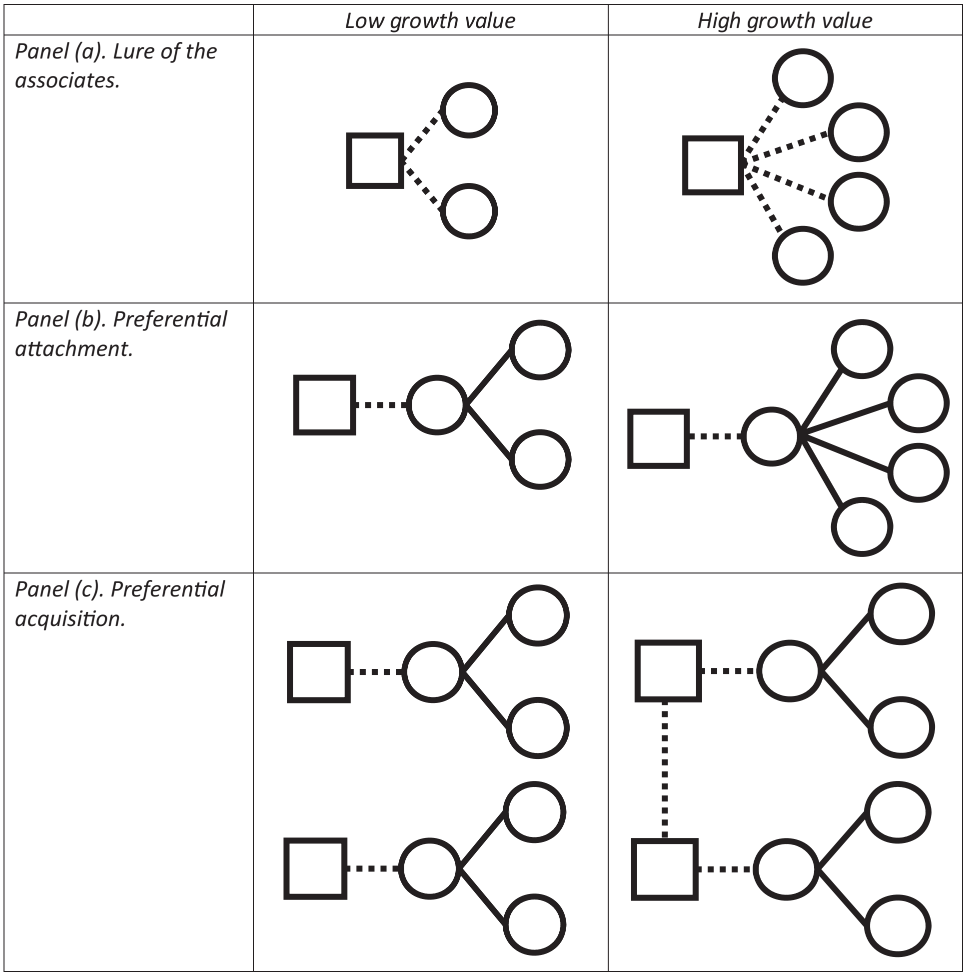

Figure 1. Squares represent novel words; circles represent known words. Links are placed between known words and novel words that are phonologically similar to each other. a: Lure of the associates (LA). The left side shows a "low" LA novel word, where few known words are phonologically similar to it; the right side shows a "high" LA novel word, where several known words are phonologically similar to it. b: Preferential attachment. The left side of the figure shows a "low" preferential attachment (PATT) novel word, which connects to a known word with few connections; the right side shows a "high" PATT novel word, which connects to a known word with several connections. c: Preferential acquisition. The left side of the figure shows a pair of "low" preferential acquisition (PACQ) novel words, which are not connected to each other; the right side shows a pair of "high" PACQ novel words, which are connected to each other.

PATT value is "*waub*" because it is phonologically similar to a word (i.e., *web*) with many phonological neighbors (i.e., *wet, ebb, ware, wed, wedge,* and *well*), whereas an example of a nonword with a low PATT value is "choing" because it is phonologically similar to a word (i.e., *choice*) with few phonological neighbors (i.e., *voice*).

It is important to emphasize the subtle—but crucial—difference between LA and PATT. In LA, the difference between a new word with high and low LA growth values is the number of known words that are phonological neighbors of the new word. However, consider the case where the novel word is only similar to one known word. A high novel PATT word is similar to a known word that has many neighbors, whereas a low novel PATT word is similar to a known word that has few neighbors. Note that the LA network growth algorithm would not predict any differences between the high novel PATT word and the low novel PATT because both words are only similar to one known word (i.e., both have a LA growth value of 1). However, the PATT network growth algorithm predicts that the high novel PATT word is more likely to be acquired—being similar to a word with many phonological neighbors might facilitate acquisition because the presence of overlapping phonological sequences (i.e., words with high phonotactic probability) might help to integrate the new phonological representation into the lexicon. Specifically, words with high phonotactic probability, being composed of commonly occurring sound sequences, may incur fewer cognitive resources when its novel phonological representation is held in STM, freeing up resources for lexical and semantic processing which facilitate word-learning (Storkel, 2001).

The PACQ network growth algorithm predicts that new words that are phonologically similar to several new words in the learning environment are better learned than new words that are phonologically similar to few new words in the learning environment (Figure 1c). Hills et al. (2009) defined the *learning environment* as the set of new words that a person is trying to learn, that is, the training set, as well as words that are already known to a learner. PACQ growth values indicate the number of presently known words *as well as words that will eventually be learned* that the novel word would be phonologically similar to (Hills et al., 2009). Another way to think about PACQ values is that they are based on the "adult" lexicon, or the "end-state" lexicon that is the target of child language acquisition (i.e., the larger, adult lexicon that the child lexicon is developing into) rather than the current state of the child/learner's lexicon, which serves as a reasonable approximation for the learning environment that children are exposed to over the course of language development, as children interact with their caregivers and listen to adults communicate with each other. Based on this definition, an example of a nonword with a high PACQ value is "mub" because it is phonologically similar to many words in the adult lexicon (i.e., *mud*, *mob*, and *moon*), whereas an example of a nonword with a low PACQ value is "zaz" because it is phonologically similar to few words in the adult lexicon (i.e., *czar*). If we expect that phonological relationships in the learning environment play a role in the acquisition of new phonological representations, akin to the findings of Hills et al. (2009) with respect to the semantic network, words with high PACQ values may be more salient in the language environment and are predicted to be learned better than words with low PACQ values.

At this point, it is important to emphasize that each of the three network growth algorithms have been adapted from generative network growth models in network science to provide testable model predictions with respect to the acquisition of semantic and phonological representations in a language network. Without necessarily considering what the network representation itself is (i.e., whether it is a network of friendships among people or phonological word forms of language), such generative network growth models provide predictions about which connections will be created when a new node is added to the network. These network growth models have been adapted in the present context to focus on how different structural aspects of the network (i.e., the connections in the network) might affect the selection of new nodes to be added to that network from a population of candidate nodes, in line with previous investigations done in the context of the semantic network (e.g., Hills et al., 2009). In this case, network growth models are applied to investigate growth in the phonological lexicon and provide us with a viable framework where one can begin to address theoretically important questions regarding how the higher-order structure of the lexicon (i.e., beyond that of a novel word's immediate connections; e.g., Storkel et al., 2006) and how the structure of the language environment modulate the role of phonological similarity in word learning. We note that these are questions that would not have been asked otherwise without the application of network analysis and network growth models to understand the processes that constrain the development of the early lexicon.

In this article, we examined how these network growth algorithms could account for the acquisition of novel phonological word-forms using a combination of archival analysis of existing datasets, and word learning experiments. In Study 1, an archival analysis of two large-scale age-of-acquisition norms available for English (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) and Dutch words (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014) was conducted to determine which network growth algorithm better predicts whether a particular word was more likely to be learned. As a brief preview, one of the key results of Study 1 indicated that the effect of the PATT growth algorithm appeared to differ over the course of language development. This intriguing result led us to conduct a word-learning experiment (Study 2) to determine if novel words with high or low growth values are more likely to be learned by participants. The word learning experiment allowed greater control over the characteristics of nonwords to explicitly test the influence of PATT and PACQ growth algorithms. Together, these different approaches reveal how growth of the phonological lexicon is driven by a network growth algorithm that initially prioritizes the acquisition of phonological word forms that are similar to highly connected words in the phonological network, but as the lexicon matures the acquisition of phonological word forms that are similar to less connected words in the phonological network is emphasized instead.

Study 1: Large-Scale Archival Analysis of Age of Acquisition Norms

We capitalized on the existence of large data sets of age-of-acquisition norms in two languages (English and Dutch) to examine which network growth algorithms were significant predictors of learning. In this section, we follow the approach taken by Hills et al. (2009) and computed network growth values for each word at different time points, which represented the "prediction" of each of the network growth algorithms, with higher growth values indicating that a given word was hypothesized to be more likely acquired in the next time step based on the network growth principle. Logistic regression was conducted to determine which network growth values were significant predictors of age-of-acquisition ratings, after controlling for the influence of the length of the word and its usage frequency in the language. The aim of the present set of analysis was to evaluate each network growth algorithm by considering whether its corresponding growth value was associated with a higher probability or likelihood of acquiring a particular word.

Method

Data and materials.

English data. To approximate when an English word was learned we used the age of acquisition (AoA) ratings reported in Kuperman et al. (2012). This dataset consisted of AoA ratings for over 30,000 English words obtained via the Amazon Mechanical Turk platform. The values in Kuperman et al. (2012) are the mean AoA ratings (referring to the year at which a particular word was thought to be learned) averaged across participants.

Phonological transcriptions for English words were obtained from the Hooiser Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984), which consisted of 19,340 phonological word-forms that were examined in the phonological network by Vitevitch (2008). This collection of words has been used extensively in previous work on speech perception and spoken word recognition (e.g., Luce & Pisoni, 1998). Approximately ~13,000 English words were found in both data sets. The phonological transcriptions of these ~13,000 words were used to construct phonological networks of English words (see below).

Dutch data. To approximate when a Dutch word was learned we used the AoA ratings reported in Brysbaert et al. (2014). This dataset consisted of AoA ratings for over 30,000 Dutch words obtained from native speakers of Dutch who lived in Ghent, Belgium. The values in Brysbaert et al. (2014) are the mean AoA ratings (referring to the year at which a particular word was thought to be learned) averaged across participants.

Phonological transcriptions for Dutch words were obtained from the CLEARPOND Dutch database (http://clearpond.northwestern.edu/; Marian, Bartolotti, Chabal, & Shook, 2012), which consisted of ~27,000 phonological word-forms. Approximately ~15,000 Dutch words were found in both data sets. The phonological transcriptions of these ~15,000 words were used to construct phonological networks of Dutch words (see below).

Longitudinal network analysis.

Network construction. In the phonological network, nodes represent phonological word forms and connections are placed between two words that are phonologically similar to each other based on the one-edit distance metric (i.e., substitution, deletion, or addition of one phoneme; Luce & Pisoni, 1998), consistent with the approach taken by Vitevitch (2008). Separate phonological networks of known words were constructed for each year (specifically, Years 3, 4, 5, 6, 7, 8, and 9).

At each year, the AoA dataset was filtered to exclude words that were yet to be learned such that this new set consisted of words that were already acquired by that point of time (i.e., known words). For instance, at year 4, 236 English words had an AoA rating <4 and these words were used to construct the "known lexicon" at Year 4. This network was used to compute the PATT and LA growth values for the unlearned words (13,234 − 236 = 12,998) at Year 4 (see below). Note that PATT and LA growth values for unlearned words could be different for different years because these growth values reflect the structural properties of the known lexicon (which itself changes over time). PACQ growth values were simply the degree of words in the final (full) network, where all the words are "known" (i.e., the "adult" lexicon). The entire dataset was used to approximate the adult lexicon for English and Dutch, respectively. To keep the analysis tractable, we restricted the analysis to words with an AoA rating less than 10 years. Therefore, "known" phonological networks were constructed for English and Dutch words that were known by Years 3, 4, 5, 6, 7, 8, and 9, respectively.

Growth values for network growth models. Growth values for the PATT, PACQ, and the LA network growth algorithms were based on the algorithms used in Hills et al. (2009). The PACQ network growth algorithm posits that novel words that are phonologically similar to several words in the learning environment are learned earlier than novel words that are phonologically similar to few words in the learning environment. We define the *learning environment* as the set of words that will be eventually known by the learner, that is, the final "adult" lexicon, which is represented by the entire word list. The growth values of yet-to-be-learned, novel words (at any given time point) were simply the degree (i.e., number of phonological neighbors) of the word in the adult network that was constructed from words in the final adult lexicon. The PACQ network growth algorithm predicts that words with high PACQ values are more likely learned as compared to words with low PACQ values.

We also computed PACQ growth values based on a smaller subset of words with AoA ratings <12 years to approximate the "child" learning environment (as in Hills et al., 2009) and conducted the same analysis using these values. As the pattern of results was qualitatively similar to the analysis conducted using growth values obtained from the adult lexicon, we report the analysis conducted using PACQ growth values obtained from the adult lexicon below.

PATT growth values indicate the average number of phonological neighbors of *known words* that the novel word is phonologically similar to. This was computed in the following manner: A known network is constructed based on the words that are known to the learner at each time point. For each novel (yet-to-be-learned) word, we obtained the list of words in the known lexicon that it would connect to if it were learned and computed the mean degree of those words in the known lexicon. The PATT network growth algorithm predicts that words with high PATT values are more likely learned as compared to words with low PATT values at each time point. One key issue to note is that PATT growth value for each novel word changes over time, because it depends on the connectivity of the words that are already known, which changes as new words are added to the learner's vocabulary.

We also computed PATT growth values based on the maximum degree of the set of known words (as in Hills et al., 2009) and conducted the same analysis using these values. As the pattern of results was qualitatively similar to the analysis conducted using PATT growth values obtained by taking the average degree of known words, we report the analysis conducted using PATT growth values obtained by taking the average degree of known words below.

LA growth values indicate the number of words in the known lexicon that the novel word is phonologically similar to. This was computed in the following manner: A known network is constructed based on the words that are known to the learner at each time step. For each novel word, we simply counted the number of words in the known lexicon that it would connect to if it were learned. Given previous findings in the word learning literature (Storkel et al., 2006), we predict that words with high LA values are more likely learned as compared to words with low LA values at each time step. It is important to note that LA and PATT network growth algorithms are different in subtle ways and lead to different growth values. Unlike PATT, LA ignores the connectivity (i.e., degree) of the known words that the novel word is "attaching to" (i.e., the number of phonological neighbors the known words have in the known lexicon) and refers to the number of new connections that will be made if the novel word were to be added to the known lexicon.

These growth values were then compiled such that each row in the final dataset consisted of a novel, "yet-to-be-learned" word at the beginning of each year (range = 3 to 9), together with its length and frequency, growth values for PATT and LA (computed based on the known network at that point of time) and PACQ (computed based on the final network), and a final column that indicated whether the word was eventually learned at the end of that year or not (binary outcome). Table 1 shows the descriptive statistics for lexical characteristics and network growth values for

Table 1
Descriptive Statistics for Lexical Characteristics and Network Growth Values for Unknown Words at a Given Year Based on Age of Acquisition Norms

Variable	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
English words (*N*)	13,202	12,998	12,523	11,823	11,007	10,000	8,755
Length							
M	6.15	6.19	6.27	6.38	6.50	6.62	6.76
SD	2.28	2.27	2.26	2.25	2.24	2.24	2.26
Min	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Max	16.00	16.00	16.00	16.00	16.00	16.00	16.00
Frequency							
M	1.86	1.84	1.78	1.71	1.63	1.56	1.47
SD	0.92	0.89	0.84	0.78	0.73	0.69	0.65
Min	0.30	0.30	0.30	0.30	0.30	0.30	0.30
Max	6.18	5.98	5.86	5.51	5.51	5.51	5.51
PATT							
M	0.005	0.124	0.442	0.772	1.014	1.181	1.256
SD	0.093	0.532	1.282	2.029	2.554	2.953	3.176
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	2.00	6.00	10.00	15.00	17.00	21.00	30.00
PACQ							
M	3.06	2.93	2.72	2.43	2.17	1.94	1.75
SD	5.70	5.56	5.32	4.99	4.64	4.33	4.12
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	36.00	36.00	36.00	36.00	36.00	36.00	33.00
LA							
M	0.03	0.19	0.42	0.66	0.82	0.93	1.01
SD	0.18	0.63	1.22	1.83	2.30	2.58	2.80
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	3.00	8.00	12.00	18.00	21.00	21.00	25.00
Dutch words (*N*)	15,024	14,972	14,646	13,865	12,583	10,675	8,203
Length							
M	7.09	7.10	7.16	7.27	7.38	7.51	7.60
SD	2.25	2.25	2.23	2.19	2.17	2.16	2.18
Min	1.00	1.00	1.00	1.00	1.00	1.00	2.00
Max	17.00	17.00	17.00	17.00	17.00	17.00	17.00
Frequency							
M	1.55	1.54	1.49	1.39	1.30	1.19	1.09
SD	1.31	1.30	1.22	1.08	0.97	0.85	0.74
Min	0.36	0.36	0.36	0.36	0.36	0.36	0.36
Max	10.59	10.10	9.14	8.61	8.39	6.48	6.48
PATT							
M	0.001	0.008	0.105	0.302	0.450	0.559	0.667
SD	0.037	0.090	0.504	1.079	1.495	1.780	2.077
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	1.00	2.00	8.00	14.00	13.50	15.00	18.00
PACQ							
M	1.75	1.73	1.62	1.42	1.28	1.12	0.99
SD	3.65	3.62	3.49	3.21	3.00	2.80	2.59
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	29.00	29.00	29.00	28.00	28.00	28.00	27.00
LA							
M	0.00	0.03	0.15	0.32	0.45	0.54	0.60
SD	0.05	0.18	0.57	1.09	1.44	1.73	1.92
Min	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max	2.00	4.00	7.00	12.00	16.00	21.00	23.00

Note. PATT = preferential attachment; PACQ = preferential acquisition; LA = lure of the associates.

unknown words at a given year based on AoA norms. Probability distributions of length and frequency of English and Dutch words analyzed in Study 1 can also be found in Appendix A.

Results

A logistic regression model was fitted to the data in order to determine which network growth algorithms were significant predictors of acquiring a word, after controlling for the length of the word and its usage frequency in the language. To put it in another way, the success of each network growth algorithm was evaluated by considering whether its corresponding growth value was associated with a higher probability or likelihood of acquiring that particular word. The dependent variable is binary (1 if the word was learned in the subsequent year and 0 if the word was not learned in the subsequent year). The predictors include word length (number of phonemes), log frequency obtained from word frequency counts based on film subtitles (SUBTLEX-NL for Dutch; Keuleers, Brysbaert, & New, 2010; SUBTLEX-EN for English; Brysbaert & New, 2009), and growth values for PATT, PACQ, and LA (all of which were mean centered and scaled), and the year of learning (3, 4, 5, 6, 7, 8, 9; mean-centered continuous variable).

The "full" model that included interaction terms between year and all other predictors (lexical predictors: length, frequency; growth values: PATT, PACQ, LA) was first constructed. The inclusion of the interaction terms allowed us to assess the possibility that the magnitude of the effect of network growth algorithms (and lexical predictors) varied over time. The *step*() function in R implemented both forward and backward stepwise search algorithms (by adding or removing individual predictors and assessing changes in Akaike information criterion [AIC]) to find the set of variables that led to the most parsimonious model that minimized AIC while maximizing model fit.

Comparing the influence of different network growth algorithms. Before looking more specifically at whether each network growth algorithm individually were significant predictors of the acquisition of new phonological representations, it is important to assess if network growth algorithms were significant predictors of probability of learning, even after taking into account the influence of length and word frequency, which were previously known to be associated with AoA. We constructed five models for the English AoA norms and five models for the Dutch AoA norms and conducted a series of likelihood ratio tests to compare the model fits to the data. Model 0 included the following interactions and their main effects: Word Length × Year and Frequency × Year. Models 1, 2, and 3 included the individual predictors of the three different network growth algorithms respectively. Model 1 (PATT) included the predictors in Model 0 and the PATT × Year interaction. Model 2 (PACQ) included the predictors in Model 0 and the PACQ × Year interaction. Model 3 (lure of associates) included the predictors in Model 0 and the LA × Year interaction. Model 4 represented the "full" model that included interaction terms between year and all other predictors (lexical predictors: length, frequency; growth values: PATT, PACQ, and LA; i.e., the "full" model later submitted to the stepwise regression procedure described below).

The following comparisons were examined using the likelihood ratio test for nested models: Model 0 versus Models 1, 2, and 3, respectively, to compare each network growth algorithm individually to the baseline model, and Model 4 versus Models 1, 2, and 3, respectively, to compare the model containing all three network growth algorithms against models with a single growth algorithm. The results of the likelihood ratio test are summarized in Table 2 below. Overall, it appears that all three network growth algorithms are statistically significant predictors of word learning beyond the influence of lexical variables. Based on this result, we then submitted this full model (Model 4) to the stepwise regression search procedure.

English AoA data. The final model obtained from the stepwise regression search is shown in Table 3. The following predictors were retained: Length, year, frequency, PATT, LA, as well as the following interactions, Length × Year, Frequency × Year, PATT × Year. The McFadden R^2 for the final logistic model was 0.259. A number of main effects and interaction terms were significant, which suggest that the effects of these predictors change over time. The following main effects were significant: length, year, frequency, PATT growth value. Length was a significant predictor of acquisition, $b = -0.248$, $p < .001$, such that longer words were associated with lower probability of acquisition. Year was a significant predictor of acquisition, $b = 0.433$, $p < .001$, indicating that words with later AoA ratings were associated with a higher probability of acquisition at a given time step. Frequency was a significant predictor of acquisition, $b = 1.446$, $p < .001$, such that highly frequent words were associated with a higher probability of acquisition. Finally, PATT growth value was a significant predictor of acquisition, $b = 0.075$, $p = .003$, such that words with greater PATT growth values were associated with a greater probability of acquisition. However, it is important to note that the main effects of length and PATT are qualified by an interaction with year and should interpreted with respect to the simple slopes analysis reported below. We further probe the nature of interaction effects in the final model by plotting

Table 2
Results of Likelihood Ratio Tests for Various Models

Model	Model 1 (PATT)	Model 2 (PACQ)	Model 3 (LA)
English			
Model 0: Baseline	$\chi^2 = 34.67$ $df = 2$ $p < .001$	$\chi^2 = 1.40$ $df = 2$ $p = .50$	$\chi^2 = 28.29$ $df = 2$ $p < .001$
Model 4: Full model (PATT, PACQ, LA)	$\chi^2 = 4.13$ $df = 4$ $p = .39$	$\chi^2 = 37.39$ $df = 4$ $p < .001$	$\chi^2 = 10.50$ $df = 4$ $p = .03$
Dutch			
Model 0: Baseline	$\chi^2 = 53.66$ $df = 2$ $p < .001$	$\chi^2 = 66.51$ $df = 2$ $p < .001$	$\chi^2 = 81.03$ $df = 2$ $p < .001$
Model 4: Full model (PATT, PACQ, LA)	$\chi^2 = 54.45$ $df = 4$ $p < .001$	$\chi^2 = 41.59$ $df = 4$ $p < .001$	$\chi^2 = 27.07$ $df = 4$ $p < .001$

Note. PATT = preferential attachment; PACQ = preferential acquisition; LA = lure of the associates.

Table 3
Final Logistic Regression Model With Lexical and Network Growth Values to Predict Acquisition (English Age-of-Acquisition Norms)

Predictors	Estimate	*SE*	*z*	*p*
(Intercept)	−4.536	0.099	−45.96	<.001***
Length	−0.248	0.013	−19.59	<.001***
Year	0.433	0.045	9.61	<.001***
Frequency	1.446	0.022	64.58	<.001***
PATT	0.075	0.025	3.00	.003**
LA	0.026	0.017	1.50	.13
Length × Year	0.031	0.005	5.83	<.001***
Frequency × Year	0.020	0.010	1.96	.05*
PATT × Year	−0.050	0.009	−5.88	<.001***

Note. PATT = preferential attachment growth values; LA = lure of associates growth values.
* $p < .05$. ** $p < .01$. *** $p < .001$.

probability curves and conducting a simple slope analysis for each interaction.

Length × Year interaction. Figure 2 shows how the probability of learning a word (with values on all other predictors [i.e., frequency, all growth values] fixed at their mean) changed over time and how it is modulated by the length of the word. Simple slopes computed at three different time points (years = 3, 6, 9) were all significant (year = 3: $b = -0.33$, $z = -13.62$, $p < .001$; year 6: $b = -0.24$, $z = -20.21$, $p < .001$; year 9: $b = -0.15$, $z = -10.34$, $p < .001$). The negative slopes indicated that shorter words were more likely to be learned as compared to long words, and this effect increased over time such that shorter words were even more likely to be learned at the later years.

Frequency × Year interaction. Figure 3 shows how the probability of learning a word (with values on all other predictors (i.e., length, all growth values) fixed at their mean) changed over time and how it is modulated by word frequency. Simple slopes computed at three different time points (years = 3, 6, 9) were all significant (year = 3: $b = 1.39$, $z = 34.13$, $p < .001$; year 6: $b = 1.45$, $z = 66.56$, $p < .001$; year 9: $b = 1.51$, $z = 43.19$, $p < .001$). The positive slopes indicated that frequent words were more likely to be learned as compared to less frequent words, and this effect increased over time such that frequent words were even more likely to be learned at the later years.

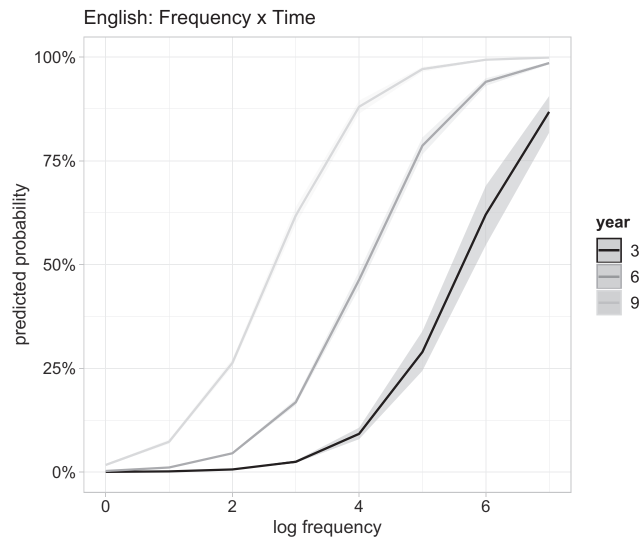

Figure 3. Predicted probability curves depicting the Frequency × Time interaction effect (English age-of-acquisition norms).

PATT Growth Value × Year interaction. Figure 4 shows how the probability of learning a word (with values on all other predictors [i.e., length, frequency, other growth values] fixed at their mean) changed over time and how it is modulated by the PATT growth value. Simple slopes computed at three different time points (years = 3, 6, 9) were all significant (year = 3: $b = 0.21$, $z = 4.86$, $p < .001$; year 6: $b = 0.06$, $z = 2.62$, $p = .01$; year 9: $b = -0.09$, $z = -3.95$,

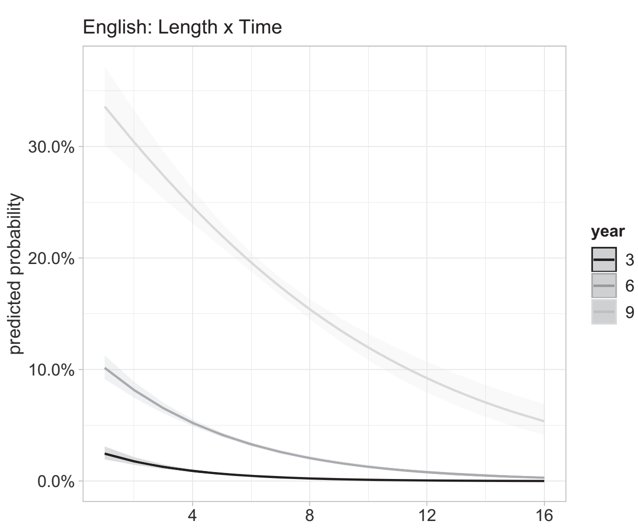

Figure 2. Predicted probability curves depicting the Length × Time interaction effect (English age-of-acquisition norms).

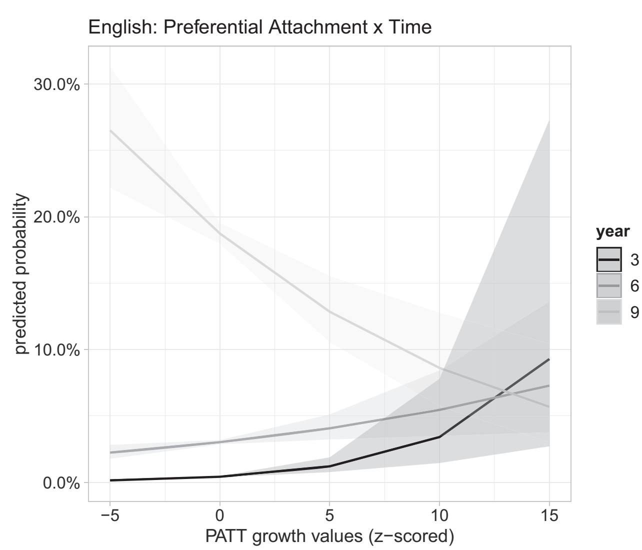

Figure 4. Predicted probability curves depicting the Preferential Attachment × Time interaction effect (English age-of-acquisition norms).

$p < .001$). The pattern of the interaction indicated that words with higher PATT values were more likely to be learned as compared to words with lower PATT values at the early stages, whereas the direction of the effect flipped later on such that words with lower PATT values were more likely to be learned as compared to words with higher PATT values at the later stages.

Dutch AoA data. The final model obtained from the stepwise regression search is shown in Table 4. The following predictors were retained: Length, year, frequency, PATT, PACQ, LA, as well as the following interactions, Length × Year, Frequency × Year, PATT × Year, PACQ × Year. The McFadden R^2 for the final logistic model was 0.235. A number of main effects and interaction terms were significant, which suggest that the effects of these predictors change over time. The following main effects were significant: length, year, frequency, PATT, PACQ, and LA growth values. Length was a significant predictor of acquisition, $b = -0.161$, $p < .001$, such that longer words were associated with lower probability of acquisition. Year was a significant predictor of acquisition, $b = 0.401$, $p < .001$, indicating that words with later AoA ratings were associated with a higher probability of acquisition at a given time step. Frequency was a significant predictor of acquisition, $b = 0.532$, $p < .001$, such that highly frequent words were associated with a higher probability of acquisition. PATT growth value was a significant predictor of acquisition, $b = 0.047$, $p = .04$, such that words with greater PATT growth values were associated with a greater probability of acquisition. PACQ growth value was a significant predictor of acquisition, $b = 0.067$, $p = .002$, such that words with greater PACQ growth values were associated with a greater probability of acquisition. Finally, LA growth value was a significant predictor of acquisition, $b = 0.049$, $p = .03$, such that words with greater LA growth values were associated with a greater probability of acquisition. However, it is important to note that the main effects of length, frequency, PATT, and PACQ are qualified by an interaction with year and should interpreted with respect to the simple slopes analysis reported below. We further probe the nature of these interactions by plotting probability curves and conducting a simple slope analysis for each interaction below. Unlike the English analysis, both PACQ and LA were significant predictors in the model. The inclusion of PACQ growth variable in the final model indicated that PACQ was a significant predictor of acquiring a new word, $z = 3.17$, $p = .002$, Wald confidence interval [1.03, 1.11], such that for every 1 *SD* increase in the PACQ growth value, the odds of learning is multiplied by 1.07, holding all other variables constant. In other words, the larger the PACQ growth value (i.e., higher phonological similarity to words in the learning environment), the higher the odds of learning the new phonological word-form. The inclusion of the LA growth variable in the final model indicated that LA was also a significant predictor of acquiring a new word, $z = 2.19$, $p = .03$, Wald confidence interval [1.01, 1.10], such that for every 1 *SD* increase in the LA growth value, the odds of learning is multiplied by 1.05, holding all other variables constant. In other words, the larger the LA growth value (i.e., more connections from the novel word to words that were already acquired), the higher the odds of learning the new phonological word-form.

Table 4
Final Logistic Regression Model With Lexical and Network Growth Values to Predict Acquisition (Dutch Age-of-Acquisition Norms)

Predictors	Estimate	*SE*	*z*	*p*
(Intercept)	−2.555	0.080	−31.92	<.001***
Length	−0.161	0.011	−15.03	<.001***
Year	0.401	0.035	11.31	<.001***
Frequency	0.532	0.011	49.63	<.001***
PATT	0.047	0.023	2.04	.04*
PACQ	0.067	0.021	3.17	.002**
LA	0.049	0.022	2.19	.03*
Length × Year	0.045	0.005	10.00	<.001***
Frequency × Year	0.038	0.006	6.78	<.001***
PATT × Year	−0.043	0.008	−5.43	<.001***
PACQ × Year	0.017	0.011	1.55	.12

Note. PATT = preferential attachment growth values; PACQ = preferential acquisition growth values; LA = lure of associates growth values.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Length × Year interaction. Figure 5 shows how the probability of learning a word (with values on all other predictors [i.e., frequency, all growth values] fixed at their mean) changed over time and how it is modulated by the length of the word. Simple slopes computed at three different time points (years = 3, 6, 9) were significant for Year 3 and Year 6, but not Year 9 (year = 3: $b = -0.28$, $z = -13.19$, $p < .001$; year 6: $b = -0.14$, $z = -15.17$, $p < .001$; year 9: $b = -0.01$, $z = -0.86$, $p = .39$). The negative slopes indicated that shorter words were more likely to be learned as compared to long words, and the magnitude of this effect decreased over time such that the effect of length was nonsignificant at the final time point.

Frequency × Year interaction. Figure 6 shows how the probability of learning a word (with values on all other predictors (i.e., length, all growth values) fixed at their mean) changed over time and how it is modulated by word frequency. Simple slopes computed at three different time points (years = 3, 6, 9) were all significant (year = 3: $b = 0.43$, $z = 20.97$, $p < .001$; year 6: $b = 0.54$, $z = 52.40$, $p < .001$; year 9: $b = 0.66$, $z = 34.21$, $p < .001$). The positive slopes indicated that frequent words were more likely to be learned as compared to less frequent words, and that the magnitude of the frequency effect on learning increased over time.

PATT Growth Value × Year interaction. Figure 7 shows how the probability of learning a word (with values on all other predictors [i.e., length, frequency, other growth values] fixed at their mean) changed over time and how it is modulated by the PATT growth value. Simple slopes computed at three different time points (years = 3, 6, 9) were significant for Years 3 and 9 but not Year 6 (year = 3: $b = 0.16$, $z = 3.91$, $p < .001$; year 6: $b = 0.03$, $z = 1.52$, $p = .13$; year 9: $b = -0.10$, $z = -5.31$, $p < .001$). The pattern of the interaction indicated that words with higher PATT values were more likely to be learned as compared to words with lower PATT values at the early stages, whereas the direction of the effect flipped later on such that words with lower PATT values were more likely to be learned as compared to words with high PATT values at the later stages.

Discussion

In Study 1 we examined how well PATT, PACQ, and the LA network growth algorithms predicted the acquisition of novel words at each year. This was done for two different data sets (English and Dutch age-of-acquisition norms), and the pattern of findings was very similar across both languages. The results dem-

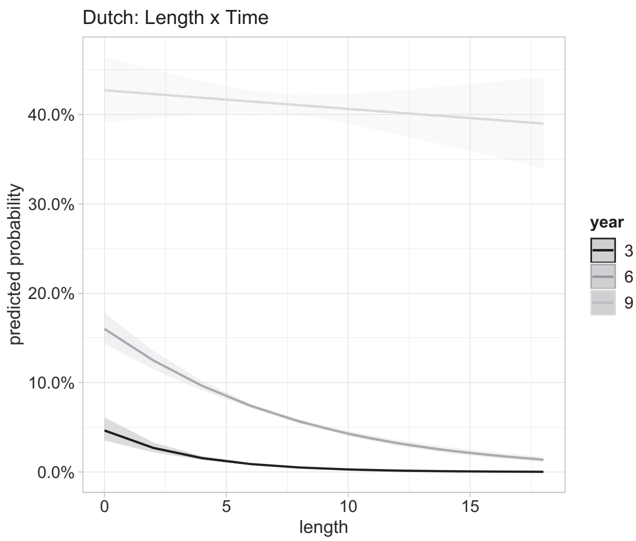

Figure 5. Predicted probability curves depicting the Length × Time interaction effect (Dutch age-of-acquisition norms).

onstrate that network growth algorithms are statistically significant predictors of word learning beyond the influence of various lexical variables. When considered individually, all three network growth algorithms were significant predictors of whether a word would be learned or not, after controlling for the effects of lexical variables (based on the results of a series of likelihood ratio tests; see Table 2). When all network growth algorithms were included in the same model and submitted to a stepwise regression procedure to search for the best performing model, PATT, and not PACQ, remained a significant predictor of word learning for the English norms, whereas all three network growth algorithms remained significant predictors of word learning for the Dutch norms.

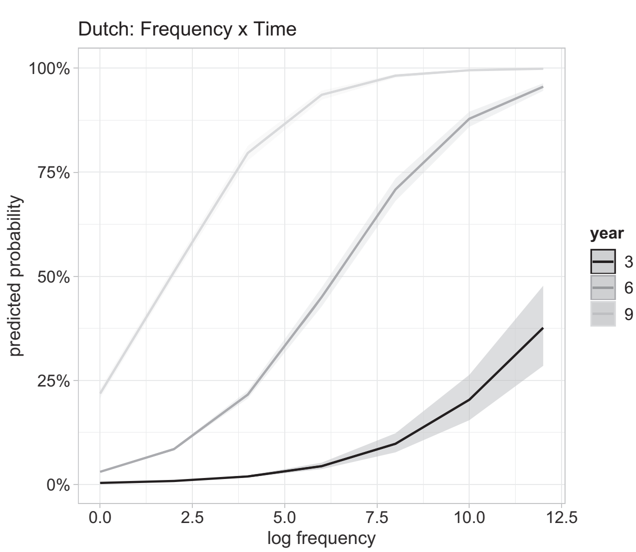

Figure 6. Predicted probability curves depicting the Frequency × Time interaction effect (Dutch age-of-acquisition norms).

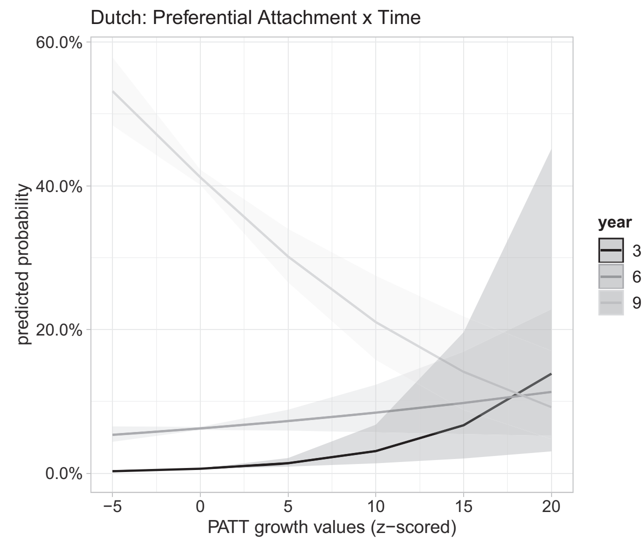

Figure 7. Predicted probability curves depicting the Preferential Attachment × Time interaction effect (Dutch age-of-acquisition norms).

The pattern of results related to the LA and PACQ growth algorithms in predicting Dutch AoA norms are in line with the predictions previously discussed in the Introduction. With respect to the LA growth algorithm, more connections from the novel word to words that were already acquired was associated with higher odds of learning the new phonological word-form, suggesting that the formation of more connections to existing lexicon enhances the process of integrating the new word form into the lexicon as previously demonstrated in prior word learning experiments (e.g., Storkel et al., 2006). With respect to the PACQ growth algorithm, higher phonological similarity to words in the learning environment was associated with higher odds of learning the new phonological word-form, consistent with previous findings in the context of the semantic network where highly connected words in the learning environment may be more salient and act as a "signal" for initiating word learning (Hills et al., 2009).

The most intriguing result from the analyses was that the effect of PATT appeared to change over time in both English and Dutch analyses. In the early stages of development the odds of learning a word increased when the growth value of PATT increased; however, in the later stages of development (∼after 6 or 7 years) the odds of learning a word decreased when the growth value of PATT increased. The directionality of the odds ratio for the PATT effect in the early years indicates that novel words that are phonologically similar to known words that are themselves phonologically similar to several other known words are more likely to be learned, where a higher degree of phonological similarity facilitates the acquisition of new words. One possibility is that novel words with high PATT growth values tend to possess phonological sequences that are common to known words in the lexicon, which may help with retaining the new representation in STM and promote integration into the lexicon.

However, at the later years the directionality of the odds ratio changed, such that novel words that are phonologically similar to known words that are themselves phonologically similar to few other known words are more likely to be learned. The change in the influence of PATT over time suggests that the PATT growth model may need to be revised to account for the observed reversal in odds ratios. The opposite pattern observed for word learning in later years might indicate that there is a cost to learning new words that sound similar to words that already have several phonological neighbors. This cost may become more apparent in a more mature lexicon where a great deal of phonological similarity already exists among known words (and is known to induce greater levels of competition among activated word forms in spoken word recognition tasks; e.g., Luce & Pisoni, 1998), making phonological similarity a less useful strategy for word learning than in the earlier years. Network scientists have observed in other real-world systems that the growth of a system can be affected by certain costs and constraints associated with creating new connections between nodes (Amaral, Scala, Barthelemy, & Stanley, 2000; Clune, Mouret, & Lipson, 2013), so this is not an unreasonable suggestion.

The present analysis provided us with some novel and interesting results. However, as is the case with all analyses of archival data, there are certain limitations. For example, the AoA ratings reported by Kuperman et al. (2012) and Brysbaert et al. (2014) were self-reports obtained from adults, so it is not clear how accurate one's memory is for exactly when a particular word was learned, especially words learned in early childhood. In addition, the growth values used in the present corpus analysis were computed on phonological networks where edges were placed between words that were phonologically similar based on the one-edit distance metric. Although this is a commonly adopted operationalization of phonological similarity that reflects people's perceived phonological similarity (Luce & Pisoni, 1998; Vitevitch & Luce, 2016), the present findings might look different should a less conservative definition of phonological similarity were used to construct network edges.

Importantly, the results of Study 1 provide specific hypotheses regarding how the mechanisms of PATT and PACQ drive word learning that we investigate empirically in the next section. Specifically, the hypotheses are that with respect to a mature lexicon, (a) unlearned words with higher PATT growth values are less likely to be acquired as compared to unlearned words with lower PATT growth values, and (b) no differences are expected for unlearned words with higher or lower PACQ growth values. To test these specific hypotheses, we conducted a word learning experiment using carefully selected stimuli matched on a number of relevant variables but varying on the growth values of the network growth algorithms. The use of made-up words (i.e., nonwords) as stimuli allowed us to control for the impact of various lexical characteristics that have been shown to influence word learning (Storkel, 2004).

Study 2: Word-Learning Experiment

A standard word-learning task was used to assess the ability of college-aged adults to learn a small set of specially created nonwords (Goldstein & Vitevitch, 2014; Gupta, 2003; Mirman, Magnuson, Estes, & Dixon, 2008; Storkel, 2001; Storkel et al., 2006; Xu & Tenenbaum, 2007). Based on the findings in Study 1, we predicted that college-aged participants in the present word-learning experiment would learn the same proportion of nonwords with high and low PACQ growth values. As for nonwords with high and low PATT growth values, we predicted that our college-age participants (who would presumably have a large, mature lexicon) would show an advantage for learning nonwords with low rather than high PATT growth values, consistent with the result observed in Study 1 for the later years that we examined. The experiments reported below have been approved by the University of Kansas Institutional Review Board.

Method

Participants. Thirty-six native American English speakers were recruited from the introductory psychology subject pool at the University of Kansas. All participants had no previous history of speech or hearing disorders and received partial course credit for their participation.

Stimuli. A total of 24 nonwords, 12 for the PATT set and 12 for the PACQ set, were selected from a corpus of legal consonant-vowel-consonant nonwords (Storkel, 2013). A list of the nonword stimuli and their relevant characteristics can be found in Appendix B. All of the selected nonwords were phonologically similar to only one known word in the network examined in Vitevitch (2008; i.e., they all had a LA growth value of 1). Two words were considered phonologically similar, or phonological neighbors of each other, if the first word could be transformed to the other via the substitution, deletion, or addition of a single phoneme in any position (Luce & Pisoni, 1998). Note that constraining the number of phonological neighbors each nonword had to 1 represented an important departure from prior word learning studies that manipulated the *number* of real word neighbors of each nonword (e.g., Storkel et al., 2006), although such a restriction limited the number of nonwords that were used in the present experiment. Recall that these studies found that nonwords that were phonologically similar to more known words were better learned than nonwords that were phonologically similar to few known words—a finding that is consistent with the LA network growth algorithm described by Hills et al. (2009). As prior work has previously shown evidence for the lure of associates growth principle in word learning (even though it was not explicitly labeled as such; Storkel & Lee, 2011; Storkel et al., 2006), we focus our present investigation on the PATT and PACQ growth algorithms, which have not been investigated within an experimental setting. In the present word learning study, we constrained all nonwords to be phonologically similar to *one* known word so that it precludes the LA network growth algorithm as a possible explanation for the results of this experiment.

The 12 PATT nonwords were divided into two groups: six nonwords had high PATT growth values and six nonwords had low PATT growth values. For the PATT set, the crucial difference between nonwords with high and low PATT growth values was the *phonological similarity of the known word* that the nonwords were phonological neighbors of [$M_{\text{high}} = 11.00$, $SD_{\text{high}} = 3.52$; $M_{\text{low}} = 5.00$, $SD_{\text{low}} = 2.10$; $t(10) = 3.59$, $p < .01$]. Given that the selected nonwords only have a single real word neighbor, the PATT growth value for each nonword is simply the degree of its real word neighbor. A high PATT word would be phonologically similar to a known word that is phonologically similar to many other known words, whereas a low PATT word would be phonologically similar to a known word that is phonologically similar to few other known words. Both high and low PATT nonwords were also matched on

PACQ growth values; $M_{\text{high}} = 2.00$, $SD_{\text{high}} = 1.10$; $M_{\text{low}} = 1.00$, $SD_{\text{low}} = 0.63$; $t < 1.94$, $p = .08$.

The 12 PACQ nonwords were also divided into two groups: six nonwords had high PACQ growth values and six nonwords had low PACQ growth values. For the PACQ set, the crucial difference between the high and low PACQ growth values was the number of nonwords each nonword is phonologically similar to within the learning set ($M_{\text{high}} = 3.00$, $SD_{\text{high}} = 0$; $M_{low} = 1.00$, $SD_{\text{low}} = 0$). A high PACQ word would be phonologically similar to *many* other nonwords in the learning set, whereas a low PACQ word would *not* be phonologically similar to any nonwords in the learning set. Both high and low PACQ nonwords were also matched on PATT growth values, $M_{\text{high}} = 11.83$, $SD_{\text{high}} = 7.41$; $M_{\text{low}} = 10.63$, $SD_{\text{low}} = 3.27$; $t < 1$, $p = .73$. Recall that the PACQ growth values refer to the degree of the target word in the learning environment, which in this case refers to the set of 12 PACQ nonwords.

PATT and PACQ nonwords were also matched on other variables such as phonotactic probability (Vitevitch & Luce, 2004), and stimulus file duration; all ts < 1.71, ps $> .10$. Table 5 shows the descriptive statistics for each set of nonwords. A list of the nonwords for each condition is included in the Appendix.

A male native speaker of American English (second author) produced the nonwords at a normal speaking rate into a high-quality microphone in an Industrial Acoustics Company sound-attenuated booth. The digital recording was transferred to a hard drive and SoundEdit16 was used to create, edit, and normalize individual sound files for each nonword.

To provide participants with a referent, each nonword was matched to a nonobject obtained from Kroll and Potter (1984). These nonobjects were constructed by combining parts of drawings of real objects to create closed figures that did not look like existing objects in the real world. Semantic set size, concreteness, and first and second word associate strength (Storkel & Adlof, 2009) of selected nonobjects were controlled across conditions in both sets; all ts < 1.55, ps $> .15$. Table 6 shows the object-likeness ratings and word associate strength ratings for nonobjects in each condition.

Procedure. Participants took part in two experimental sessions that were scheduled between 2 to 5 days apart ($M = 2.64$ days, $SD = 0.96$ days). In the first session, participants learned either the PATT or the PACQ nonword set, and the other set in the second session. The order in which participants learned each set of nonwords was counterbalanced. Each participant was tested individually. Stimuli were presented via BeyerDynamic DT100 headphones at a comfortable listening level and PsyScope 1.2.2 was used to randomize and control the presentation of stimuli.

The word learning paradigm, which is an established paradigm that has been successfully used in several previous word learning studies (Stamer & Vitevitch, 2012; Storkel et al., 2006), consisted of two components: a learning phase and a test session. In the learning phase, participants listened to nonwords while looking at pictures of nonobjects. Each nonword was paired with a nonobject and used in short sentences that were played over the headphones. The following carrier phrases were used: “Look at the (nonword),” “Don’t forget the (nonword),” “Listen closely, it’s called a (nonword),” “This is a (nonword),” and “Remember, it’s a (nonword).”

The test session consisted of a picture-naming task where each nonobject was presented on the screen and participants were instructed to respond with the nonword that was paired with that nonobject. Participants were instructed to say “don’t know” or “pass” if they were unable to name the nonobject. Verbal responses were recorded for offline scoring of accuracy. No corrective feedback was provided to the participants.

During each experimental session, participants performed a baseline picture-naming task, followed by three rounds of learning and testing phases (see Table 7). The baseline picture-naming task was conducted prior to any learning sessions to ensure that participants did not know the names of the nonobjects. In the first training session, one exposure of nonwords and nonobjects was presented. An exposure consisted of presenting the nonobject on the screen while the nonword associated with it was embedded within a single short sentence that was played over the headphones. In the subsequent two training sessions, five exposures of the nonwords were presented. The nonobject was presented on the screen while the nonword associated with the nonobject was embedded in five different short sentences that played over the headphones. At all three test sessions, participants performed the picture-naming task.

Results

Table 8 shows the raw accuracy data compiled from all participants for each training block of the two word learning experiments. Overall, accuracy increased over blocks, indicating that participants were able to successfully acquire the nonword labels for nonobjects. A logistic regression model was fit to the data in

Table 5
Descriptive Statistics for Characteristics of Nonwords

	Preferential attachment		Preferential acquisition	
Characteristic	High	Low	High	Low
Segment probability	0.034 (0.014)	0.035 (0.015)	0.094 (0.0085)	0.083 (0.037)
Biphone probability	0.00040 (0.00064)	0.00038 (0.0060)	0.00078 (0.00062)	0.00087 (0.00079)
Onset duration (ms)	57 (3)	57 (2)	54 (2)	57 (2)
Word duration (ms)	586 (42)	588 (48)	597 (52)	564 (57)
File duration (ms)	700 (41)	704 (46)	707 (55)	675 (58)
Growth values of network growth algorithms				
PATT (degree of real word associate)	11.00 (3.52)	5.00 (2.10)	11.83 (7.41)	10.67 (3.27)
PACQ (degree based on learning environment)	2.00 (1.10)	1.00 (0.63)	3.00 (0)	1.00 (0)

Note. PATT = preferential attachment growth values; PACQ = preferential acquisition growth values. Standard deviations are provided in parentheses.

Table 6
Descriptive Statistics for Characteristics of Selected Nonobjects

Characteristic	Preferential attachment		Preferential acquisition	
	High	Low	High	Low
Semantic set size	10.33 (0.52)	10.50 (0.55)	10.67 (0.52)	10.33 (0.52)
Concreteness ratings	4.58 (0.71)	4.65 (1.42)	4.67 (0.74)	4.22 (1.00)
First word associate strength	0.17 (0.07)	0.13 (0.07)	0.11 (0.03)	0.18 (0.11)
Second word associate strength	0.12 (0.04)	0.09 (0.05)	0.08 (0.03)	0.10 (0.3)

Note. Standard deviations are provided in parentheses.

order to determine if PATT and PACQ growth values were significant predictors of correctly producing the nonword in the word learning task, after controlling for various covariates. Separate regression models were constructed for the experiment with PATT nonwords and the experiment with PACQ nonwords. For both models the dependent variable was binary (0 if the nonword was incorrectly produced or the participant said, “don’t know” and 1 if the nonword was correctly produced). The covariates included phonotactic probability (segment and biphone probability), nonword duration (as obtained from the audio recordings), trial number within each block, and the counterbalancing condition that each participant was assigned to (i.e., the order in which the two experiments were completed; PATT followed by PACQ or PACQ followed by PATT). The key predictors of interest were the nonwords’ growth values for PATT and PACQ, and the test block (baseline, 1, 2, 3). Analyzing the data via a generalized linear model allowed us to include growth value as a continuous variable in the model.

PACQ nonwords. The generalized linear model for the PACQ word learning experiment is shown in Table 9. The following predictors were significant: trial, segment and biphone probability, and block. Cox and Snell’s R^2 for the model was 0.218. Notice that the PACQ growth variable was not a significant predictor, $p = .186$, indicating that the value of the PACQ growth was not a significant predictor of whether the nonword was likely to be correctly produced or not.

PATT nonwords. The generalized linear model for the PATT word learning experiment is shown in Table 10. The following predictors were significant: Trial, duration, block, and PATT growth value. Cox and Snell’s R^2 for the model was 0.221. The PATT growth variable was as significant predictor of correctly producing the nonword, $z = -3.88$, $p < .001$, Wald confidence interval [0.65, 0.87], such that for every 1 *SD* increase in the PATT growth value, the odds of producing the nonword is multiplied by 0.75, holding all other variables constant. In other words, the larger the PATT growth value, the lower the odds of learning the nonword.

Table 7
Timeline of Experimental Sessions for Study 2

Session 1	Session 2
Nonword set 1	Nonword set 2
Baseline test	Baseline test
First training (1 exposure)	First training (1 exposure)
First test	First test
Second training (5 exposures)	Second training (5 exposures)
Second test	Second test
Third training (5 exposures)	Third training (5 exposures)
Third test	Third test

Note. Session 2 took place 2 to 5 days after Session 1. The order in which participants learned the preferential attachment growth values and preferential acquisition growth values nonwords was counterbalanced.

Discussion

The results from the present word-learning experiment were consistent with the findings reported from the analysis of the AoA ratings reported in Study 1. Recall that in Study 1, the effect of PACQ on the odds of learning a word was not significant for English words (although it was significant for Dutch words, we note that the effect size was very small). In the present word-learning experiment, there was no difference in the proportion of high and low PACQ nonwords learned by the end of the experiment.

As for PATT, in Study 1 the effect of PATT changed over time, such that in the earlier years words with higher PATT growth values were more likely to be learned, whereas at the later years words with lower PATT growth values were more likely to be learned. This pattern was found for both English and Dutch words. Given that the participants were college students who presumably possess a large and stable lexicon, the present finding that words with lower PATT growth values were more likely to be learned is consistent with the result in Study 1 whereby words with lower PATT growth values were more likely learned than words with higher PATT growth values in the later years of development.

The principle of PATT is an important notion in network science as many complex networks investigated by network scientists exhibit a scale-free degree distribution, which is characterized by the presence of a few, highly connected “hubs” in a network, as well as many nodes in the network with very few connections.

Table 8
Proportion of Correct Trials Across Each Block and for Each Condition

Condition	Baseline	Block 1	Block 2	Block 3
Preferential attachment				
High	0.000	0.056	0.231	0.398
Low	0.000	0.079	0.236	0.491
Preferential acquisition				
High	0.000	0.083	0.329	0.472
Low	0.000	0.083	0.301	0.486

Table 9
Logistic Regression Model for the Preferential Acquisition Word Learning Experiment

Predictors	Odds ratios	Confidence interval	z	p
(Intercept)	0.04	[0.02, 0.06]	−14.57	<.001***
Counterbalance order	1.05	[0.81, 1.36]	0.35	.725
Trial	0.93	[0.89, 0.96]	−3.89	<.001***
Duration	1.08	[0.91, 1.29]	0.87	.386
Segment probability	1.4	[1.07, 1.84]	2.42	.016*
Biphone probability	0.63	[0.47, 0.83]	−3.2	.001**
Block	3.62	[3.12, 4.24]	16.46	<.001***
PACQ value	0.89	[0.74, 1.06]	−1.32	.186

Note. PACQ = preferential acquisition growth values.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Barabási and Albert (1999) suggested that such scale-free distributions arise in these networks because the systems grow over time following the principle of PATT. The claim of Barabási and Albert is intriguing because it suggests that complex networks from very different domains may develop in a similar manner, whereby new nodes are more likely to connect to existing nodes with several connections rather than to nodes with few connections (Steyvers and Tenenbaum (2005) suggested that the semantic network is one such system). Note, however, that other work suggests that a number of network growth algorithms—not just growth via PATT—can lead to networks with scale-free degree distributions (Fox Keller, 2005).

Notice that the effect of PATT observed in many other systems is different from the effect of PATT that we observed in the archival analysis and the present word-learning experiment. In most real-world networks, new nodes are more likely to connect to existing nodes with several connections rather than to nodes with few connections (leading to a scale-free network). In our word-learning experiment (and in the later years of the archival analysis in Study 1) a word was more likely to be learned if it was similar to a word that had few connections instead of being similar to a word that had many connections.

Further note that the phonological networks of English (Vitevitch, 2008) and several other languages (Arbesman, Strogatz, & Vitevitch, 2010) do not exhibit a scale-free degree distribution. Just as Amaral et al. (2000; see also Clune et al., 2013) suggested that the growth of a system can be affected by certain costs and constraints associated with creating new connections between nodes, Arbesman et al. (2010) suggested that certain linguistic constraints, such as the phoneme inventory of the language or phonotactic restrictions, may have affected growth in the phonological lexicons, leading to a degree distribution that is not scale-free.

The results of Study 1 and Study 2 suggest that phonological similarity may be another "cost" to consider in the growth of the phonological network. Indeed, the work of Luce and Pisoni (1998) and others on spoken word recognition has demonstrated that there are processing costs in adults for words being phonologically similar to many other words. Perhaps the "processing costs" associated with being phonologically similar to many words exerts pressure—at later stages of development—on the word-learning process to instead acquire word-forms that would not further crowd the "phonological space" of a more mature lexicon.

General Discussion

In the present article, we examined how various network growth algorithms (PATT, PACQ, LA) influenced the acquisition of new phonological word-forms. In Study 1, an analysis of age-of-acquisition ratings from English and Dutch showed that growth values of words based on PATT significantly predicted the odds of learning those words at subsequent time steps, and the direction of the effect reversed over time. Words with higher PATT values were more likely to be learned at earlier years whereas words with lower PATT values were more likely to be learned at later years.

In Study 2, a word learning experiment showed that novel words with low PATT growth values compared to novel words with high PATT growth values were more likely to be learned. The results of these studies suggested that an inverse variant of PATT influenced the acquisition of new phonological word-forms. In the inverse PATT growth algorithm, words that are phonologically similar to known words with few phonological neighbors were more likely to be learned than words that are phonologically similar to known words with many phonological neighbors.

The present article was to some extent motivated by the work done by Hills and colleagues (2009; see also Hills et al., 2010), who found that PACQ, rather than PATT or LA network growth algorithms, was associated with the growth of early normative semantic networks. As the learning of new words involves the acquisition of new semantic as well as phonological information, we felt that it was important to investigate how these growth algorithms might also play a role in the development of the phonological network. Interestingly, the present results suggest that the network growth algorithms underlying the acquisition of phonological word-forms are quite different from the ones that underlie acquisition in the semantic network. PATT rather than PACQ appeared to have a greater influence on the growth of the phonological language network. Furthermore, although the PATT network growth algorithm was associated with the acquisition of new phonological word-forms in early phonological networks, an inverse form of the PATT network growth algorithm appeared to drive the acquisition of new phonological word-forms as the lexicon matures.

Recall that PATT predicts that new nodes are more likely to connect to nodes that have several connections (Barabási & Albert, 1999). In the present context, PATT predicts that words are more likely to be learned if they are phonologically (or semantically) similar to words that are themselves phonologically (or semantically) similar to several other words in the language network. The

Table 10
Logistic Regression Model for the Preferential Attachment Word Learning Experiment

Predictors	Odds ratios	Confidence interval	z	p
(Intercept)	0.02	[0.01, 0.03]	−15.26	<.001***
Counterbalance order	1.13	[0.85, 1.49]	0.84	.402
Trial	0.95	[0.91, 0.99]	−2.53	.012*
Duration	1.59	[1.36, 1.86]	5.85	<.001***
Segment probability	0.91	[0.78, 1.05]	−1.28	.201
Biphone probability	0.93	[0.79, 1.10]	−0.81	.416
Block	3.99	[3.37, 4.77]	15.55	<.001***
PATT value	0.75	[0.65, 0.87]	−3.88	<.001***

Note. PATT = preferential attachment growth values.
* $p < .05$. *** $p < .001$.

inverse PATT network growth algorithm predicts that words are more likely to be learned if they are phonologically (or semantically) similar to words that are phonologically (or semantically) similar to *few* other words in the language network.

In the case of PACQ, nodes that connect to several other words in the learning environment are more likely to be integrated into the network—words that are phonologically (or semantically) similar to several other words in the learning environment (i.e., set of novel words) are more likely to be integrated into the mental lexicon (Hills et al., 2009). The finding that PACQ was associated with the growth of early semantic networks by Hills and colleagues, but seemed to play less of a role in the growth of phonological networks in the present work may indicate that environmental factors are relatively more influential in facilitating the acquisition of semantic information as compared to phonological word-forms, at least in the early stages of language development. That is, the presence of certain important objects or concepts in the environment drives acquisition in the semantic network toward the adult state of the lexicon more than it does in the phonological network. One is more likely to acquire the name of an object or a concept if caregivers and other people in the ambient environment are discussing those important objects or concepts (Tan & Schafer, 2005).

In contrast, the structure of the existing lexical network may be more influential in the acquisition of phonological word-forms than in the acquisition of semantic information. Indeed, Shriberg (1993) has observed a developmental trend in the acquisition of phonemes by children; the early-8 sounds are acquired (/m, b, n, w, d, p, j, h/), then the middle-8 (/t, ŋ, k, g, f, v, ʧ, ʤ/), then the late-8 (/ʃ, s, θ, ð, r, z, l, ʒ/). Typically, one does not acquire one of the late-8 sounds before one acquires the early-8 sounds. In the case of phonology, then, there appears to be some dependence on previously acquired information. In the case of semantic information, however, one does not need to have already acquired the concept for *dog* before one can acquire the concept for *cat*. One can acquire the concept for *cat* before *dog* or can acquire both concepts simultaneously.

Although it is generally recognized that there are many processes involved in the acquisition of a phonological word-form at a single point in time (e.g., triggering, configuration, and engagement in Storkel & Lee (2011); see also Leach and Samuel (2007) and Davis and Gaskell (2009)), the present results suggest that there also might be many different learning mechanisms employed in the acquisition of phonological word-forms over time. Previous studies together with the present findings suggest that the portfolio of different learning mechanisms that learners possess includes LA (e.g., Storkel & Lee), PACQ, PATT in the early stages, and the novel mechanism that we called *inverse PATT*. Relying on a portfolio of learning mechanisms rather than just a single learning mechanism may allow for the acquisition of a "cocktail" of novel words with varying levels of phonological similarity to already known words. For instance, the LA growth algorithm might be important for acquiring novel words that sound similar to several known words, whereas inverse PATT might be important for acquiring novel words that are phonologically distinctive. In fact, the structure of the phonological language network itself reflects the presence of words that occupy densely populated phonological spaces as well as words that occupy more sparsely populated phonological spaces. Given that both near- and far-phonological neighbors have been shown to affect the process of spoken word recognition (e.g., Vitevitch, Chan, & Goldstein, 2014), it is important to consider how various growth algorithms and learning strategies interact to give rise to the unique structure of the phonological language network.

Interestingly, the dominant strategy employed to acquire phonological word-forms may change as the structure of the phonological lexicon itself changes. Early in development new words were more likely to attach to existing words with more connections (i.e., many phonological neighbors), whereas later in development new words were more likely to attach to existing words with few connections (i.e., few phonological neighbors). As shown in Study 1, in the earlier years, words that were phonologically similar to known words with several phonological neighbors were more likely to be learned. However, by approximately 6 or 7 years of age, the reverse was true—words that were phonologically similar to known words with few phonological neighbors were more likely to be learned. Perhaps as the phonological space in the growing lexicon becomes more crowded, the processing costs associated with discriminating among many phonological neighbors during spoken-word recognition (Luce & Pisoni, 1998) may trigger a change in which network growth algorithm becomes the dominate strategy used in that stage of development (for other changes that crowding of the phonological space might initiate, see the lexical restructuring hypothesis; Garlock, Walley, & Metsala, 2001; Metsala & Walley, 1998). Although the influence of certain constraints or costs on growth has been examined previously by network scientists—financial and space limitations constrain the expansion of air transportation networks (Amaral et al., 2000), and the skull limits growth of networks in the brain (Mitchison & Barlow, 1991)—the present findings point to a new area for future exploration, namely how several different costs introduced at different time-points might shape future growth.

The observation that learners may possess a portfolio of different learning mechanisms for the acquisition of phonological word-forms has interesting implications for phonological language disorders. Perhaps some learners, instead of responding to the pressure to change to a different learning mechanism, may persist in using the same strategy they have been using, even though it is no longer optimal to do so. Consider that Beckage et al. (2011) found that children who are late talkers have a bias toward acquiring words that are semantically different from known words, whereas Hills et al. (2009) found that typically developing children tended to first acquire words that are semantically similar to known words. Beckage et al. (2011) suggested that the bias to acquire words that were semantically different from known words resulted in the formation of a mental lexicon with less connectivity among words, which may have led to inefficient lexical retrieval and processing later in life. Further investigation is required to determine if a similar use of a nonoptimal strategy might underlie certain phonological language disorders.

The present results point to some interesting differences and similarities in how the semantic and phonological networks grow. An interesting direction for future research is to consider how "costs" in one network might influence growth in the other. Relatedly, what processing costs are introduced when the forms in the phonological network are first linked with their meanings in the semantic network in the early stages of development (Naigles, 2002)? The structure of the early phonological network may play a crucial role in scaffolding the acquisition of new semantic information, and vice versa. Given that "structure always affects function" (Strogatz, 2001; p. 268), we believe the present results demonstrate that network science provides researchers with a powerful set of computational tools to examine the structure of the mental lexicon, and to examine how the growth of that structure over time might further affect other language-related processes.

Context of the Research

Network science provides a way for researchers to see how the parts of a system are related to each other, and how the way in which those parts are organized might influence processing in that system. The present work examined how the network of words we know, referred to as the *mental lexicon*, grows over time as we learn new words throughout our lifetime. Through two different approaches (large-scale archival analysis and laboratory-based experiments) we found that several algorithms are involved in the learning of new words, and that the optimal algorithm at a given time may be influenced by how "crowded" the word-space currently is. Future work may examine if persisting with a nonoptimal algorithm may play a factor in certain well-studied speech sound and language disorders. Growing networks may also hold much promise for researchers studying other areas of development.

References

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. Oxford, UK: Wiley.

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11149–11152. http://dx.doi.org/10.1073/pnas.200327197

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering, 20,* 679–685. http://dx.doi.org/10.1142/S021812741002596X

Barabási, A.-L. (2009). Scale-free networks: A decade and beyond. *Science, 325,* 412–413. http://dx.doi.org/10.1126/science.1173299

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286,* 509–512. http://dx.doi.org/10.1126/science.286.5439.509

Beckage, N., & Colunga, E. (2013). Using the words toddlers know now to predict the words they will learn next. *Proceedings of the Annual Meeting of the Cognitive Science Society*. Advance online publication. Retrieved from https://escholarship.org/uc/item/8sk4t2c0

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE, 6*(5), e19348. http://dx.doi.org/10.1371/journal.pone.0019348

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition, 140,* 122–134. http://dx.doi.org/10.1016/j.cognition.2015.03.013

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

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica, 150,* 80–84. http://dx.doi.org/10.1016/j.actpsy.2014.04.010

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1934–1949. http://dx.doi.org/10.1037/a0016902

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science, 34,* 685–697. http://dx.doi.org/10.1111/j.1551-6709.2010.01100.x

Clune, J., Mouret, J.-B., & Lipson, H. (2013). The evolutionary origins of modularity. *Proceedings of the Royal Society B: Biological Sciences, 280,* 20122863. http://dx.doi.org/10.1098/rspb.2012.2863

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 3773–3800. http://dx.doi.org/10.1098/rstb.2009.0111

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology, 53,* 1560–1570. http://dx.doi.org/10.1037/dev0000347

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: Network distinctiveness predicts age of acquisition. *Cognitive Science, 41*(Suppl. 1), 120–140. http://dx.doi.org/10.1111/cogs.12350

Fox Keller, E. F. (2005). Revisiting "scale-free" networks. *BioEssays, 27,* 1060–1068. http://dx.doi.org/10.1002/bies.20294

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45,* 468–492. http://dx.doi.org/10.1006/jmla.2000.2784

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132. http://dx.doi.org/10.1016/S0010-0277(03)00070-2

Gaskell, M. G., & Ellis, A. W. (2009). Word learning and lexical development across the lifespan. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 3607–3615. http://dx.doi.org/10.1098/rstb.2009.0213

Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: How groups of similar sounding words facilitate acquisition. *Frontiers in Psychology, 5,* 1307. http://dx.doi.org/10.3389/fpsyg.2014.01307

Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 56,* 1213–1236. http://dx.doi.org/10.1080/02724980343000071

Gupta, P., & Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 3755–3771. http://dx.doi.org/10.1098/rstb.2009.0132

Hansson, K., Forsberg, J., Löfqvist, A., Mäki-Torkko, E., & Sahlén, B. (2004). Working memory and novel word learning in children with hearing impairment and children with specific language impairment. *International Journal of Language & Communication Disorders, 39,* 401–422. http://dx.doi.org/10.1080/13682820410001669887

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63,* 259–273. http://dx.doi.org/10.1016/j.jml.2010.06.002

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science, 20,* 729–739. http://dx.doi.org/10.1111/j.1467-9280.2009.02365.x

Hills, T. T., & Siew, C. S. Q. (2018). Filling gaps in early word learning. *Nature Human Behaviour, 2,* 622–623. http://dx.doi.org/10.1038/s41562-018-0428-y

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42,* 643–650. http://dx.doi.org/10.3758/BRM.42.3.643

Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal* *of Verbal Learning & Verbal Behavior, 23,* 39–66. http://dx.doi.org/10.1016/S0022-5371(84)90499-7

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44,* 978–990. http://dx.doi.org/10.3758/s13428-012-0210-4

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55,* 306–353. http://dx.doi.org/10.1016/j.cogpsych.2007.01.001

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19,* 1–36. http://dx.doi.org/10.1097/00003446-199802000-00001

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE, 7*(8), e43230. http://dx.doi.org/10.1371/journal.pone.0043230

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. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum Publishers.

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108,* 271–280. http://dx.doi.org/10.1016/j.cognition.2008.02.003

Mitchison, G., & Barlow, H. B. (1991). Neuronal branching patterns and the economy of cortical wiring. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 245,* 151–158. http://dx.doi.org/10.1098/rspb.1991.0102

Naigles, L. R. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition, 86,* 157–199. http://dx.doi.org/10.1016/S0010-0277(02)00177-4

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language, and Hearing Research, 48,* 439–458. http://dx.doi.org/10.1044/1092-4388(2005/030)

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

Peters, R., & Borovsky, A. (2019). Modeling early lexico-semantic network development: Perceptual features matter most. *Journal of Experimental Psychology: General, 148,* 763–782. http://dx.doi.org/10.1037/xge0000596

Sailor, K. M. (2013). Is vocabulary growth influenced by the relations among words in a language learner's vocabulary? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1657–1662. http://dx.doi.org/10.1037/a0032993

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech & Hearing Research, 36,* 105–140. http://dx.doi.org/10.1044/jshr.3601.105

Siew, C. S., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive Network Science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity, 2019,* 2108423. http://dx.doi.org/10.1155/2019/2108423

Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour, 2,* 682–692. http://dx.doi.org/10.1038/s41562-018-0422-4

Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism, 15,* 490–502. http://dx.doi.org/10.1017/S1366728911000216

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports, 7,* 46730. http://dx.doi.org/10.1038/srep46730

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41–78. http://dx.doi.org/10.1207/s15516709cog2901_3

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44,* 1321–1337. http://dx.doi.org/10.1044/1092-4388(2001/103)

Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research, 46,* 1312–1323. http://dx.doi.org/10.1044/1092-4388(2003/102)

Storkel, H. L. (2004). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research, 47,* 1194–1212. http://dx.doi.org/10.1044/1092-4388(2004/088)

Storkel, H. L. (2013). A corpus of consonant-vowel-consonant real words and nonwords: Comparison of phonotactic probability, neighborhood density, and consonant age of acquisition. *Behavior Research Methods, 45,* 1159–1167. http://dx.doi.org/10.3758/s13428-012-0309-7

Storkel, H. L., & Adlof, S. M. (2009). Adult and child semantic neighbors of the Kroll and Potter (1984) nonobjects. *Journal of Speech, Language, and Hearing Research, 52,* 289–305. http://dx.doi.org/10.1044/1092-4388(2009/07-0174)

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. http://dx.doi.org/10.1044/1092-4388(2006/085)

Storkel, H. L., & Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26,* 191–211. http://dx.doi.org/10.1080/01690961003787609

Strogatz, S. H. (2001). Exploring complex networks. *Nature, 410,* 268–276. http://dx.doi.org/10.1038/35065725

Tan, S. H., & Schafer, G. (2005). Toddlers' novel word learning: Effects of phonological representation, vocabulary size and parents' ostensive behaviour. *First Language, 25,* 131–155. http://dx.doi.org/10.1177/0142723705050338

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research, 51,* 408–422. http://dx.doi.org/10.1044/1092-4388(2008/030)

Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology, 68,* 1–32. http://dx.doi.org/10.1016/j.cogpsych.2013.10.002

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-term and short-term memory. *Journal of Memory and Language, 67,* 30–44. http://dx.doi.org/10.1016/j.jml.2012.02.008

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 & Computers, 36,* 481–487. http://dx.doi.org/10.3758/BF03195594

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2,* 75–94. http://dx.doi.org/10.1146/annurev-linguistics-030514-124832

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114,* 245–272. http://dx.doi.org/10.1037/0033-295X.114.2.245

Appendix A

Cumulative Probability Distributions of Lexical Characteristics of English and Dutch Words Analyzed in Study 1

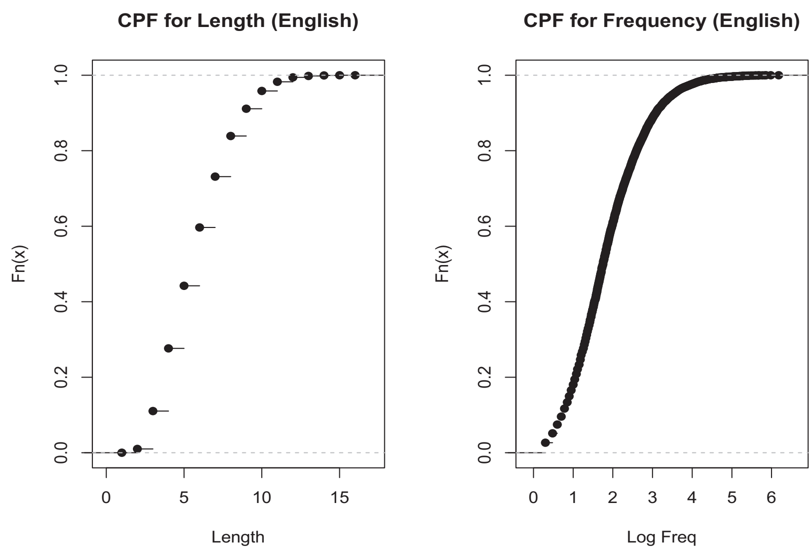

Figure A1. Cumulative probability distributions of length and log frequency for English words analyzed in Study 1. CPF = cumulative probability function.

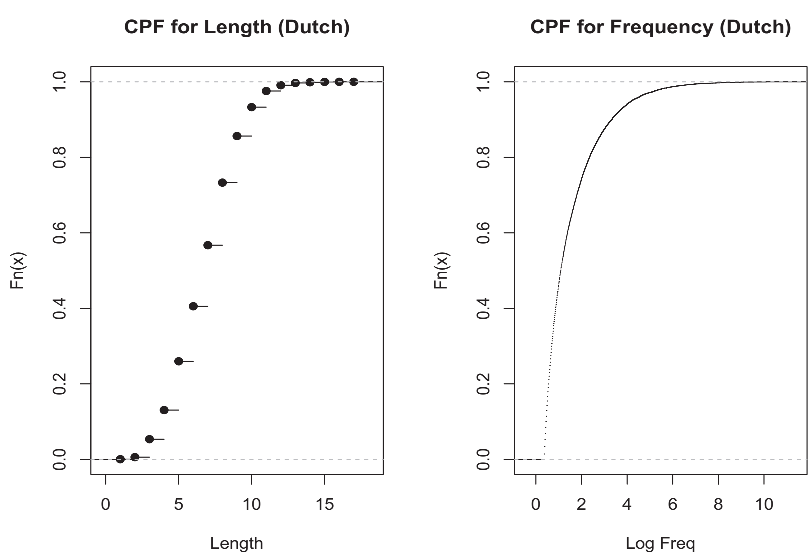

Figure A2. Cumulative probability distributions of length and log frequency for Dutch words analyzed in Study 1. CPF = cumulative probability function.

(*Appendices continue*)

Appendix B

Nonword List for Study 2 (in IPA) With Growth Values and Other Characteristics

IPA	Segment probability	Biphone probability	Word duration (ms)	PATT growth value	PACQ growth value
			PATT nonwords		
			Low PATT nonwords		
tʃoɪŋ	0.0241	0.0001	612	3	2
tʃoɪʃ	0.0201	0.0001	560	3	1
θaʊf	0.0362	0.0001	531	8	0
ðʊm	0.0626	0.0002	576	5	1
jʊŋ	0.0299	0.0016	581	7	1
noɪŋ	0.0390	0.0002	669	4	1
			High PATT nonwords		
ðʊtʃ	0.0211	0.0001	573	9	3
zʊtʃ	0.0208	0.0001	597	9	3
waʊŋ	0.0417	0.0001	633	18	1
waʊb	0.0559	0.0002	627	10	1
jʊtʃ	0.0261	0.0017	522	9	3
zetʃ	0.0398	0.0002	562	11	1
			PACQ nonwords		
			Low PACQ nonwords		
θɑs	0.1461	0.0024	559	8	1
ʤʊtʃ	0.0320	0.0004	599	9	1
zɑz	0.0832	0.0009	585	9	1
θεθ	0.0870	0.0008	633	10	1
ʤoŋ	0.0748	0.0003	536	11	1
dʊŋ	0.0738	0.0004	469	17	1
			High PACQ nonwords		
vɑθ	0.0903	0.0011	602	4	3
vɑf	0.1026	0.0013	642	6	3
vɑv	0.1065	0.0016	649	6	3
mʊb	0.0934	0.0002	511	15	3
mʊz	0.0875	0.0002	563	19	3
mʊg	0.0853	0.0003	617	21	3

Note. IPA = international phonetic alphabet; PATT = preferential attachment growth values; PACQ = preferential acquisition growth values.

Received December 3, 2018
Revision received April 9, 2020
Accepted May 5, 2020 ■