Univerzita Karlova

Filozofická fakulta

Habilitační práce
[Phonological networks and their growth in second language learners]

2021

Eva Maria Luef

Contents

https://docs.google.com/spreadsheets/d/1sQjKNzfjzte5vFvn73CIeqyBucs2UDXR/edit?usp=sharing&ouid=102485829946183460861&rtpof=true&sd=true

List of figures

24. Edge strength (assessed through weighted degree centrality) increases in the giant component as the learner networks grow, while it stays relatively constant in the islands.
25. Average path length and diameter of giant components in the learner and BNC networks.
26. Distributions of network measures in the learner networks.
27. Distributions of network measures in the BNC network.
28. Phonemic distribution in the six proficiency levels of L2 speakers and the BNC data of L1 speakers.
29. Community structure in the giant component of the A1 network.
30. Communities of the C2 giant component.
31. Communities of the BNC network: giant component in the middle, and other larger islands/communities nearby.
32. Random network growth (R package "igraph", function "erdos.renyi.game").
33. Random network growth (R package "igraph", function "barabasi.game").
34. Hub-and-spoke network typology resulting from a close approximation of the "winner-takes-all" mechanism of node growth.
35. According to percolation theory, each node has an equal 0.03 change of being selected for removal. This leads to a higher likelihood of small-degree nodes being selected as they are more numerous.
36. Bose-Einstein condensation in a network with two super-hubs taking all incoming links.
37. The introduction of new words can lead to new neighbor formation in pre-existing words in a phonological network (data from the A1, above, and A2, below, networks of the present study).
38. In the A1 learner network, the new word "hit" has one network community (orange) but another community of individual connection points ("hot", "hat", i.e., the lure-of-the-associates dynamic).
39. Word growth (i.e., addition of new phonological neighbors) according to network parts in the learner networks.
40. Growth distribution across the proficiency networks (heavy tails were cut off). Proficiency levels indicate starting point of growth development, i.e., 'A1 growth' refers to growth occurring between the A1 and A2 level.
41. Standardized node growth in relation to phonemic length and per proficiency level. Kruskal-Wallis test statistics indicated by asterisks (within-proficiency-group comparisons).
42. Standardized node growth and lexical frequency rate of words. Kruskal-Wallis test statistics indicated for each proficiency level.
43. Standardized node growth and phonotactic probability of words. Kruskal-Wallis test statistics indicated for each proficiency level.
44. Standardized node growth and clustering coefficients of words. Kruskal-Wallis test statistics indicated for each proficiency level.
45. Standardized node growth and degree/weighted degree. Kruskal-Wallis test statistics indicated per proficiency level.
46. Node growth and network parts (giant component, islands, singletons).
47. Standardized growth rate and community sizes across the proficiency levels. Kruskal-Wallis test statistics indicated per proficiency level.
48. Probability of preferential node attachment in the different learner networks. PA values were normalized by 100.000.
49. Density general node fitness across network (PC 1).

List of tables

1. Introduction

Lexical knowledge is a crucial pillar of linguistic competence, upon which all other linguistic functions depend. Decades of psycholinguistic research have explored the cognitive representations of words in our minds and their internal organization, the so-called mental lexicon. Of particular interest, here, have been the pathways of word acquisition and retrieval, the functional components of the mental vocabulary, and the interrelations between words encoded in the vocabulary. Research of the mental lexicon is voluminous and has significantly advanced our understanding of how the human mind processes words in its first and other learned languages.

The primary method of investigation to understand the inner workings of the mental lexicon has been the reconstruction of linguistic processes through psycholinguistic tasks, such as word recognition, or the charting of developmental trajectories of word learning by focussing on sub-groups of words and their formal and functional neighborhoods (i.e., semantic or phonological neighborhoods). This "bottom-up" approach to word processing has significantly shaped our current knowledge in the field and built a solid theoretical foundation to capture phenomena observed in experimental tests over the decades. Recent advances in the mathematical domain of network sciences have now offered promising new scientific avenues to investigate a "top-down" approach to the mental lexicon. By studying words as belonging to a large and largely interconnected network, a new understanding of word processing can be gained from a more holistic perspective of word memory and learning. Novel insights into functional and developmental patterns can come from viewing the mental lexicon as a complex system beyond the sum of its parts: performance of one part may depend on that of another to achieve a relevant outcome, creating an interrelated web. This bird's-eye view of the mental lexicon as a complex system can facilitate the study of the grander structure of word connections, from which new patterns of hierarchical relationships and lexical access dynamics can be inferred, leading to more predictive models of the factors influencing lexical processing.

The research field of lexical network science is relatively new and researchers working within this theoretical framework have investigated a small number of languages involving first-language users. Second and foreign languages are underrepresented, with no study charting the whole lexicon of language learners from the network perspective. This book will attempt to fill the knowledge gap by providing an overview of lexical word form networks of learners of English as a second language[1]. For contrast, a lexical word form network of British English first-language users will be presented, in order to be able to draw comparisons between first and second languages. A specific focus of this book will be on mathematical modelling of network-theoretical concepts in relation to their psycholinguistic realities and the question of what network science can contribute to theories of word learning and lexical access. Central insights into the dynamics of word learning at various proficiency levels arise by placing them in the wider context of growth algorithms governing evolving (=growing) networks. The core chapters include computations of network-relevant measures and mathematical descriptions of word form networks, which are intended to provide a first glimpse at second-language-learner networks and serve as reference points for future studies focussing on word memory and learning from the network viewpoint. The goal of this book is a network-theoretical description of lexical knowledge and its growth in learners of English as a second language (ESL). The structure of the book is as follows.

The second chapter begins with an overview of word form relationships in the mental lexicon as defined in common psycholinguistic theories of lexical access. The central notion is 'phonological neighbor', a measure of the relationship between word forms. Much explanatory weight of network science is placed on the quality and quantity of relationships between entities (words) along known similarity dimensions, and phonological neighbors are the logical starting point for a network-theoretical approach to word forms. The chapter will survey various facets of the mathematical concepts of network science and their application to phonological network

[1] The term "second language" will be used throughout this book in the psycholinguistic sense of any language learned after the first (native) language of a speaker, with the exception of bilingualism (see, e.g., Gass & Selinker, 2008). ESL or "English as a second language" and EFL or "English as a foreign language" will thus be used as synonymous terms.

description, with a special focus on the different levels of network organization, including micro, meso, and macro levels of analysis. Characteristics of individual nodes, small clusters of connected nodes (communities), and the overall topology of a network are of systemic relevance for network connectivity and can provide information about lexical processing in phonological networks. Also covered will be theories on activation spreading (diffusion) in lexical networks, since patterns of co-activation are expected to follow network principles and thus differ from predictions based on the traditional models of lexical access.

Chapter 3 presents the findings of the network computations for the phonological networks of learners of English as a second language. Vocabulary data for each of the six CEFR proficiency levels gathered from the *Cambridge Learner Corpus*, with which phonologically neighboring words were calculated in an Oracle database, formed the basis of network construction. In addition to the learner networks, a phonological network of British first-language users (data from the *British National Corpus*) was constructed for comparative purposes. Network-mathematical analyses of the seven networks are presented, including micro-, meso-, and macro-analytical detail. Implications for lexical access are discussed.

In the fourth chapter, network growth algorithms that are of potential significance for phonological networks will be reviewed. A particular focus will be on scale-free networks and how new links can be accumulated consistently with the scale-free assumption of phonological networks. A discussion of various factors impacting the growth of scale-free networks will follow, including, for instance, uniform and preferential attachment processes, fitness models, and aging effects. The chapter ends with a review of previous findings on the growth of phonological networks.

The network growth algorithms are explored in the learner networks in chapter 5, where gains in vocabulary from one CEFR proficiency level to the next will be analyzed from the theoretical perspective of evolving networks. Growth rates across distinct network parts, communities, and individual nodes, as well as the contributions of growth algorithms arising from various evolving network theories, will be investigated. Furthermore, growth development over the course of language learning is charted in micro, meso, and macro levels of the respective networks. Various theoretical extensions of the Barabási-Albert evolving network model (e.g., initial attractiveness) are also tested in the networks.

Chapter 6 replicates a previous phonological network growth study of first-language American English and Dutch, which was conducted by Cynthia Siew and Michael Vitevitch in 2020. Due to computational differences, their findings are not directly comparable to the results presented in chapter 5 of this work. In order to ensure comparability, network growth algorithms of the present data were re-calculated in accordance with the alternative computations as used by Siew and Vitevitch. Regression analysis showed similar results for both studies, strengthening the growth hypothesis suggested by Siew and Vitevitch.

The conclusion summarizes the findings and presents future directions for the application of network sciences to the study of the mental lexicon and word learning. The potential of network theoretical approaches to lexical organization and lexical access in language users has only begun to be explored, with new theories currently being developed that aim to accommodate traditional knowledge about the mental lexicon and new hypotheses derived from network sciences.

A two-part appendix of this book includes network graphs and the phonological neighborhood data used for the present study. The latter is presented in an online format where interested researchers are provided with measures of phonological neighborhood density in English as a second language as a resource for linguistic experimentation.

Hopefully this book will inspire researchers to apply network-mathematical notions to the psycholinguistic study of word relationships in the mental lexicon. This integrated approach across research specialties can be intellectually fruitful and lead to a deeper understanding of the cognitive underpinnings of linguistic representations in the human mind. Embracing different theoretical approaches and innovative research questions will get us further to the goal of elucidating the nature of words in the human mind.

2. The mental lexicon as a phonological network

2.1. Traditional views of lexical connectivity

The mental lexicon is the human repository of lexical knowledge (Oldfield, 1966). It is the cognitive system that organizes lexical activity and forms the basis of linguistic expression by providing storage to all vocabulary items that are known by a language user (Dóczi, 2020). A lexical representation is suggested to contain information about its form, meaning, and syntactic properties that is accessible upon lexical access of a given word (Yelland, 1994). How words are represented and processed in the mental lexicon is not only crucial for theories on language acquisition and development but can more generally shed light upon universal principles by which humans mentally categorize language. The mental lexicon is best conceptualized as an ideal, abstract notion, rather than a concrete list of word knowledge (Aitchison, 2012; He & Deng, 2015). Generally, it can be seen as a dynamical memory system supporting linguistic processing that is continuously adapting in the face of experience.

Theories on lexical access in the mental lexicon distinguish between word production and word perception. In production, a meaning is provided (i.e., the semantic concept the speaker wants to convey) and the appropriate phonological representation has to be identified. In perception, the input is the phonological representation and the listener has to match it with existing words in the lexicon to retrieve the meaning. This dichotomous search system poses problems for theoretical accounts of word storage, as optimization of word accessibility in production and perception are seemingly incompatible (Levelt, 2001). While some researchers have postulated two separate mental lexica for production and perception processes, a single lexicon with duplicate sound-meaning pairings is also plausible (see Aitchison, 2012; Fay & Cutler, 1977, for a discussion). Alternatively, the mental lexicon could either be organized according to optimal production or to optimal perception, thereby biasing word storage toward a particular type of speech process.

Virtually all psycholinguistic accounts of lexical processing propose that lexical information is primarily organized along similarity principles, both semantic and word form similarity (see Buchwald, 2011, for a review). The former refers to words sharing semantic features, while the latter refers to words sharing phonological forms. Even though the two processes work in tandem, they are conceptually separate from one another and associated with different neural processing pathways (Ullman, 2007). Studies have consistently demonstrated an advantage of phonological similarity in word learning, demonstrating the strong influence of phonological word form on lexical processing in first language acquisition and foreign language learning (e.g., Aitchison, 2012; Arutiunian & Lopukhina, 2020; Beckage & Colunga, 2019; Dell & Gordon, 2003; Fourtassi, Bian, & Frank, 2020; Gahl, Yao, & Johnson, 2012; Harley & Bown, 1998; Havas et al., 2018; James & Burke, 2000; Siew & Vitevitch, 2016; Vitevitch & Luce, 2016).

2.1.1. Phonological neighbors

The similarity bias in the phonological domain is governed by 'phonological neighbors', a well-studied notion of lexical relationships (Goldrick, Folk, & Rapp, 2010; Landauer & Streeter, 1973). In their seminal study of lexical frequency, Streeter and Landauer (1973) defined phonological (in their term "lexical") similarity as the distance of one piece of information (a phoneme or grapheme) between two words. What was referred to as "neighbors" and "similarity neighborhoods" have evolved into today's concepts of "phonological neighbors" and "phonological neighborhoods" (see, e.g., Vitevitch & Luce, 2016). Phonological neighbors are commonly defined as sharing the majority of phonological segments and differing by one phoneme substitution, deletion, or addition (the so-called Hamming or Levenshtein distance, Landauer & Streeter, 1973; Luce & Pisoni, 1998). What lies at the core of lexical activation is competition for activation between segments and thus phonological neighbors. When a target word becomes activated (through production or perception), other words that share phonemes with the target word also become co-activated. Since co-activation spreads through common phonemes, the more phonemes are shared within a neighborhood, the more activation spreads within the neighborhood. This is generally referred to as the "phonological neighborhood effect" (Vitevitch & Luce, 2016). Differences arise in relation to speech production and perception processes: in perception, competition among lexical candidates leads to slower access of the target word (Luce & Pisoni, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Vitevitch, 2002a; Vitevitch & Rodriguez, 2004). In speech production the opposite is observable, and words from dense neighborhoods are produced faster and more accurately (Vitevitch, 2002b; Vitevitch, Armbruster, & Chu, 2004; Vitevitch & Sommers, 2003). As mentioned, the crucial distinction here is that in speech recognition, a listener has little accompanying semantic information available and thus has to guess a phonological word form from phonemic input alone. In production, speakers have semantic information at their disposal, which they can use to prevent co-activation of phonological neighbors. It has been suggested that in speech production, articulation-relevant features of phonological representations become strengthened through co-activation (Vitevitch, 2002b). In general, phonological neighborhood effects are a well-documented phenomenon in psycholinguistic research, can be observed in different languages and populations (e.g., Arutiunian & Lopukhina, 2020; Gordon, 2002; Marian & Blumenfeld, 2006; Stamer & Vitevitch, 2012), and constitute a central component of lexical processing (see Vitevitch & Luce, 2016, for an overview of neighborhood effects in perception and production).

An important question arising from phonological neighborhoods in the mental lexicon concerns word learning, or the addition of new words to the mental lexicon. Words which find many potential neighbors or 'anchor words' in the vocabulary of a learner are more easily and rapidly integrated (Gaskell & Dumay, 2003; Storkel, Armbruster, & Hogan, 2006). Thus, dense phonological neighborhoods with numerous words connected via the one-segment distance can exert a pull-effect on new words that share phonological features with these neighborhoods. In fact, there are indications that clusters of phonologically related words constitute particularly strong attraction points for new words (Stamer & Vitevitch, 2012; Storkel et al., 2006). While adult word learning seems to be facilitated by high-density neighborhoods, it is impaired when target words reside in sparse neighborhoods (Storkel et al., 2006). A phonological homogeneity bias is evident here that essentially skews the learner's perceptions and word memory in a way that favors larger clusters of phonologically similar words. While the majority of word learning research has focussed on first languages (or 'L1'), phonologically guided word learning has also been demonstrated in second language ('L2') learning (e.g., Bialystok, 2010;

Kaushanskaya, Yoo, & Van Hecke, 2013; Leach & Samuel, 2007; Smits, Sandra, Martensen, & Dijkstra, 2009; Stamer & Vitevitch, 2012; Yates, 2013), and new L2 words that are phonologically similar to many words that are already in a learner's L2 vocabulary will be learned faster and more accurately (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010). Vice versa, new words entering sparse phonological neighborhoods in a learner's vocabulary are more difficult to acquire and retain. The phonological similarity bias in learning is not restricted to one language but operates across different languages, as evidenced by parallel activation of L1 phonological forms when L2 is being processed (Broersma & Cutler, 2008). Phonologically similar L1 and L2 words are frequently co-activated, even when their activation is irrelevant to the task (Carrasco-Ortiz, Midgley, & Frenck-Mestre, 2012; Schulpen, Dijkstra, Schriefers, & Hasper, 2003). Third language ('L3') studies have yielded similar results, too. L3 words tend to activate words from the first and second languages of speakers (Van Hell & Dijkstra, 2002), which suggests that similarity in word forms across languages can provide benefits for lexical processing in L3 learners (Mulík, Carrasco-Ortiz, & Amengual, 2018). While co-activation between L3 and L1 lexemes is commonly observed, that between L3 and L2 seems to be more dependent on the proficiency level of the learner (Mulík et al., 2018), and this has implications for phonological pull-effects during word learning: higher L2 proficiency leads to increased L3 neighborhood effects. The typological and phonological relationship of L1, L2, and L3 languages certainly plays a role here. In addition, cross-language homophones (Carrasco-Ortiz et al., 2012; Haigh & Jared, 2007) and cognates (Dijkstra et al., 2010; Dijkstra, Timmermans, & Schriefers, 2000; Midgley, Holcomb, & Grainger, 2011; Van Heuven, Dijkstra, & Grainger, 1998) need to be considered as they may also be affected by phonological neighborhood effects.

2.1.1.1. Quantifying phonological neighbors

The one-segment phonological distance has proven to be a useful concept for psycholinguistics over the last few decades. However, more detailed quantifications of phonological neighbors have been introduced. One line of research focusses on locus-oriented notions of phonological neighborhoods that consider the serial order of phonemes in words, essentially proposing that not all phonemes are equal when it comes to phonological neighborhoods (e.g., Desroches, Newman, & Joanisse, 2009; Simmons & Magnuson, 2018). Typically, word-initial phonemes are attributed a higher conceptual importance in the sense that stronger neighborhood connections exist between words that share onsets (so-called "cohort effects), such as *cat-cab*; rhyme neighbors differing in the onset phoneme (for instance *cat-hat*) show weaker competition effects (Simmons & Magnuson, 2018). In addition to locus of phoneme substitution/deletion/addition, another fact that has been suggested to be important in phonological neighborhood measures is the so-called P-metric that counts the phonemic possibilities that exist for a word to form neighbors (Vitevitch, 2007). As exemplified by Vitevitch (2007), the English word *mop* has three phoneme positions where neighbors can form (P=3), e.g., *hop, map, mock*. Its phonological neighbor word *mob*, however, has only two phoneme positions for neighborhood formation (P=2), e.g., *rob, mock*. Contrasting findings regarding the phonological neighborhood spread have been presented by the literature. Yates (2009) found that words where numerous phoneme positions can be changed to create neighbors were responded to faster, as opposed to words where only a limited number of phoneme positions are changeable. Vitevitch (2007), however, reported the opposite: words with smaller P-values were recognized faster, supporting activation-competition theories. Fewer neighbor formation possibilities mean less cognitive effort involved in processing words. Higher degrees of certainty in word recognition (in the case of a smaller P) correlates with faster recognition rates, whereas more uncertainty due to a higher rate of variation probability (in the case of a larger P) slows down processing. Even though these findings are not in agreement, they nonetheless show that the probability of phoneme overlap with neighbors (across different phonemic positions) has an effect on lexical access and potentially the structural organization of word forms in the mental lexicon.

Although the Levenshtein distance is the dominant metric of phonological similarity, different ways of quantifying phonological similarity have been suggested, such as the 75% similarity grouping (Kapatsinski, 2006). Here, the overall number of segments per word factors into neighborhood similarity and phonological neighbors may differ by up to 25% from the target word. Perceived phonological similarity is tightly linked with the notion of phonological confusability and the question of how individual language users rate similarity of phonemes. Phonological confusability may encompass a wider range of word similarity relationships apart from the one-segment neighborhood, including the 75% metric or PLD20, which gives the mean number of steps that are required to transform a word into its 20 closest neighbors (Suarez, Tan, Yap, & Goh, 2011). Suarez and colleagues show that co-activation extends to the wider neighborhood separated by more than one phoneme distance (also see Chan & Vitevitch, 2009, for similar findings), even when no one-phoneme neighbors exist. What they proved was “neighborhood effects without neighbors” (Suarez et al., 2011: p. 605).

Phonological similarity can also be described across different phonetic dimensions, with feature-based analyses describing the degree of closeness of phonological neighbors (Bailey & Hahn, 2001). Voiced and unvoiced variants of a consonant are arguably much closer in phonological distance than a vowel and a consonant, as in *bat-pat* vs. *ball-boy*. Fricke, Baese-Berk, and Goldrick (2016) point out that the English word *cod* has a multitude of neighbors (27 overall) with which it shares different phonological features: *God* differs only in the voicing parameter of the word-initial plosive, while the word-initial sibilant in *sod* represents a larger phonetic distance to the target word *cod*. The authors showed that position-specific similarity of segments can predict activation spreading in a phonological neighborhood (in their case, in word-initial position). These findings demonstrate that further quantification of phonological neighbors can inform lexical processes.

Phonology-based models of lexical access tend to view the mental lexicon as "a collection of arbitrarily ordered phonological representations and the process of lexical retrieval as a special instance of pattern matching" (Chan & Vitevitch, 2009: p. 1934). The majority of current models of spoken word recognition share the assumption that phonological overlap is the central force driving competition and activation in lexical processing (Weber & Scharenborg, 2012). Phonemic input activates all similar phonemes within a phonological neighborhood and words containing those shared phonemes compete for activation. The way phonological connections between words can further or impede activation spreading is a crucial question in theoretical models of lexical access.

2.1.1.1.1. Spoken word recognition

Models of spoken word recognition rely on various notions of phonological neighbors. One of the earliest models, the cohort model of lexical access, focusses on word-initial segments (Marslen-Wilson, 1987; Marslen-Wilson & Warren, 1994). The model predicts co-activation based on temporal phonemic overlap starting at the initial phoneme and proceeding with each successive, similar phoneme in a serial manner as speech unfolds in time. Here, phonological neighbors are those words sharing onset phonemes, and non-onset phonological neighbors are excluded as candidate words early on in chronological phonemic perception. This model is based on the assumption that the human brain, and thus the mental lexicon, operates according to the principle of greatest efficiency (Marslen-Wilson & Welsh, 1978). Initially, the range of lexical candidates is large (with activation of the first few phonemes) but it rapidly declines and produces mismatches which are then excluded from the candidate list. Word-initial phonological cohorts are proposed to be co-activated and word choice is narrowed down with each new perceived phoneme. Upon hearing /s/, the whole cohort of s-initial words is activated and then gradually, with each further phoneme, different words are dismissed until the target word is retrieved. Words will be recognized once they have reached a unique identifying phoneme, for example the English phonemic string /fɛb/ clearly identifies the only English word that begins with it, *February* (Weber & Scharenborg, 2012). After this initial access process, the integration stage checks for syntactic and semantic suitability of a word, removing contextual mismatches from the cohort. The special status of the word-initial phonological portion for lexical processing has been confirmed by numerous studies (e.g., Friedrich, Felder, Lahiri, & Eulitz, 2013; Treiman & Danis, 1988; Vitevitch, 2002a). High onset density or a high number of phonological neighbors sharing the same onset phoneme generally slows down lexical processing, as a large number of competitor words become activated (Vitevitch, 2002a). No such effect has been determined for rhyme neighbors, underscoring the phoneme-wise left-to-right activation in speech recognition (Sevald & Dell, 1994; Vitevitch, 2002a). Competition remains high at word onsets but eases up as more phonemes are added to the word selection process (Chen & Mirman, 2014). In the cohort model, competition is restricted to phonemic access but it is not explicitly postulated that words in a cohort compete with one another. According to Marslen-Wilson (1987, p. 84), “the timing of word-recognition processes is not affected by the number of alternatives that need to be considered”, thus the speed and accuracy of word retrieval is not influenced by the number of competitors.

The successor model Cohort II was devised to account for lexical frequency of words (high-frequency words are recognized faster than low-frequency ones) and to consider phonological confusability, for instance *nobility* being activated by *mobility* (Marslen-Wilson, 1990; Marslen-Wilson, Brown, & Tyler, 1988). The cohort model has been challenged by findings that (English) listeners can rarely uniquely identify a word before offset (Bard, Shillcock, & Altmann, 1988; Luce, 1986). As a consequence, the notion of onset matching as the single most important mechanism of lexical access has been questioned (Weber & Scharenborg, 2012).

The cohort model defines the structure of the mental lexicon as one-segment distance neighbors in a right-unfolding way (see Figure 1). The structural organization predicted by the cohort theory would resemble a computerized feed-forward string matching, however it is not clear how different strings (=words) would be related to one another. Since cohorts are formed depending on the initial phoneme, each word-initial phoneme in a language would constitute the first layer of cohort. Word-initial biphones, triphones, and so forth, all constitute their own cohorts. This means that each word is a member of different cohorts (e.g., *bean* belongs to the /b/-cohort, the /bi/-cohort, and the /bin/-cohort), and the probability of belonging to different cohorts increases with phonemic length.

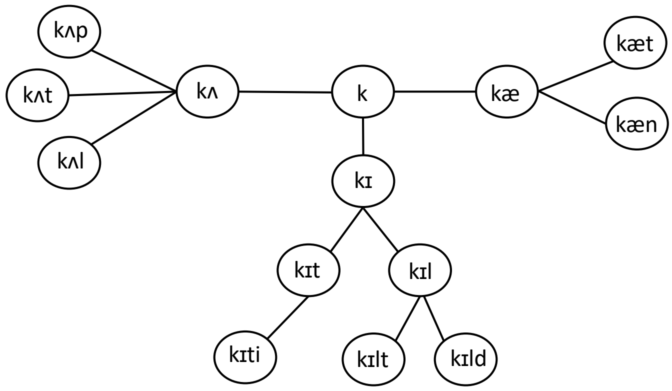


Figure 1: A network graph of word relationships as predicted by the cohort model. The structural organization of the mental lexicon links words together via temporally ordered segments, creating clusters of serially matched phonemic strings. Activation is shared in groups of overlapping phonemes.

While the special role of word-initial phonological segments in speech processing is accepted by the majority of literature, other models of word recognition acknowledge a contribution of non-initial phonological segments to neighborhood formation. A famous activation-competition model of spoken word recognition, the Neighborhood Activation Model (or NAM, Luce & Pisoni, 1998) and its derived connectionist instantiation PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000), posit that phonological co-activation spreads within a group of words that overlap in the majority but differ in a minimal number (N=1) of phonological segments. In the original NAM model, neighborhoods may be established through any segmental position in a word, and phonemic differences in neighborhood formation do not impact the strength of a neighborhood. An addendum to the model, however, recognizes different variations of acoustic-phonetic distances between phonological neighbors, and thus neighborhood strength. As Goldinger, Luce and Pisoni (1989) showed, phonetically close neighbors have an amplified effect on phonemic competition by inhibiting word recognition (see Gahl et al., 2012; Scarborough, 2013; Suarez et al., 2011, for similar findings). For instance, *cap* and *cab* are more influential neighbors and share more activation (and competition), as opposed to *cab* and *fab*.

In general, NAM assumes that competition arises between the co-activated lexical candidates from which the best-fitting word is then chosen for final selection. Since lexical selection is a competitive process, words with strengthened activation, resulting for instance from high frequency rates, facilitate word recognition (Frisch, 2011). The neighborhood probability equation follows

$$p=(\text{target} * \text{frequency}_t) / (\text{target} * \text{frequency}_t)+(\sum(\text{neighbors}_j * \text{frequency}_{Nj})$$

Here, the activation level of the target word *t*, the sum of neighbor word probabilities (=neighbors$_j$, i.e., the overall level of activity in the lexical neighborhood), and lexical frequency information are taken into account (Chan & Vitevitch, 2009; Luce & Pisoni, 1998). Low-density neighborhoods (i.e., words that have few neighbors) experience less competition and thus faster target word recognition rates, which leads to those words being responded to and recognized more quickly as opposed to words with a high number of neighbors. Numerous studies have confirmed the NAM predictions for word recognition, earning it its prominent place in spoken word recognition (e.g., Goh, Suarez, Yap, & Tan, 2009; Luce et al., 2000; Vitevitch, 2002c; Ziegler, Muneaux, & Grainger, 2003). Figure 2 schematizes neighborhood connectivity as outlined by NAM.

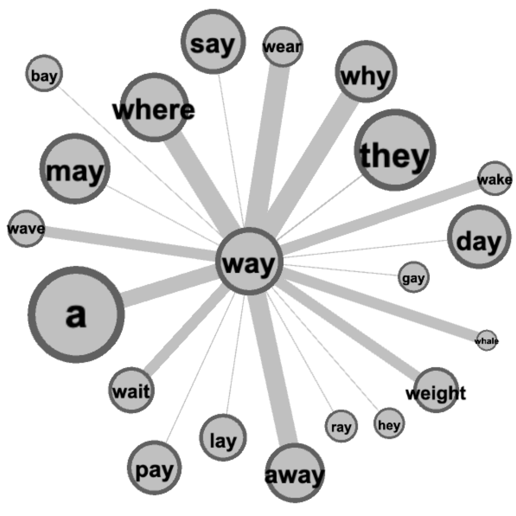


Figure 2: Phonological neighborhood structure according to NAM; node size according to lexical frequency rate (larger nodes are more frequent words), links according to phonological distance (thicker links indicate closer distance, Downey, Sun, & Norquest, 2017). Graph constructed with Gephi using B2 learner data of the present study (see chapter 3 for details).

Luce and Pisoni (1998: p. 1) explicitly acknowledge a "structural organization of the lexicon" based on "similarity relations among the sound patterns of spoken words" but do not provide hints as to what (if any) larger structural design there might be beyond linking words in one-segment neighborhoods. It is a logical assumption that each of the neighbors of *way* in Figure 2 has its own neighborhood, with potential implications for further spreading of co-activation outside of the immediate neighborhood. Ultimately, a large number of words in a lexicon could be interlinked in one large web, however, NAM does not address this issue nor does that factor into the NAM account of activation spreading. The immediate, one-phoneme-distance neighborhood of a target word constitutes the central tenet of activation spreading in the model.

Recognizing the crucial role of phonetic features in phonological neighborhoods, the TRACE model of speech recognition was developed, which posits that competitor words are activated that match any part of the speech input not by phonological segments but by features (McClelland & Elman, 1986). TRACE is a connectionist interactive-activation model with three layers: feature, phoneme, and word. Multidimensional features of phonemes serve as the input (e.g., frication, nasality, back vowel, front vowel), which are then channelled up to the next layer, the phoneme layer (see Strauss, Harris, & Magnuson, 2007, for an overview). Phonemes best matching the features level (e.g., frication: /s/, /f/) will be selected and activation will then be spread to the next layer, where the word will be selected. Words with higher activation levels then inhibit activation in other, non-matching words. Due to its focus on phonetic features rather than phonemes, TRACE can account for underspecification, phonological variation (e.g., dialects), and mispronunciation of target words (Mayor & Plunkett, 2009). Special weighting is assigned to higher frequency units in the model (Hannagan, Magnuson, & Grainger, 2013). TRACE predicts that phonemic activation becomes stronger over time and as a word receives more activation it will activate its phonemes more strongly. Differences in the time course of phoneme activation predict that activation of words overlapping in word-initial segments (i.e., onset neighbors) will occur earlier (Weber & Scharenborg, 2012), mirroring a crucial dynamic proposed by the cohort model.

The structural organization of the mental lexicon that is assumed by TRACE is phonological feature sharing based on phonological detail. Phonological neighborhoods are established according to phonological features (e.g., nasality, frontness in vowels), link together words that are defined by a distance of more than one phonological segment, and may vary with acoustic-phonetic detail of a particular speaker (for instance, idiosyncratic phonological detail). Figure 3 schematizes the phonological structure underlying the TRACE model.

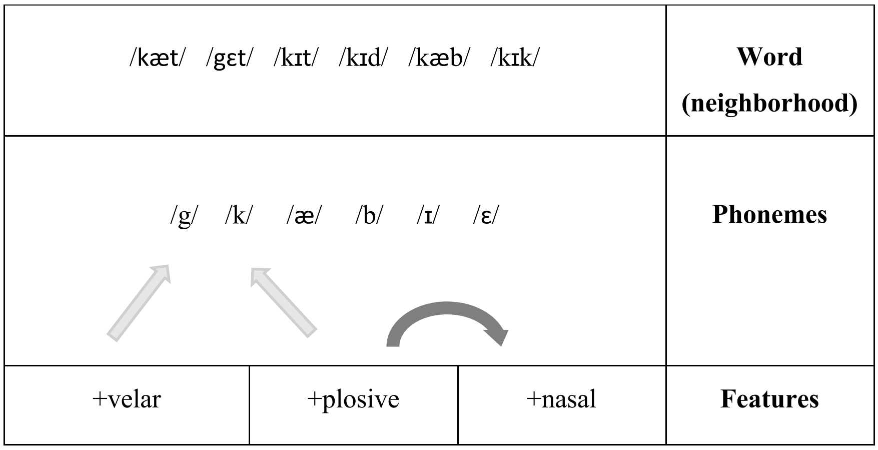


Figure 3: The three levels of processing elements in the TRACE model. Spreading activation indicated in light grey, inhibition indicated in dark grey.

Phonological neighbors are defined differently in this model as compared to NAM and the cohort model, with the sharing of phonetic features rather than phonemes being the criterion. Thus, word pairs such as *cat-bad* become co-activated as neighboring lexical competitors (Hannagan et al., 2013). The implication for the organization of the mental lexicon is that phonological neighborhoods can potentially have a wider dimension, encompassing a greater range of sounds that share features.

These classical models of spoken word recognition make different predictions about the spread of co-activation in the mental lexicon. Connectionist models, such as TRACE, propose inhibitory mechanisms between lexical representations. Due to the bottom-up phonemic activation, candidate words are able to inhibit each other depending on similarity of the input. Once the bottom-up activation coupled with the lateral inhibition focusses on a target word, it becomes the last activated candidate. According to NAM and the cohort model, activated candidate words do not interact at the lexical level but these models propose decision rules that determine which lexical entry received the most activation relative to the other activated candidate words. These models only make predictions about immediate phonological neighbors and how they influence and compete with one another. TRACE suggests a wider net of neighboring words going beyond the one-phonological-segment distance but does not make any predictions concerning activation spreading in neighbors of neighboring words. The larger organizational design of all words contained in a lexicon and their relationships to one another via extended neighborhoods remains unaddressed by these models.

2.1.1.1.2. Spoken word production

A salient feature of phonological neighborhood effects is that they fulfil dual functions. In spoken word recognition, neighbors inhibit lexical processing. Spoken word production, on the other hand, is facilitated by neighbors (Chen & Mirman, 2012; Dell & Gordon, 2003). Explanations for this discrepancy can be found in models of speech processing, in particular interactive models where lexical and phonological levels of word recognition provide feedback to one another. This interaction is captured by Dell's interactive two-step model of lexical access and retrieval (Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997) where lexical and phonological retrieval are distinct and ordered categories but interact through bi-directional spreading of activation. This means that semantic information (semantic features) can influence phonological retrieval and phonological information can affect lexical retrieval (Dell, Martin, & Schwartz, 2007). In word production, the first step is lexical selection which maps the conceptual representation of a word to a lexical representation (often referred to as 'lemma', Foygel & Dell, 2000). Phonological information is not required at this point (Levelt, Roelofs, & Meyer, 1999). Next, phonological encoding is initiated and the phonemes used for building the target word are retrieved. Phonological encoding describes the process of constructing the phonological form of a target token before articulatory gestures can be prepared in spoken word production (Caramazza, Costa, Miozzo, & Bi, 2001; Dell, 1986; Levelt et al., 1999). Phonological components (phonemes) of target words are sequentially activated after speakers have selected words (Oppermann, Jescheniak, & Schriefers, 2010), mostly independently of the whole word representation (O'Séaghdha & Frazer, 2014; Roelofs, 2006). This phonological pre-activation (also referred to as phonological preparation, see Li, Wang, & Davis, 2017) is initiated at the start of the target word, and thus the onset phoneme(s) are attributed a crucial role in the process (Meyer, 1991; O'Séaghdha, Chen, & Chen, 2010). In word production, the initial semantic activation provides a baseline activation, which is then further boosted by activation of phonological neighbors. Word recognition, on the other hand, begins with the activation of phonological segments and boosts activation of all phonological neighbors, including the target word, and thus activation spreads more evenly within the phonological neighborhood and is less focussed on the target word. Competition for activation is greater in word recognition than in production, as suggested by the literature. The two-step model postulates separate modules for lexical and phonological detail that are linked via interactive routes of information flow. Figure 4 shows the interactive-feedback model as proposed by Dell (1986).

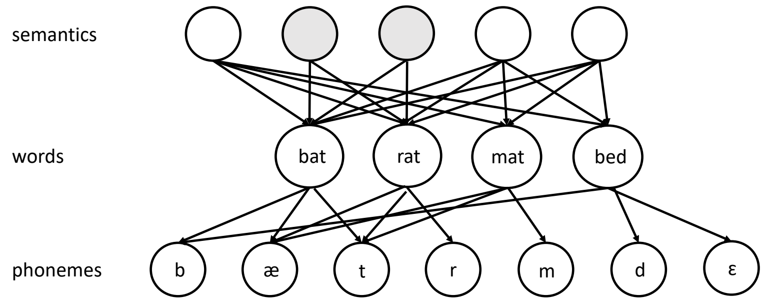


Figure 4: Interactive model of speech production.

One unresolved question in speech production models concerns the flow of information from the semantic to the phonological domain and whether it can be characterized as discrete, cascading, or fully interactive (Schriefers & Vigliocco, 2015). This has direct implications for neighborhood activation. According to discrete serial models of lexical access, the target lemma and a set of semantically related other lemmas are initially activated. After exclusion of the non-target lemmas, phonological encoding of the target is initiated, and non-targets are not phonologically encoded (Levelt, 1999). In cascading models (e.g., Peterson & Savoy, 1998), the activated set of initial lemmas send some activation to phonological encoding before the final target lemma has been selected, thereby spreading phonological activation among competing lemmas. Interactive models (e.g., Dell, 1986) assume feedback spreading between the phonological and the lemma level, a process through which activation will be spread among competitor lemmas at the lemma stage, in addition to the phonological forms sending activation back to the lemmas and thus spreading co-activation among phonologically similar forms. There is some evidence that semantic competitors receive co-activation (as predicted by cascading and interactive models), for instance phonological activation has been shown to spread between near-synonyms like ‘couch’ and ‘sofa’ (Jescheniak & Schriefers, 1998; Peterson & Savoy, 1998) but others have pointed out that this may just be a special case of activation spreading under certain circumstances (Levelt, 1999; Schriefers & Vigliocco, 2015). The assumption of feedback from the phonological to the lemma level has been supported by the ‘lexical bias effect’, which refers to the fact that phonemic errors tend to lead to existing rather than non-words (Nooteboom, 2005). Feedback spreading from the level of the phonological segments to the higher lemma level can explain this phenomenon, while discrete serial models would predict independence of phonological errors from an existing word.

2.1.2. The "cauldron of lexical soup"

The notion of phonological neighborhood is essential for any theory of speech production or perception, as the mental lexicon is known to be structured according to phonological similarity. While quantification and qualification of phonological neighbors may differ in classic psycholinguistic models, one-phoneme difference measurements with phonetic distance considerations are the predominant notions for which theories on lexical access allow predictions. Words constituting so-called minimal pairs that differ in one phoneme receive enough co-activation to arise as lexical competitors for target words. Further, phonetic closeness contributes as well and minimal pairs differing by fewer phonetic features share more co-activation. This logic concerning phonological neighborhood activation spreading is commonly applied in investigations of words and their immediate neighborhoods (such as in experimentation involving particular words) but is largely ignored in the larger context of how all words in a lexicon relate to one another. Due to their processing-oriented nature, the traditional lexical access models take into account only the nearest neighborhood of a word but disregard the larger structure linking different neighborhoods together, which may also have an effect on activation spreading, as evidenced by neighborhood effects without neighbors (Suarez et al., 2011).

Although lexical access models of speech production and perception differ in terms of their basic architecture, they are united by the prediction that phonological pattern matching is the guiding principle. As Chan and Vitevitch (2009) put it, the mental lexicon is presumed to resemble a "cauldron of lexical soup" (p. 1944) where similar phonological forms compete for selection. This means that the lexical models are able to account for neighborhood density effects through virtue of phonological matching without explicitly investigating (or acknowledging) any non-arbitrary phonological structure underlying the collection of all words present in the mental lexicon. Psycholinguistic models have focussed on acquisition, recognition, and production of words, but the issue of how language-related information is represented in the mental lexicon, including the relationship that exists between the stored words, has received little research attention. Word memory itself is an issue separate from (albeit intertwined with) the processing-centric bias in contemporary cognitive psychology (Vitevitch, 2021).

2.1.3. Complex systems approach to the lexicon

Multi-agent systems, consisting of numerous constituent parts, may give rise to collective behaviors at the holistic level of the entire system, which can be unpredictable from solely the individual constituent interactions (Siegenfeld & Bar-Yam, 2020). Such systems can be defined as complex systems (Boccara, 2010), and in contrast to reductionist modelling where the focus is on a part of a system or a small-sized model of it, the complex systems approach takes into account the dependencies, relationships, competitions, and other forms of interactions between local constituents, to which the emergence of a global behavior of a system is linked. A complex system is thus more than "the sum of its parts" (Kane & Higham, 2015), as new behaviors appear when individual parts come together to form larger structures. Similarities to Chaos Theory exist (Rickles, Hawe, & Shiell, 2007), and complex (linguistic) systems may develop ways to self-organize, resulting in emergent principles on the macroscopic level (Hohenberger & Peltzer-Karpf, 2009).

The mental lexicon has been described as a complex system (Siew, 2013; Vitevitch, 2008; Vitevitch, Chan, & Roodenrys, 2012). While the majority of studies of the mental lexicon has focussed on specific and localized lexical neighborhoods within a given lexicon (see section 2.1.1.), Aitchison (2012: ch. 9) acknowledged that a larger structure of word relations beyond the immediate neighborhood must exist and refers to the mental lexicon in terms of a "gigantic multidimensional cobweb". However, no further mentioning of the functional principles or structural organization underlying this gigantic web is made.

An effort to understand a complex system entails the theoretical ability to model it, and network science has become a popular methodological approach applied to the description and analysis of complex systems (Turnbull et al., 2018). Applying the tools of networks science (or graph theory) to phonological connections between words in the mental lexicon permit the modelling of the organizational structure of lexical entries in the mental lexicon from a holistic, "top-down" approach. This line of research has only just emerged but it has already been shown that phonological forms in the mental lexicon are organized in a specific, predictable way and that the structure of the organization – the network design – plays a role in lexical processing (e.g., Benham, Goffman, & Schweickert, 2018; Chan & Vitevitch, 2009; Lara-Martinez, Quintana-

Obregon, Reyes-Manzano, Lopez-Rodriguez, & Guzman-Vargas, 2021; Levy et al., 2021; Neergaard, Luo, & Huang, 2019; Shoemark, Goldwater, Kirby, & Sarkar, 2016; Siew & Vitevitch, 2016, 2020a; Turnbull, 2021; Turnbull & Peperkamp, 2017; Vitevitch, 2008, 2021). Through network science, connectivity of words in a given lexicon can be modelled from a global perspective, patterns of interactions within and between lexical sub-groups can be correlated with emerging behaviors at the level of the lexicon, and fundamental principles underlying the organization (and constant re-organization through learning) of the system and its growth can be studied.

2.2. Phonological networks

Networks science (or graph theory) has been used in cognitive psychology for decades (Kauffman, 1993; van Hemmen & Schulten, 1995), and its utility for the study of linguistic processes in the human mind was recognized early on (Estes, 1975; Feather, 1971; for a review see Siew, Wulff, Beckage, & Kenett, 2019). With network science, the structure of dyadic relationships between entities can be modelled and the influence of the structure on cognitive processes (for instance, word retrieval) can be mathematically quantified (Castro & Siew, 2020; Cong & Haitao, 2014). This is similar to connectionism, which investigates processing of information in human cognition and behavior from the viewpoint of large networks of interactive units (see Joanisse & McClelland, 2015). While connectionism models processes, such as memory categorization or pattern recognition, network theory models the relationships between cognitive units, such as words. The two approaches are largely distinct.

In the network sciences, the main network-theoretical concepts are *nodes* (or vertices) and *edges* (or ties) and their relationship to one another is the essence of network connectivity. A basic assumption about networks is that the transmission of information can only occur between connected nodes. To which and to how many neighbors a node passes its information (or a portion of it) is restricted by the number of neighbors and the strength of the relationships between nodes (Wang et al., 2011). Describing the mental lexicon as a network of phonological word forms requires an understanding of how information (i.e., lexical activation) flows within networks and what role the structural properties of the different levels of network analyses (micro, meso, macro) play.

Studies of word relationships have long relied on graph theoretical architecture, even without explicitly acknowledging it. For instance, Quillian (1967) and Collins and Loftus (1975) constructed semantic-based representations of the mental lexicon that schematized words as nodes and the relationships between them as edges (see Figure 5). Collins and Loftus conceptualized word relationships as two-fold: semantic and phonological (referred to as “lexical” by the authors). Phonological word relationships were presumed to be organized along phonemic similarity, with phonemic properties specified according to their position in a word. Closeness of word relationships was expressed via edge length, with shorter length distances between two nodes indicating a closer semantic or phonological relationship.

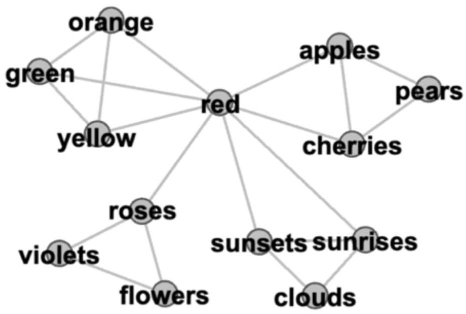


Figure 5: Semantic network, adapted from Collins and Loftus (1975).

Starting in the late 2000s, network-theoretical approaches were beginning to be applied more widely in lexical research, and networks of phonological word forms slowly emerged (Gruenenfelder & Pisoni, 2009; Stella & Brede, 2015; Vitevitch, 2008). In such networks, nodes are words and in case of a phonological relationship between two words, an edge is placed between them. Most commonly, the relationship measure is the one-segment Levenshtein distance (see section 2.1.1.) and words are linked if they differ by one phoneme.

Research on semantic and phonological networks has demonstrated that network-theoretical models of the mental pathways can predict word acquisition (Beckage & Colunga, 2016; Beckage, Smith, & Hills, 2011; Carlson, Sonderegger, & Bane, 2014; Hills, Maouene, Maouene, Sheya, & Smith, 2009b; Vitevitch & Castro, 2015), storage (De Deyne, Kenett, Anaki, Faust, & Navarro, 2017; Storkel, 2002; Vitevitch, 2008), and retrieval (De Deyne et al., 2017; iCancho & Solé, 2001; Vitevitch & Castro, 2015). Beckage, Aguilar, and Colunga (2015) modeled lexical acquisition based on network theory and compared it to a different growth model based on an equal probability for each word to be learned (‘bag-of-words model’, see

Beckage et al., 2011), and they demonstrated superiority of the network growth model for word form (phonological) learning. The network science approach in studies on word learning has gained popularity recently, and the last few years have seen a rise of network-theoretical studies investigating phonological relationships in the mental lexicon. What follows now is an overview of the most crucial network measures, their application to phonological networks, and their meaning for psycholinguistic reality. This will be followed by mathematical modelling of those measures in phonological networks of English-as-a-second-language ('ESL') learners in chapter 3.

2.2.1. Network connectivity

Network measures can relate to different slices (or aspects) of a network. In macro analysis, the focus is on the whole network and its topology. Meso analysis investigates sub-groups of nodes or communities in a network, and micro analysis has as its focus the individual constituents (nodes).

2.2.1.1. Micro-level centrality

One way to understand network cohesion is to look at the role of individual nodes in it, and the location of nodes is generally referred to as centrality (Scott, 1991). Centrality is a crucial property of networks as it influences dynamical processes, such as activation spreading, synchronization, and information transmission between nodes. Many metrics have been suggested that quantify aspects of node centrality in networks (see, e.g. Rodriguez, 2019), and the ones that have known or potential implications for phonological networks are discussed below (for mathematical details see Barabási, 2016; Rodriguez, 2019; Wasserman & Faust, 1994).

2.2.1.1.1. Degree centrality

This measure counts the number of links of a node and relates it proportionally to all other links in a network, and is equivalent to the notion of phonological neighbors in phonological networks. It can be calculated with the following equation

$$C_D(u)=\sum_{v\in V\setminus\{u\}} a_{uv}$$

where N is the total number of nodes in the network, *v* represents a particular node, and *u* is the node for which degree is measured. If the two nodes are linked, the adjacency matrix value

a_{uv}=1. Degree centrality of a node is calculated by starting with 1 and adding values up to N. A node can potentially have a degree of up to N-1 if it is linked to all other nodes in a network. Degree distribution refers to the probability that a node has *k* links to other nodes (see Figure 6 for a schematization).

In weighted networks nodes are linked via different edge strengths, and the notion of degree centrality has been expanded to include the sum of all edge strengths between each node and its neighbors (Roberts, 1976). In phonological networks this can represent the phonological/phonetic distance between phonological neighbors (e.g., Fricke et al., 2016). Words with many phonologically close neighbors (e.g., *bat*: pat, bad, bag) show higher weighted degrees than neighboring words with greater phonological distance scores (e.g., *fog*: dog, fig, for). The equation for weighted degree is expressed as

$$C_D(u)=\sum_{v\in V\setminus\{u\}} w_{uv}$$

with *w* being the weighted adjacency matrix, in which $w_{uv}>1$ if nodes *u* and *v* are linked, and the value representing the exact weight of the edge (see Opsahl, Agneessens, & Skvoretz, 2010).

2.2.1.1.2. Closeness centrality

This measure accounts for the length of the shortest paths from a given node to all other nodes in a network and is a useful estimate of the speed of information flow through one node to all others (see Figure 6). Closeness of a node is defined as

$$C_{Cl}(v)=1/\sum_{w\in V} d_{v,w}$$

where the distance matrix value $d_{v,w}$ represents the shortest path between nodes v and w, and the sum of path lengths from node v to all other nodes is accounted for. Closeness centrality scores range from 0 to 1, with 1 indicating that a node is close to all other nodes (Metcalf & Casey, 2016). While there are conceptual similarities to degree centrality, closeness centrality is a mathematically different measure and not correlated with degree centrality (Ko, Lee, & Park, 2008). In highly-connected networks, a large number of nodes may show similar closeness centralities, and the measure is thus best applied to sub-groups (or clusters) of nodes in a network (Salavati, Abdollahpouri, & Manbari, 2019). Eccentricity is the opposite of closeness centrality and measures the maximum distance of a node from any other node in the network; the maximum eccentricity value is the graph diameter. Phonological word forms that show high closeness centrality can potentially spread co-activation more efficiently in phonological networks, as they show short path distances to numerous other nodes. Some studies have reported an effect of closeness centrality on lexical processing (Goldstein & Vitevitch, 2017; Iyengar, Veni Madhavan, Zweig, & Natarajan, 2012) but the effects seem counterintuitive to standard lexical processing theories: Goldstein and Vitevitch (2017) found a recognition advantage for words with high closeness centrality. Contrary to the many-neighbors-disadvantage postulated by models of word recognition (e.g., Vitevitch & Rodriguez, 2004), words that are close to many others in the lexicon were retrieved faster and more accurately in the study by Goldstein and Vitevitch (2017). The authors explain their findings by hypothesizing an accumulation of partial activation benefits over time that strengthens cognitive representations of central words with high closeness centrality scores, eventually leading to a lexical access advantage. Goldstein and Vitevitch (2017) hypothesize that an accrual of partial activation through activation of neighbors may take place over time as the lexicon ages. Specifically, they state that "this accrued partial activation can yield processing benefits in the future" (p. 8). A derived implication in the context of phonological network growth is that during the course of vocabulary build-up, words with high closeness centrality scores become strengthened, benefit from a lexical processing advantage, and play a more central role in network growth.

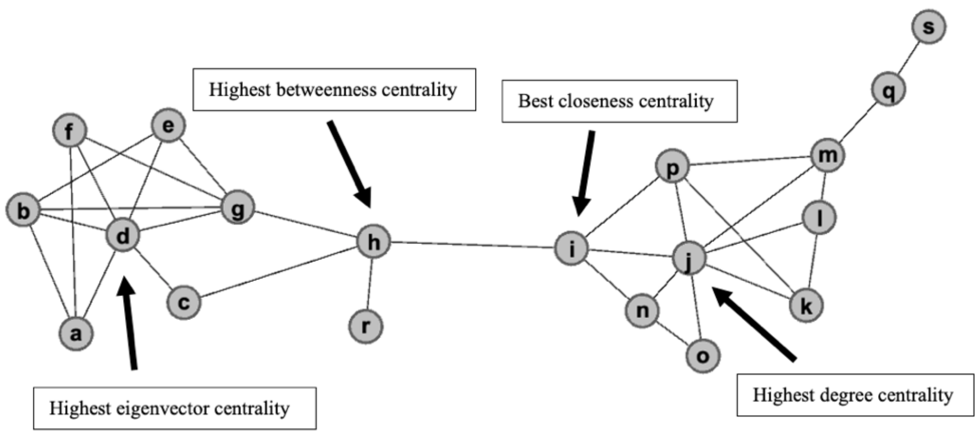


Figure 6: Degree, closeness, betweenness, and eigenvector centralities in a network.

2.2.1.1.3. Betweenness centrality

Betweenness measures the times a given node lies on the shortest path between other nodes; specifically, it measures the percentage of shortest paths that lead through a node (Goldbeck, 2015; see Figure 6). Nodes of high betweenness centrality function as 'bridges' (Gerometta, 2015) that link groups of nodes and they represent crucial points that govern the information flow in a network. The following equation yields betweenness scores

$$C_B(v)= \sum_{s\neq v\neq t} (\sigma_{st}(v)/ \sigma_{st})$$

where σ_{st} stands for the total number of geodesic distances from node *s* to node *t* and $\sigma_{st}(v)$ is the number of those paths that lead through *v*. Weighted networks take into account edge weights, and a node's betweenness score in a weighted network is given by the sum of the weights of its adjacent edges

$$C_{wB}(i)=\sum_{j=1}^{N} a_{ij}w_{ij}$$

where a_{ij} and w_{ij} are adjacency and weight matrices between the nodes *i* and *j*. Here, a higher betweenness score is assigned when a node has many strong (high weight) edges. In phonological networks, some words occupy key positions of higher betweenness centrality. Previous research suggests that words in strategic key positions of a phonological network are strengthened by repeated co-activation (or partial activation) that they receive through activation of nearby words, and this results in lexical processing advantages (Vitevitch & Goldstein, 2014). Words with high betweenness centrality play an important role in the community structure of a network (i.e., subgrouping of nodes according to similarity, Siew, 2013) and create cohesion between phonological subgroups in the network (Gerometta, 2015).

2.2.1.1.4. Eigenvector centrality

This centrality measure assigns nodes a score based on their influence in a network; this is achieved by taking into account the number of neighbors and their centrality in the network (Bonacich, 1972, 2007; see Figure 6). Highly-connected nodes (=high degree centrality) that are linked to other highly-connected nodes are deemed most influential in a network. Thus, this centrality measure is appropriate when it is assumed that the status of a node is a positive function of the status of its neighbors (Bonacich & Lloyd, 2015). There are a number of efficient but mathematically complicated algorithms to calculate eigenvalue and eigenvectors for networks, incorporating features such as edge centrality (Xu, Feng, & Qi, 2021), different interaction patterns between nodes (Carreras, Miorandi, Canright, & Engo-Monsen, 2007), the number of interactions between nodes (Katz centrality: Katz, 1953), and link quality between nodes (PageRank: Page, Brin, Motwani, & Winograd, 1998). Generally, links with high-scoring eigenvector centrality nodes will contribute more to the eigenvector centrality score of a particular node than links with low-scoring nodes, following the logic that not all neighbors are equal. The information that is available to a node is limited to what the neighbors pass on, and highly central neighbors possess more information. In phonological networks, high eigenvector centrality of a node could indicate more efficient activation spreading through highly-central node connections. Vitevitch et al. (2011) and Chan and Vitevitch (2009) suggest that lexical retrieval could be facilitated by eigencentrality-centered search algorithms (similar to what Griffiths, Steyvers, & Firl, 2007, described for semantic search).

2.2.1.2. Assortativity by degree

Another graph-theoretical concept that has been shown to have an effect on phonological networks is assortativity by degree (also known as 'assortative mixing by degree' or 'homophily', see Peel, Delvenne, & Lambiotte, 2018). It is the extent to which nodes link with other nodes that have similar degrees, and it is calculated as a correlation between two node degrees. The most popular mathematical approach to assortativity by degree is Pearson's assortativity coefficient, also known as degree correlation r (Newman, 2002a), where Pearson's correlation coefficient is calculated for degree centrality of two linked nodes. Positively assorted networks show clustering of high-degree nodes, whereas negatively assorted networks (=disassortative networks) show a propensity for high-degree nodes to link up with low-degree nodes. Assortativity is generally high in phonological networks (r=0.56-0.76, in Arbesman et al., 2010; Vitevitch, 2008), which means that high-density neighborhoods tend to be linked to other high-density neighborhoods, whereas low-density neighborhoods tend to be linked to other low-density neighborhoods (see Gravino, Servedio, Barrat, & Loreto, 2012; Van Rensbergen, Storms, & De Deyne, 2015, for similar findings in semantic networks). Vitevitch (2008) explains that assortative mixing in phonological networks could represent advantages for lexical processing: if disassortative mixing was the rule, the distribution of highly connected nodes throughout the network would result in wide activation spreading among a large number of nodes. This means that a large number of lexical competitors would have to be processed for each word recognition event, leading to slower and more laborious speech recognition. Activation spreading and lexical competition remain more localized in a highly assortative network, making lexical retrieval more efficient and quicker. The implications of high degree assortativity and lexical processing are twofold. Accumulation of phonological neighbors facilitates production and inhibits retrieval of words (Chen & Mirman, 2012; Dell & Gordon, 2003; see section 2.1.). Speech perception being generally slowed by co-activation of numerous neighbors (Luce & Pisoni, 1998; McClelland & Rogers, 2003; Vitevitch & Luce, 2016), Arbesman et al. (2010) argue that low rates of degree assortativity in phonological networks strengthen lexical representations and reduce retrieval errors. Inversely, high assortativity in phonological networks could indicate a lexical organization that favors production processes.

Assortativity also has implications for network failure, and networks with high positive assortative mixing are not easily disrupted by targeted removal of high-degree nodes (Newman, 2002a, see section 4.2.2. for a discussion of network robustness). If one highly connected word is lost in an assortative network, only a small impact on the overall connectivity and activation spreading patterns within a network is measurable (McClelland, Rumelhart, & Hinton, 1986). If one highly connected node is lost in a disassortative network, major implications arise for overall connectivity, with network break-down as a likely outcome (Newman, 2002a). This prediction was confirmed in a network study on misperceived words, so-called "slip of the ear" errors, in American English. Vitevitch, Chan, and Goldstein (2014) found a correlation between neighborhood density/degree of words that are produced by speakers and those misheard by listeners. The fact that the misheard word and the correct word were from similar assortative neighborhoods indicate that speech recognition utilizes assortative patterns in a phonological neighborhood.

2.2.1.3. Macro-level analysis

Macro analyses of networks traditionally deal with a range of graph theoretical constructs, including the *giant component*, *islands*, and *singleton nodes* or hermits (Barabási, 2016). Of significant interest in network science is the giant component, which is the largest connected portion of a network (see Figure 6) that allows each node to reach any other node by "traversing a suitable path of intermediate collaborators" (Newman, 2004, p. 5202). The size of the giant component determines the geodesic (=shortest) distance between two random nodes (Barabási & Albert, 1999), and provides information about the topological make-up of a network. Giant component formation is not necessarily related to the number of nodes in a network, but small networks may have large giant components. Assortative mixing has implications for giant component growth: positive assortative mixing by degree may result in the giant component gaining in size over time, as high-degree nodes are linked to other high-degree nodes (Newman, 2002a). Disassortative mixing by degree, on the other hand, may lead to the dynamic of high-degree nodes tending not to join the giant component, maintaining instead as a small proportion of the network.

Islands, or "lexical islands" in language networks, are smaller, linked components that are not connected to the giant component (Arbesman, Strogatz, & Vitevitch, 2010). Singleton nodes, or "lexical hermits", are isolated nodes with no connections. Figure 7 shows the macro level network components of the A2 learner network (see Appendix A for further graphs).

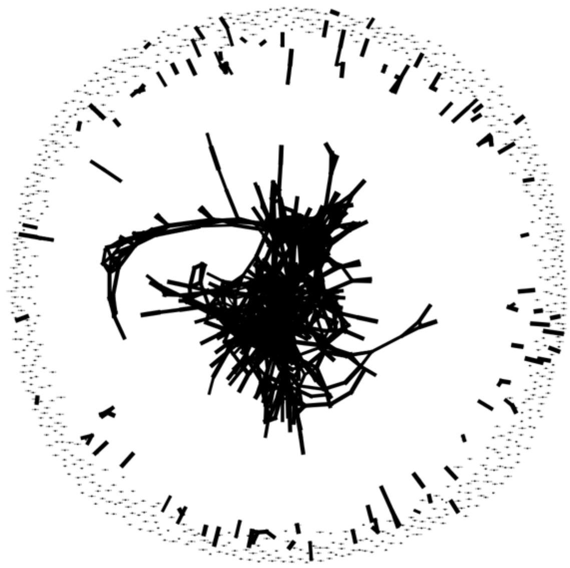


Figure 7: The giant component of the A2 learner network of the present study at the center, surrounded by lexical islands and singleton words (=grey dots). Yifan-Hu visualization in Gephi.

The beta index (β) is a simple connectivity measure of the ratio of edges to vertices within a network (Arbesman et al., 2010). It is expressed by the relationship of the number of edges (e) over the number or nodes (v), thus $\beta = e/v$, and is a good indicator of network efficiency (Linehan, Gross, & Finn, 1995). Beta values of >1 signify more complex levels of connectivity (Haggett & Chorley, 1972), as opposed to $\beta=1$ (=circular connectivity) or $\beta<1$ (tree-like dendogrammatic network). An efficient network with high β links a large number of nodes together, minimizing average path lengths and geodesic distance. In addition to a high β index, an efficient network where nodes can easily link to one another and exchange or spread information is also characterized by short path lengths between nodes and a small network diameter, defined as the longest path of the shortest paths across a network (Latora &

Marchiori, 2001). The mean path length (ℓ) is the average of all path lengths in a network and estimated as

$$\ell = 1/[N(N-1)] \sum_{i \neq j \in V} d_{ij}$$

where N is the number of nodes in a network and d_{ij} denotes the shortest distance between nodes *i* and *j* (see Watts & Strogatz, 1998). The network diameter δ is found by calculating all the shortest paths between all nodes following the equation

$$\delta = \max\{s(i, j)\}ij$$

with s(*i,j*) being the number of edges in the shortest path from node *i* to node *j*. The longest of all paths is then chosen as the network diameter.

In phonological networks, path length and diameter of the giant component are crucial measures of how densely words are crowded in phonological space in the lexicon. Word are linked beyond the immediate neighborhood and even distant words, connected to a target via multiple paths, can become lexical competitors. Vitevitch, Goldstein, and Johnson (2016) showed that phonological associations more frequently contained words from the distant neighborhood when the target word had a sparse neighborhood (low-degree target word). Here, activation was suggested to spread out from the immediate neighborhood, with only few near neighbors containing the activation diffusion close to the target word. In semantic word processing, different effects of near and distant semantic neighbors on a target word have been suggested, with near neighbors exerting an inhibitory effect but distant neighbors exerting a facilitative effect on semantic processing (Mirman & Magnuson, 2008). The authors explain this as a competition effect, where semantically similar words (=near neighbors) heighten competition, while distant neighbors create a graviational gradient that helps identify the correct "attractor" (=target word). The authors argue that this dynamic should also apply to phonological processing.

Many networks show a tendency to form links between neighboring nodes, a process called clustering. Node clustering is quantified by the clustering coefficient (CC) and calculated using the equation

$$CC = 2e_n/(k_n(k_n-1))$$

with k_n being the number of neighbors per node (n) and e_n the number of connected pairs between all neighbors (Watts & Strogatz, 1998). Hence, CC=0 if none of the neighbors of a node are linked, and CC=1 if all are. In probabilistic terms, the clustering coefficient expresses the likelihood of a connection between two arbitrary neighbors of a node. Clustering coefficients in phonological networks assess the number of phonological neighbors of a target word that are also neighbors of one another (Goldstein & Vitevitch, 2014). Words with low clustering coefficients (i.e., low-C words) were found to be recognized faster and more accurately in spoken word recognition (Chan & Vitevitch, 2009; Yates, 2013), findings that mirror neighborhood density and word recognition (e.g., Luce & Pisoni, 1998; Vitevitch & Rodriguez, 2004). Neighborhood density and the clustering coefficient are different measures, with neighborhood density accounting for the number of neighbors and the clustering coefficient accounting for neighborhood connections between neighbors. The two measures are generally not correlated (Chan & Vitevitch, 2009; Yates, 2013). In the realm of word production, the clustering coefficient behaved differently compared to neighborhood density, and it was shown that low-C words are produced faster and more accurately (Chan & Vitevitch, 2010). High clustering coefficients of phonological neighborhoods seem to interfere with lexical retrieval and production, a finding potentially linked to the diffusion of activation across a lexical network (see more on the activation diffusion theory in section 2.3.). In terms of language acquisition, clustering coefficients have been shown to be indicative in children acquiring the semantics of their first language: normally developing children show higher clustering coefficients in their semantic networks than those delayed in their semantic development (Beckage et al., 2011).

A macro-level network measure closely related to the clustering coefficient is transitivity, the probability of adjacent nodes being connected (Luce & Perry, 1949; Newman, Strogatz, & Watts, 2002). Values are bounded between zero and one, with high values indicating good transitivity. While conceptually similar, transitivity differs mathematically from the clustering coefficient and typically leads to different values (Schank & Wagner, 2005). Specifically, the consideration of triangles (or triads) of nodes constitutes a crucial difference in the two measures. Triads are subgraphs between three nodes in a graph, which can all be linked to one another (the triangle), not linked at all (the null triad), or show some links (e.g., one-edge-subgraph). The clustering coefficient takes into account the ratio of the number of triangles of a node in relation to the number of potential triangles that could involve that node, whereas transitivity accounts for the total number of triangles in a network divided by the total number of triads (Estrada, 2016). A transitive triad is a group of three nodes that are completely interconnected; an intransitive triad may only have two edges. The transitivity equation follows

$$T=3t/P_2$$

where t represents the triangles in a network and

$$P_2 = \sum_{i=1}^{N} k_i(k_i - 1)/2 \quad \text{(see Wasserman \& Faust, 1994)}$$

Network transitivity is a related measure to the clustering coefficient and frequently used in undirected graphs (Newman et al., 2002), however the two can yield different results (see Schank & Wagner, 2005, for a mathematical description). Both measures have been calculated in studies on phonological networks (Arbesman et al., 2010; Chan & Vitevitch, 2009). In the present study, global transitivity was calculated to serve as an indicator of network connectivity (the probability of adjacent nodes to be connected) across the whole network, while clustering coefficients were calculated for each individual node, estimating the interconnectedness of a node's immediate neighborhood.

Networks that are characterized by short average path lengths and high clustering coefficients or transitivity are known as small-world networks (Watts & Strogatz, 1998). Such networks show exceptionally dense connectivity patterns, which enable fast and efficient information propagation (see Figures 8a and 8b). The small-world coefficient σ proposed by Humphries and Gurney (2008) compares the ratio of network transitivity and path length to their equivalents in a random network. The conditions for a small-world network are defined as $C \gg C^{rand}$ and $L \approx L^{rand}$, which results in $\sigma > 1$.

7a 7b

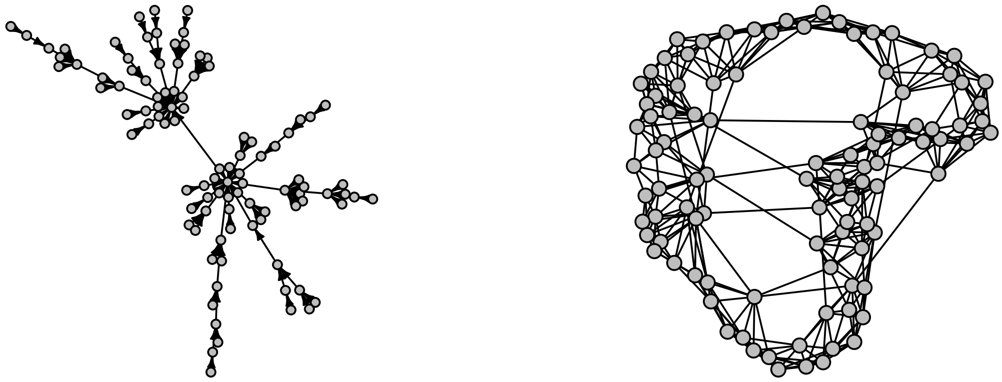


Figure 8: Scale-free (a) and small-world (b) networks (constructed with R package "igraph", functions "barabasi.game", "watts.strogatz.game").

Phonological networks display small-world structuring, with ensuing rapid, accurate, and robust lexical processing (Gerometta, 2015; Vitevitch, 2008). Small world structuring is important for lexical retrieval and phonological networks structured in that way are optimally designed for rapid and robust lexical retrieval processes (Goldstein & Vitevitch, 2014). In a small-world phonological network, the extreme inter-connectedness of phonological neighbors and the short distances that connect all words could mean that phonological activation spreads more widely and potentially encompasses neighboring neighborhoods. As a target word becomes activated, and neighbors become co-activated, their neighbors may also receive some of the co-activation, potentially bouncing activation back and forth in the tightly connected network. Figure 9 schematizes this activation diffusion in a small-world phonological network.

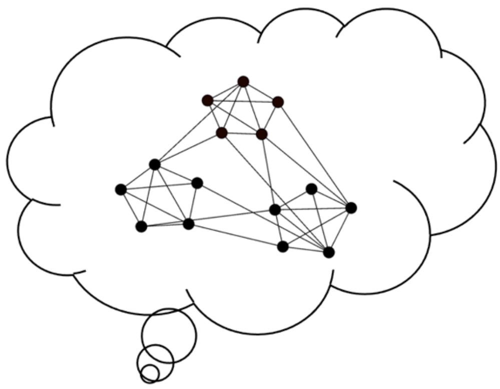


Figure 9: Lexical activation diffusion in a small-world phonological network.

There is some debate over the concept of small-worldness in phonological networks. Kapatsinski (2006) argues that small-worldness and/or short average path lengths are not required for efficient lexical retrieval, as lexical search is generally restricted to small localized neighborhoods that share sublexical similarity features, with short path lengths across the giant component or the overall network playing a minor role.

2.2.1.4. Meso-level analysis

In addition to the macro and micro level analyses that are commonly performed on networks, another layer of mesoscopic dimension rests between the two. Here, the focus is on smaller sub-graphs embedded within the larger network, a natural tendency of nodes to cluster together in so-called *communities* (Ravasz & Barabási, 2003; see Figure 10). Communities are sub-groups of nodes in a network and can play an important role for network structure (Barabási, 2016). They can partition a complex network into smaller, more interpretable parts and reveal coarse-grained network interactions and relations (Hoffmann, Peel, Lambiotte, & Jones, 2020).

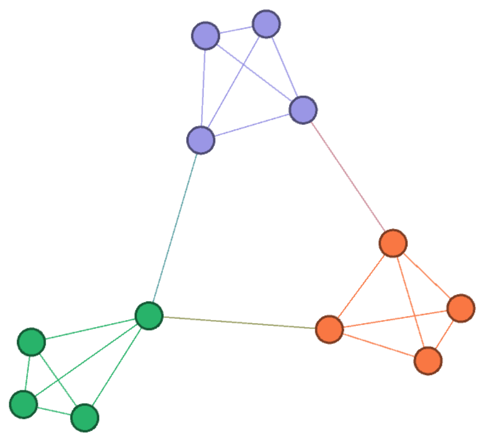


Figure 10: Three communities in a simple network.

In graph theoretical terms, all nodes of a community must be linked to one another or be able to be reached through other nodes of the community. In addition, nodes belonging to a community show a higher probability of being linked to other nodes of that community than to nodes outside of the community (Porter, Onnela, & Mucha, 2009). Nodes of high betweenness centrality tend to play an important role in community formation, as they link the different sub-groups together. There are various methods for community detection, including modularity (Newman & Girvan, 2004). The starting point here is maximization of difference between actual number and expected number of edges in a community, which needs to be optimized at a later point. One of the most popular optimization methods is the so-called Louvain algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) and it has proven useful in community detection in phonological networks (Siew, 2013). A modularity value *Q* quantifies the density of links within communities as compared to the density of links between communities (Newman, 2006). It is mathematically defined as

$$Q=\sum_{i=1}^{k}(e_{ii}-a_i^2)$$

with e_{ii} representing the probability that edge *e* is in module *i*, and a_i^2 representing the probability that a random edge would be classified into module *i*. *Q* ranges from -1 to 1. A high *Q* value close to 1 indicates that more edges are found within a module than would be expected by chance (Rosvall, Delvenne, Schaub, & Lambiotte, 2017) and is a good indicator that community structure is well optimized within a network. The Louvain method (a so-called greedy heuristic, Blondel et al., 2008) first puts each node in its own cluster and then repeatedly merges two clusters that increase modularity by the largest amount. The process ends when all merges lead to reduction of modularity (or *Q*).

The only study of communities in phonological networks (Siew, 2013) demonstrated that community size can predict various lexical characteristics of words in the lexicon of American English first-language users. Large communities house phonemically shorter and more frequent words, show higher phonological neighborhood density (i.e., node degree), and are acquired relatively early during childhood language acquisition (see Siew, 2013). This has implications for language learning/acquisition: the early stages of language learning may be geared toward building robust communities as a foundation for lexical growth in the future. In addition, community structure in phonological networks is suggested to influence the flow of lexical activation among neighboring words and communities: communities can better contain activation, thereby limiting the phonological search space and lexical competition (Siew, 2013). If activation remains within a community with minimal spill-out, each word within the community receives a larger share of the activation, while at the same time keeping competitor words out of the activation loop. Community formation could play an important role in activation-competition restriction and optimizing lexical retrieval to high efficiency. There is undoubtedly great utility of the network community concept for phonological networks of second language learners in terms of both learning trajectory and lexical retrieval efficiency in the mental lexicon.

2.3. Diffusion of lexical activation

Diffusion theory (or diffusion dynamics) is the analysis of how an innovation or a bit of information spreads through an interconnected network, and different diffusion models have been proposed in the network literature (Dearing, 2008; Gomez-Rodriguez, Leskovec, & Krause, 2012; Ren, Yang, Yang, Xu, & Yang, 2012; Rogers, 2003; Zhang & Gan, 2018). Diffusion processes typically consist of three components: the population where they unfold, the mechanisms governing their evolution and development, and the content of the diffusion (Milli, Rossetti, Pedreschi, & Giannotti, 2018). There are two main branches of diffusion modelling, characterized by different predictions for overall network behavior: (1) purposeful information spreading and (2) non-selective information spreading (Wang et al., 2011). In human social networks, information spreading tends to be purposeful, as not every person will pass on information to all of the people with whom they interact. Such selective spread of information to specific individuals is complex to model and various factors need to be accounted for that can explain people's motivations to share information with select others (see, e.g., Alexy, George, & Salter, 2013). Models of non-selective spreading, on the other hand, follow the simple rule that each node will potentially infect all of its neighboring nodes (e.g., Keeling & Eames, 2005). Generally, a number of factors can play a role in spreading patterns but nodes do not generally possess the ability to deny the transmission of information to another node upon interaction. While non-selective network propagation models stem from viral spreading models (including computer viruses, Kephart & White, 1991), there are crucial differences to activation spreading in phonological networks. In viral spreading, connected social agents or computers do not always transmit the virus: if a person does not meet their friends, they cannot infect them; if a computer is equipped with an anti-virus software, they may prevent being infected by a connected machine. In phonological networks, however, shared phonemes will always become co-activated by neighboring words upon lexical access/retrieval and production. There is no (known) property of segments or words that may prevent co-activation of shared phonemes under particular circumstances analogous to purposeful spreading in social networks. There is also no mechanism that could make a neighboring node resistant to activation reception. Thus, activation of phonological word forms will always reliably lead to co-activation of shared phonemes with neighbors in a phonological network.

Two categories of virus propagation models are commonly recognized: homogeneous and heterogeneous models. In homogeneous networks, all nodes are fully interconnected and viruses can be propagated without dependence on network typology (Zhu & Cen, 2017). Heterogeneous networks are not fully interconnected and their typological make-up plays an important role for determining how viral infection can spread between nodes (Kjaergaard, Brander, & Poulsen, 2010): not all nodes can reach all other nodes and viruses can remain localized to specific parts of the network. Phonological networks can certainly be characterized as heterogeneous: information from islands or singleton nodes cannot spread to the giant component.

A number of network-relevant measures have been implicated in activation spreading in phonological networks, chief among them the overall network typology. The sizes of the giant component, islands, and the ratio of singleton nodes have paramount influence on the ability of individual nodes to exchange activation. Larger giant components mean more (and potentially more wide-spread, see below discussion on clustering coefficient) diffusion of activation, encompassing a larger array of words (Siew & Vitevitch, 2016). If a network contains many small islands and/or unconnected singleton nodes, activation will be diffused only minimally. In terms of lexical processing, the latter scenario means less lexical competition as there are fewer co-activated words. This can have great lexical retrieval benefits (Siew & Vitevitch, 2016). Clustering coefficients of network parts or whole networks (small-worldness) also have crucial impact on activation diffusion. Networks with low clustering coefficients generally show wider spread of information because there are fewer densely (or exclusively) interconnected regions of the network. In contrast, networks with high clustering coefficients will restrict information transmission to the interconnected regions, which are in turn less strongly linked to other regions of the network (Naug, 2008; Newman, 2003a). Individual network neighborhoods can also have activation-restricting or activation-propagating features. Figure 11 shows how activation can be localized to singleton nodes or less densely connected neighborhoods, and how it can be propagated by a highly interconnected neighborhood.

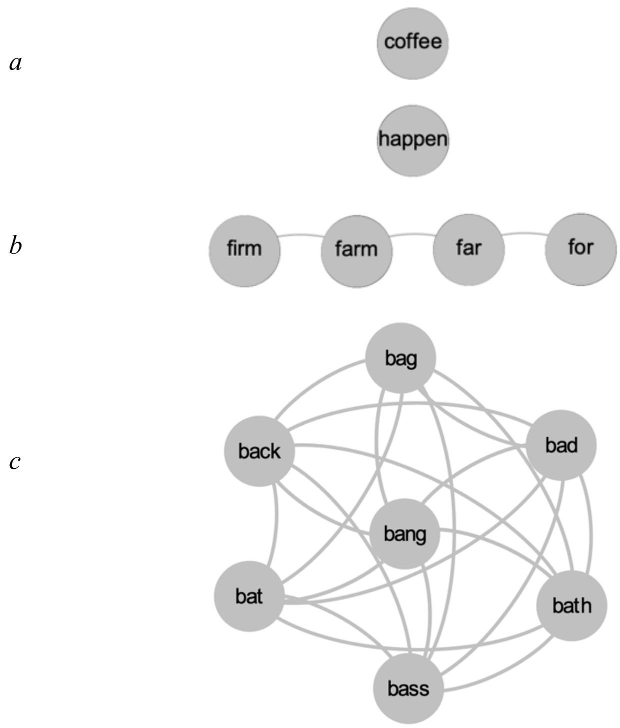


Figure 11: Activation-restricting network parts (a, b) are smaller, less-connected entities. Activation-propagating network parts (c) are larger, interconnected entities.

An activation diffusion theory for phonological networks was proposed by Chan and Vitevitch (2009) and further elaborated on by Vitevitch and colleagues (2011). Activation starts with a node and spreads along phonological similarity paths to neighboring words, from which it is passed on further to neighbors of neighbors. Some of the activation spreads back to the originally activated node, which is then given an additional activation boost. Over time, activation keeps building up in a specific node, which is then retrieved as the target node. The more interconnected the neighbors are (such as in Figure 11c), the more activation bounces back and forth in the neighborhoods. The "badge" neighborhood displayed in Figure 11c is highly interconnected and characterized by a high clustering coefficient (see Chan & Vitevitch, 2009). Here, activation is spread widely among the neighboring words, making it more difficult for listeners to hone in on the target word among the lexical competitors. In the case of the neighborhood shown in Figure 11b, activation of "firm" spreads to "farm", and from there some of it goes back to "firm" and some spreads further to "far". The overall portion of the activation that "firm" receives is much larger compared to any word in the 11c neighborhood, making lexical retrieval faster and more efficient.

The diffusion model has direct implications for lexical processing of words residing in different parts of the phonological network. Words in the giant component are hypothesized to be more laborious to retrieve, due to the wider activation spreading through the greater neighborhood. As shown by Siew and Vitevitch (2016) in their study of lexical retrieval in relation to phonological network partitioning, words in lexical islands and singleton nodes are indeed easier to retrieve. When a node resides in the giant component, the activation diffuses out to a greater degree, with each node in the giant component receiving less share of the activation, rendering lexical retrieval a more difficult and time-consuming task (Siew & Vitevitch, 2016). A similar study by Vitevitch and Castro (2015) reports higher accuracy in picture naming of words in islands and even singleton words. Thus, lexical islands (and singleton words) possess a clear lexical retrieval advantage over giant component words (Siew & Vitevitch, 2016). Based on this, it would be expected that giant component size of a phonological network plays a role in speed and accuracy of lexical access. In addition, network diameter seems relevant: farther distances in the giant component could mean that activation ebbs faster, not reaching many distant neighbors. By contrast, if a giant component has a short diameter, activation diffusion could reach farther and potentially include many distant neighbors of the target word. The findings reported by Vitevitch and his colleagues (2015, 2016) are difficult to reconcile with psycholinguistic models of lexical processing, which would predict a processing advantage in the highly connected giant component that contains shorter and more frequent words (Siew & Vitevitch, 2016). Following the rationale that word length and frequency are correlated (Zipf's law) and the known lexical processing benefits that frequent words entail (see, e.g., Brysbaert, Mandera, & Keuleers, 2017, for a review of the word frequency effect), it would be expected that giant component words are retrieved more efficiently. In addition, the larger share of co-activation that giant component words receive through phoneme-sharing should lead to giant component words being processed more rapidly and accurately by listeners. The diffusion model of lexical activation spreading can help explain the discrepancy between expected and actual findings in terms of lexical access in giant component words, underscoring the need to integrate a network-scientific approach in theories of lexical access.

Vitevitch and colleagues (2011) ran a computer-simulated version of their proposed lexical diffusion process and attributed an initial burst of activation to a target node, which was programmed to retain a proportional share of it and then pass on activation to the nearest neighbors. The portion of the activation that was passed on to the neighbors was equally split between all neighbors, which in turn gave off activation to their own neighbors (=neighbors of neighbors of target node), all the while retaining a portion of the activation. This simple model of activation diffusion was able to mimic lexical retrieval delays for highly connected words, as observed by Siew and Vitevitch (2016). Siew (2019) created a simulation model for the spreading pattern modelled after Vitevitch et al. (2011) and was able reproduce the exact results. While the activation diffusion model is preliminary and based on 24 mini networks (Vitevitch et al., 2011), it nonetheless presents an intriguing view of lexical processing being potentially analogous to diffusion dynamics. One important caveat concerns the distance between the phonological neighbors and the fact that some neighbors are more closely related than others. Ergo, activation diffusion should take into account the phonological or phonetic distance between phonological neighborhoods. This is also acknowledged by Vitevitch and colleagues (2011).

Figure 12 illustrates how lexical activation mediated by phonological distances can alter diffusion dynamics in a network. Edge numbers indicate similarity scores between phonological neighbors based on a feature-weighted distance measurement for comparative linguistics developed by Kondrak (2000) and Downey et al. (2008). Similarity scores range from 0 (=no similarity) to 100 (=identical) [see section 3.1.1. for more details on the similarity computations].

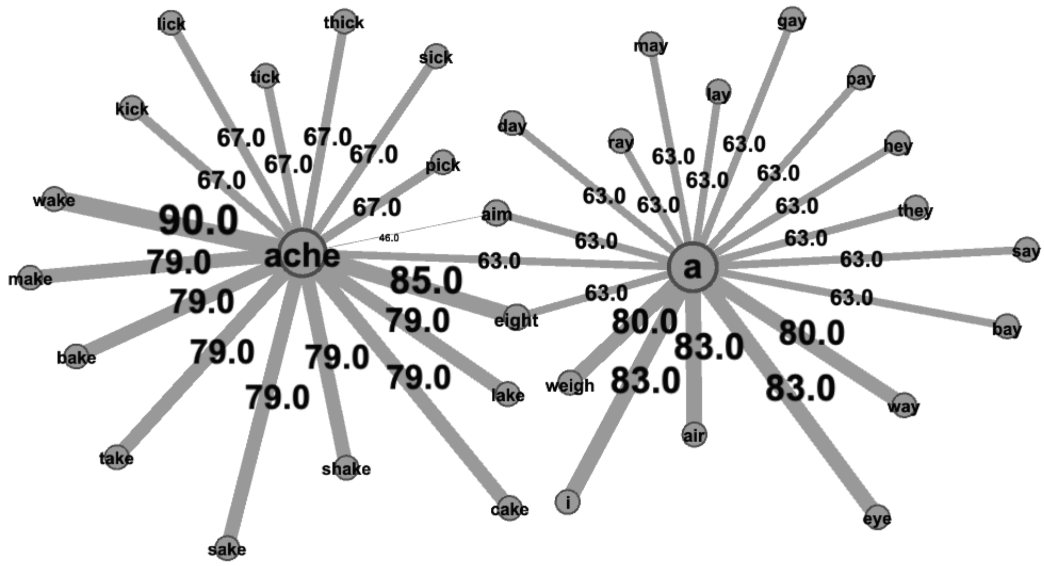


Figure 12: Phonological neighborhoods of "a" (left) and "ache" (right) with phonological distance measurements. Constructed from the B2 learner network of the present study.

Even though the sum of all phonological distances is relatively equal in the two neighborhoods shown in Figure 11 (“ache”=1228, mean=73, “a”=1239, mean=68), the “ache” neighborhood is characterized by a larger number of close relationships between words. In the “ache” neighborhood, nine words have phonological distances of >79, while in the “a” neighborhood, only five words have such a close phonological distance. Conversely, the “a”-neighborhood houses a higher number of words which are phonologically more distant to the target word (N=13), while the “ache”-neighborhood has fewer of those words (N=7). According to the current account of the diffusion framework for phonological processing by Vitevitch and colleagues (2011), activation diffuses similarly through both neighborhoods and is essentially only restricted by the number of neighbors there are (i.e., degree per node), which is quite similar in the two neighborhoods (k”a”=18, k”ache”=16). However, it is reasonable to assume that phonologically closer neighbors share a larger portion of the overall activation or, at the very least, activation spreads faster through phonologically closer neighbors. This would naturally result in modified diffusion patterns from what was outlined by Vitevitch and colleagues (2011). For instance, more activation would spread from “ache” to “a” than vice versa, since “ache” passes on more activation to “eight”, which then in turn passes on a smaller portion of that to “a”. Such fine-grained activation patterns accounting for phonological distances can contribute to the diffusion hypothesis of phonological networks.

Vitevitch and colleagues (2011) identified a number of network measures that can serve as predictors for activation spreading: degree, clustering coefficient, and network density. Words with low degree (=few neighbors) and low clustering coefficients in their immediate neighborhood received the largest proportion of overall activation in their neighborhood faster than words with many neighbors and high clustering coefficients in their neighborhood. In addition, denser networks with more interconnected neighborhoods diffused activation in a way that target word identification (via the largest share of activation accrual in the simulation) was impeded. The authors suggest that their model of activation spreading can account for the observed phenomena in phonological network connectivity and lexical access that have been reported by Chan and Vitevitch (2009).

In the diffusion model of phonological activation spreading, activation “spreads unimpeded between connected nodes” (Vitevitch et al., 2011: p. 11) and does not decay over time. This, however, could lead to what has been referred to as “heat death problem” (Berg & Schade, 1992), an overactivation of too many (irrelevant) words leading to inefficient processing.

Memory decays with time and node activation levels may too. The initial burst of activation in a phonological neighborhood may not be fully retained until the activation back-and-forth in the neighborhood is finished. A target node must keep the largest share of the activation to be selected as the lexical candidate, and accrual of activation will be impeded by activation decay. A spreading activation model by Bock and Levelt (1994) proposes a similar procedure but with activation decay: an activation signals spreads through the mental lexicon along associations (in semantic and phonological aspects), activation bounces back between associated words and finally focusses on the target word. In addition, inhibitory connections may be present (Baronchelli, iCancho, Pastor-Satorras, Chater, & Christiansen, 2013; Dóczi, 2019). Most network diffusion theories incorporate a time component, which predicts activation transmission lessening over time (Liu & Kuan, 2016; Rogers, 2003).

Recent research has shown that network science can be useful to gain knowledge on lexical processing by being able to go beyond the localized measure of neighborhood density and describe the phonological word form lexicon from a global perspective (Siew & Vitevitch, 2016). As demonstrated by the lexical activation diffusion hypothesis, the lexical activation of a target word is not only dependent on its immediate neighborhood but the larger interlinked parts of which the neighborhood is a member. The embedding of phonological neighborhoods within a larger network structure has opened new research avenues that may potentially explain some of the discrepancies in studies on phonological neighborhood density from the past. For instance, neighborhood density in Spanish and Russian was shown to have the opposite effect from what is known for English, and Spanish and Russian words in dense phonological neighborhoods impair word production but facilitate recognition (Arutiunian & Lopukhina, 2020; Vitevitch & Stamer, 2006; also see, e.g., Neergaard, Britton, & Huang, 2019; Sadat, Martin, Costa, & Alario, 2014; Vitevitch & Rodriguez, 2004, for conflicting results on phonological neighborhood density in relation to lexical retrieval/production in various languages). Future studies may show that predictions arising from the global rather than localized network view of the different lexica may be able to resolve this issue and find common linguistic universals in relation to neighborhood characteristics and lexical processing.

A largely unexplored avenue in the linguistic network sciences is the question of how second language learners construct and grow their networks. A small number of studies have explored semantic and syntactic word associations in learners of different languages, where network graphs were drawn based on linked words as elicited in experimental settings with first- and second-language users of French (Wilks & Meara, 2002; Wilks, Meara, & Wolter, 2005) and English (Jiang, Yu, & Liu, 2019; Schur, 2007; Tanaka & Takahashi, 2019). The results reported by Wilks and colleagues indicate higher semantic network density in L1 as compared to L2 language users, which influences linguistic performance in word-association tasks. The authors argue for an inclusion of network theoretical notions in the study of second language word acquisition in the semantic domain. Phonological networks of word forms in the mental lexicon of second language learners have not been explored, and the network methodological approach can help expand our knowledge of how phonological associations are formed in the developing lexicon during second language learning. Since the utility of phonological networks for the study of the L1 lexicon has produced crucial new insights into lexical access and acquisition of word form associations, such knowledge on the L2 lexicon can potentially enhance our understanding of L2 lexical processes. The present study was designed to address the question of phonological network structure in English L2 learners through various proficiency levels, with a special focus on network growth algorithms and their application and utility for descriptions of vocabulary learning. The expected findings can illuminate how the English L2 word form lexicon is structured and what implications the structure may have on lexical processing at different stages of language learning.

3. Learner networks

3.1. Drawing phonological networks

Phonological networks of language learner vocabulary data critically rely on accurate vocabulary lists that closely correspond to the words learners actually know at the investigated proficiency stages. Previous studies constructed phonological networks of L1 English by using psycholinguistic word knowledge dictionaries (e.g. Hoosier Mental Lexicon, see Nusbaum, Pisoni, & Davis, 1984) as their sample of English words known by adults (Siew, 2013; Vitevitch, 2021; Vitevitch & Castro, 2015). Alternatively, the Merriam-Webster Pocket Dictionary from 1964 has been used for American English phonological networks (Arbesman et al., 2010). It is understood that this approach is a less-than-perfect but relatively good approximation of an adult speaker's vocabulary as not every speaker is familiar with 100% of the lexicon data but it can be expected that a majority of words are known on some (passive) level (Vitevitch, 2008). A number of studies have constructed phonological networks from language corpora, for instance the Spanish LEXESP database (Sebastián Gallés, Cuetos Vega, Carreiras Valiña, & Martí Antonin, 2000) or the CALLHOME Mandarin Chinese database of telephone transcriptions (Huang, Bian, Wu, & McLemore, 1997). A phonological network of aphasic speech was constructed with data retrieved from a dataset of aphasic naming tests (Philadelphia Naming Test) consisting of 31,600 utterances (Stella, 2020), and a recent study of semantic and phonological networks of English used a free-association word databank (see De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019; Levy et al., 2021) as its basis. Developmental studies of phonological networks in L1-learning children have utilized different age of acquisition norms for different languages to obtain vocabulary lists per age group, for instance the MacArthur-Bates Communicative Development Inventory (Dale & Fenson, 1996), consisting of 680 words, (see Beckage et al., 2015). Siew and Vitevitch (2020a) used the age-of-acquisition norms reported by Kuperman and colleagues (2012) for English and Brysbaert and colleagues (2014) for Dutch to construct networks with close to 13,000 and 15,000 words, respectively.

3.1.1. Learner lexica

Capturing the entire vocabulary of a specific proficiency stage of English L2 learners is difficult, as there is large variation between individual speakers. Usage-based learner corpora can hardly reflect the whole lexical knowledge of a learner, as only a limited number of all known words are elicited by the tasks of the corpus study (Gráf, 2017). This inherent corpus-study problem equally applies to all types of linguistic corpus research but is particularly pronounced in a network study that relies on an accurate representation of vocabulary knowledge.

The present study gathered data from the Cambridge Learner Corpus (CLC), the largest corpus of English as a foreign language that consists of a collection of written student exams compiled in collaboration with the Cambridge English Language Assessment. It contains language data from over 180,000 students from approximately 200 countries and 138 different first languages and was built to specifically describe English learners' competencies at various stages of the language learning process (McEnery, Brezina, Gablasova, & Banerjee, 2019). Based on it, the English Vocabulary Profile (EVP) word lists, a freely available online resource on vocabulary knowledge of L2 English learners (https://www.englishprofile.org/wordlists) was developed, which was further supplemented with data from leading coursebooks, vocabulary books, and vocabulary lists for standardized examinations (Key English Test or KET, Preliminary English Test or PET). The EVP provides specific wordlists for each of the proficiency levels that are, for the most part, empirically based on the CLC, and thus a good approximation of what learners actually know. There are sub-categories for American and British English varieties.

For the present study, the EVP word lists were used to estimate the vocabulary sizes of learners of English at the proficiency levels defined by the 'Common European Framework of Reference for Languages', ranging from A1 as the lowest to C2 as the highest level. In order to compile vocabulary lists for each of the six proficiency levels, the British English word lists on the EVP website were filtered accordingly and vocabulary data for each level was copied onto a spread sheet. The EVP word lists only contain base words and thus manual sorting and correcting of the data was kept to a minimum. The following cleaning processes were applied: remove duplicate word forms that share meaning (e.g., "address"-noun, "address"-verb; note that semantically non-related words, such as"four/for" were kept in the data), dissolve phrases in constituent parts (e.g., "think", "about", "something"), remove abbreviations (e.g., "Dr", "Mrs"). The only inflections left in the word lists were all forms of the verbs "be" and "have" (e.g., "is", "had"). Clitics were dissolved and noted as separate words (e.g., "it's" → "it", "is"). Acronyms, such as "p.m." or "CD" were left in the sample. Table 1 shows the resulting final vocabulary size of each proficiency level.

Table 1: Vocabulary size estimation for each CEFR proficiency level based on the English Vocabulary Profile.

A1	A2	B1	B2	C1	C2
606	1483	2901	4621	5632	6714

In a next step, phonological transcriptions of all words were extracted with the online *Tophonetics* tool (https://tophonetics.com/) set for British English. Approximately 10% of the transcriptions were compared to the transcriptions given by the EVP website and 100% correspondence was noted. Vowel length symbols and stress indicators were removed from the IPA transcriptions. This was followed by further transcribing all IPA-transcribed words in X-SAMPA (Extended Speech Assessment Methods Phonetic Alphabet or machine-readable phonetic transcriptions) with the use of the online *Phoneverter* tool (http://phonetictools.altervista.org/phonverter/). This was necessary for neighbor determination (see below). Approximately 10% of the automatic X-SAMPA transcriptions were compared to manually coded transcriptions according to the SAMPA coding scheme developed by Wells (1997) and correspondence was 98%.

A number of lexical variables were calculated for each word at each proficiency level. Lexical frequency rates of the words were extracted from the British National Corpus (see https://www.english-corpora.org/bnc/) and log-transformed [LOG(x+1) in order to account for frequency rates of '0']. The resulting values were normalized by a factor of 100 and converted to integer numbers for improved graphic display. Phonological length of words (i.e., number of phonemes) was determined with the Microsoft Excel function "len", which counted the number of IPA symbols that make up the phonological transcription of a target word (minus stress and length symbols). Phonotactic probabilities were computed with the *Kansas*

University Phonotactic Probability Calculator for English (https://calculator.ku.edu/phonotactic/about). Each word was manually transcribed in computer-readable 'Klattese' as required by the software (Vitevitch & Luce, 2004) before being entered into the online calculator. As the calculator uses an American English database, all words had to be transcribed in American English first (also using *Tophonetics*, see above). The Kansas University phonotactic probability calculator computes the sum of all positional probabilities of each segment per word, in addition to the sum of all biphone probabilities (Pearson's correlation between the average of both measures was r=0.6 in the learner data). The average biphone probabilities per word were selected as the variable "phonotactic probability" in the analysis.

Phonological neighbors are defined as two words that differ in only one segment via additions, deletions, and substitutions of phonemes (Levenshtein, 1966; Turnbull & Peperkamp, 2017; Vitevitch, 2008). Levenshtein distances were determined pairwise for each word and with each other word in a learners' lexicon. For example, in the A1 network of 606 words, the phonological distance of each of the 606 words to each of the other 605 words was determined. This resulted in a very large number of possible combinations, so that the use of an Oracle Database, a multi-modal database management system capable of handling millions of rows of data, was necessary. In Oracle 12c (see Bryla, 2015), the function "Edit Distance" (or "Levenshtein Distance") was implemented to compare the string similarities of paired X-SAMPA-transcribed words. The output was the number of character changes (so-called insertions, updates, and deletes) that are required to transform the first string into the second. This number is referred to as the "edit distance" and only edit distances of "1" were extracted for further analysis, as these represented the phonological neighbors. Appendix B lists all phonological neighborhoods calculated for each proficiency level.

After establishment of the phonological neighbors for each of the proficiency level word lists, the phonetic/phonological distances between phonological neighbors were further quantified. Therefore, a feature-weighted distance measurement was computed based on evolutionary applications to language distances and comparisons. The open-source R software package "alineR" (Downey et al., 2017) calculates similarity scores of word pairs according to a set of feature weighting parameters that quantifies phonetic distances following the method of Kondrak (2000) and Downey et al. (2008). The alineR package considers twelve features: syllabic, voice, lateral, high, manner, long, place, nasal, aspirated, back, retroflex, and round.

The input in alineR are pairs of IPA-transcribed words for which a similarity score ranging between 0 (= no distance between two words) and 1 (=maximal phonetic distance) is then calculated. For the present study, the alineR results were transformed to reflect percentage agreement between two words: the values were reverse scaled and multiplied by 100. Thus, 0 reflects the least similarity and a value of 100 reflects the highest similarity between two segments. For instance, a value of 0.3 in alineR was transformed to a value of 70, so that similarity scores correspond to percentages and, in this case, 70% similarity was measured. Essentially, the alineR output of a *distance score* was converted to a *similarity score*. This rescaling was necessary in order to get more intuitive edge strengths for the phonological networks: thicker edges show a closer relationship between two words, thinner edges indicate a more distant relationship. The phonetic distance between phonological neighbors is valuable in that it can provide finer-grained distance measurements that play a role in linguistic processing (Goldinger et al., 1989).

3.1.2. L1 British English (BNC) network

The British National Corpus (or BNC, 2007) consists of approximately 100 million words collected from written and spoken language samples, specifically designed to represent a wide cross-section of British English from the latter part of the 20^{th} century. Its latest edition was released in 2007. Data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

A word list containing all words in the BNC was downloaded and further processed in an identical way as described for the learner corpora. The word list was manually sorted, removing abbreviations, duplicates etc. The final word count was 5376. The L1 language user vocabulary was smaller than that of the learners at the C1 and C2 levels (C1=5632 words, C2=6714 words), and this is most likely due to the fact that a large portion of the learner data stems from specific vocabulary tests, whereas the BNC data of the L1 British speakers was gathered from naturally occurring speech and texts.

As a next step, phonological transcriptions and X-SAMPA transcriptions were created, lexical frequency rates were calculated and log-transformed, phonological length per word was counted, and phonotactic probabilities (biphones) were computed (see details on these steps in section 3.1.1.). With the use of an Oracle database, phonological neighbors of one-segment edit distances were calculated (data provided in Appendix B), and phonological distances were determined with the help of the alineR package in R (see section 3.1.1.).

3.2. Network construction

Networks of learners' lexica and the L1 British users' lexicon were constructed with *Gephi*, an open-source software for network and graph analysis, widely used in network analysis in sociology, biology, and genomics (Bastian, Heymann, & Jacomy, 2009) and the R packages "qgraph" and "igraph" (Csardi & Nepusz, 2006; Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). Separate networks were created for each proficiency level (referred to as "A1 network", "B2 network" etc., up to "C2 network") and for the L1 British English data from the BNC ("BNC network"). In the undirected graphs, words were entered as nodes, and weighted edges connected phonologically neighboring words according to their phonological distance. With Gephi, the following statistical details were calculated for each network:

- network size (number of nodes)
- number of edges
  - the β coefficient was manually calculated by dividing the number of edges by the number of nodes, once for the whole network and once for the giant component
- size of the giant component (number of nodes and edges, percentage of overall network size)
- number and size of the islands
- number of singleton nodes
- average path length ℓ (whole network, giant component)
- clustering coefficient per node
- degree centrality per node
- weighted degree centrality per node
- closeness centrality per node
- betweenness centrality per node
- eigenvector centrality per node

Using the R package "igraph" (Csardi & Nepusz, 2006), the following variables were calculated:

- network diameter
- Pearson’s assortativity coefficient
- global transitivity of network

Small-world coefficients were calculated for each of the seven networks with the function “smallworldness” of the R package “qgraph” (Epskamp et al., 2012), taking into account the global transitivity of the network (Newman, 2003b), with the iterations set to “1000” per network.

Community detection in the learner networks was done with the “Modularity” function in *Gephi*. First, modules were created with the resolution set to “2” (as suggested by Siew, 2013, for phonological networks) and the inclusion of edge weights. Singleton nodes were excluded from the analysis, as each of them constitutes their own community. Modularity values (Q) were obtained, and each node was assigned a modularity class. These modularity classes were then extracted to spread sheets, where their sizes and the average values of the investigated network measures (degree, clustering coefficient, etc.) were calculated.

Generally, network comparisons are hampered by the fact that network size influences centrality measures. Closeness centrality, for instance, is highly dependent on network size as larger networks tend to have lower centrality measures due to the fact that larger networks have more connections that need to be traversed to get from one node to another. Freeman (1979) suggested normalizing centrality values in order to reduce the effect of network sizes. For the statistical calculations, all calculated measures were normalized (z-scored) in order to establish comparability between the differently sized learner networks.

3.2.1. Distribution fitting

Distributions of node degree, preferential attachment values, node fitness, and most other network measures are crucial considerations in the network sciences (see debate in Broido & Clauset, 2019; Zhou, Meng, & Stanley, 2020). Especially power law distributions are frequently the subject of network investigations, as they represent specific assumptions about network structure in general and growth possibilities in particular (see section 4.1.1.1.). In such distributions, the mathematical relationship between variables is such that changes in one variable lead to proportional relative changes in the other variable, independent of the initial size of the variable quantities (Clauset, Shalizi, & Newman, 2009). Power laws are mathematically represented with the function

$$y=cx^{-\alpha}$$

with α being the power law exponent, and *c* an overall scale (or normalization). With changes in *x*, *y* either decreases or increases. When power laws are calculated for distributions, the exponent α is positive, and *y* decreases as *x* increases (Milojević, 2010). This means that events with a high value of a quantity are typically rare. Power law distributions can represent what is known as the 'Pareto principle' or the '80:20 rule' stating that 20% of causes lead to 80% of phenomena (Sanders, 1987). Perline (2005) suggested three different variants of power law distributions: (1) a strong power law, with all values of a variable falling into the category of a power law, (2) a weak power law, where only a part of the distribution follows a power law, and (3) a false power law, where a truncated part of the distribution can mimic a power law (such as in log-normal distributions, where only the upper tail approximates a power law).

The probability of a distribution following a power law is calculated by comparing distribution probabilities involving other heavy-tailed distribution types (e.g., discrete log-normal, discrete exponential) and selecting the best fit based on comparative tests. The procedure of distribution fitting in the present study started with an investigation of good candidates among a predefined set of heavy-tailed distributions, including log-normal, log-logistic, log-gamma, weibull, exponential, Pareto, and Burr. Using the R packages "fitdistrplus" (Delignette-Muller & Dutang, 2015) and "actuar" (specifically for log-logistic, Pareto, and Burr fitting) and the "plotdist" function, plots of the empirical cumulative distribution function (CDF plot) and the histogram on a density scale were obtained per investigated network (i.e., proficiency level networks and BNC network). In addition to these classical goodness-of-fit plots, a Q-Q plot emphasizing a lack-of-fit at the distribution tails and a P-P plot emphasizing a lack-of-fit at the distribution center were plotted per network. Next, different goodness-of-fit statistics were calculated with "fitdistrplus" in order to compare the fit of best-seeming candidate distributions as judged by the plots. Goodness-of-fit statistics measure the distance between a fitted parametric distribution and the empirical distribution (between the fitted cumulative distribution function F and the empirical distribution function F_n). The following goodness-of-fit statistics were calculated with the R function "gofstat": Cramer-von Mises, Kolmogorov-Smirnov, and Anderson-Darling statistics (as recommended by D'Agostino & Stephens, 1986).

Distributions were compared by classical penalized criteria based on the log-likelihood (Akaike Information Criterion or AIC and Bayesian Information Criterion or BIC), with lower values indicating a preference for a particular distribution (Delignette-Muller & Dutang, 2015). The smallest values per goodness-of-fit statistic and goodness-of-fit criteria (AIC, BIC) were selected as the best fitting distribution for a given network.

As power laws can be difficult to mathematically quantify due to the fluctuations in the heavy tail and the fact that not all values of a variable x may follow a power law, the R package "poweRlaw" specifically designed to detect power laws and discriminate them from other, similar heavy-tail distributions (specifically for continuous data: power law with cutoff, exponential, stretched exponential, and log-normal; for discrete data: Yule, exponential, Poisson) was additionally used. The package is based on Clauset and colleagues' (2009) approach, who propose a rigorous power law validation procedure with maximum-likelihood fitting methods and goodness-of-fit tests based on the Kolmogorov-Smirnov statistic and likelihood ratios (see Gillespie, 2014, for details). A Monte-Carlo bootstrapping p-value is then calculated to evaluate the plausibility of a power law; and lastly, the power law model is compared to a set of alternative models. In practice, heavy-tailed distributions as identified by the first distribution fitting procedure (with "fitdistrplus", see above) were further subjected to specific power law calculations with "poweRlaw". If power law was the best-fitting of the distributions compared by the package, a power law distribution was assumed for the data. This power law fitting procedure has increasingly been used in linguistic literature on statistical distributions in recent years (see,e.g. Macklin-Cordes & Round, 2020).

When histograms indicated a bi- or multimodal distribution, the R package "diptest" was used, which utilizes Hartigan's dip test to determine whether a distribution has more than one mode (see Maechler, 2021). P-values of less than 0.05 indicate significant bimodality, and values greater than 0.05 but less than 0.1 indicate bimodality with marginal significance (Hartigan & Hartigan, 1985).

3.3. Macro-level network measures

The learner networks and the BNC network differ in quite a few of the central network measures. In some regards, the C2 and BNC networks are similar (e.g., assortativity coefficient, average degree in giant component, transitivity and β of the giant component), in others the BNC network stands out and does not resemble any of the learner networks (e.g., giant component size). Table 2 summarizes the main macro level network measures that were calculated for the learner and BNC networks.

Table 2: Overview of network measures in the ESL and BNC networks.

	A1	A2	B1	B2	C1	C2	BNC
Network size (number of words)	606	1483	2901	4621	5632	6714	5378
Connectivity (number of edges)	583	1784	3748	5889	6717	8129	4866
Giant Component size (%)	44%	42%	38%	35%	31%	30%	13.2%
Pearson assortativity coefficient r	0.57	0.59	0.65	0.67	0.68	0.68	0.71
Average degree	1.9	2.4	2.6	2.5	2.4	2.4	1.8
Average degree in GC	4.1	5.3	6.2	6.7	6.7	7	7.2
Edges in GC (%)	94%	93%	92%	90%	89%	88%	52.3%
Average path length	7.42	6.03	5.9	5.9	6	6	5.1
Clustering coefficient	0.42	0.37	0.35	0.34	0.34	0.34	0.38
Ratio of edges to vertices/β	0.96	1.2	1.3	1.3	1.2	1.2	0.9
Ratio of edges to vertices/β (GC)	2.04	2.6	3.1	3.3	3.4	3.5	3.6

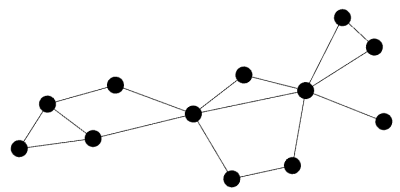
3.3.1. Network connectivity

The A1 lexicon consists of 606 nodes (i.e., words) and 583 edges (i.e., links between words), which equals a beta (β) value of 0.96. This represents a rather sparsely connected language network, as compared to the adult native L1 American English network, which has a β value of 1.61 in Arbesman, Strogatz, and Vitevitch (2010). 94% of all edges can be found in the A1 giant component, and β of the giant component is 2.04 (in adult L1 English β= 4.55, Arbesman et al., 2010). Word increase per lexicon is at first accompanied by higher connectivity between the nodes (β increase from A1 to A2 to B1) but connectivity trails off at later proficiency stages (β decrease/plateau from B2 to C1 to C2). The β value of the giant component increases continuously and proportionally, showing a steady development from A1 to C2. Number of edges found in the giant component decreases as the networks grow. Compared to the L1 American English network described by Arbesman et al. (2010), in which β=4.55, the learner networks show substantially lower β values in the giant components. The BNC giant component contains 52.3% of all network edges, with a β value of 3.6, the highest of all seven networks.

The high β index of 1.6 calculated by Arbesman and colleagues (2010) for American English and that of 4.55 for the giant component is not reached by even the highest ESL learner proficiency level, but the BNC network comes close. This shows that tight interlinking of a high number of phonologically neighboring words is much more pronounced in networks of L1 English users (American and British), with L2 English networks being more sparsely interlinked, even absent differences in lexicon size (note that the BNC network is smaller than the C1 and C2 networks). The giant components of the L1 English networks are especially tightly knit units, a dynamic possibly resulting from their smaller sizes.

3.3.2. Giant components, lexical islands, and singleton words

Like most networks, phonological networks can be divided into discrete network parts: the giant component as the largest cluster of linked entities, the smaller linked entities called islands, and non-linked stand-alone phonological isolate words. Of these, the giant component plays an outsized role as it represents an expanded phonological neighborhood where all words are connected through multiple degrees of phonological neighbors. Studies have proven that phonological activation spreads not only in the immediate one-segment-distance phonological neighborhood but reaches farther, potentially activating neighborhoods multiple phonological segments away (Chan & Vitevitch, 2009; Suarez et al., 2011). Therefore, the density of connections in the giant component of a lexical network, the overall diameter of the giant component (i.e., geodesic distance between farthest points), as well as the average phonological neighbor per word (=degree) in the giant component are crucial measures that can give insights into patterns of activation spreading of phonological word forms in the lexical processing of language learners. Figure 13a-b depicts two same-sized networks differing by diameter, degree centrality, and β values. In 13a activation can spread quickly through the entire network, encompassing a larger number of nodes in a short time, whereas activation spreading in 13b involves more steps to reach each node.



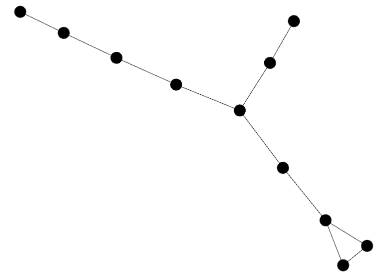


Figure 13a-b: Two same-sized networks differing by degree centrality, diameter, and β coefficient.

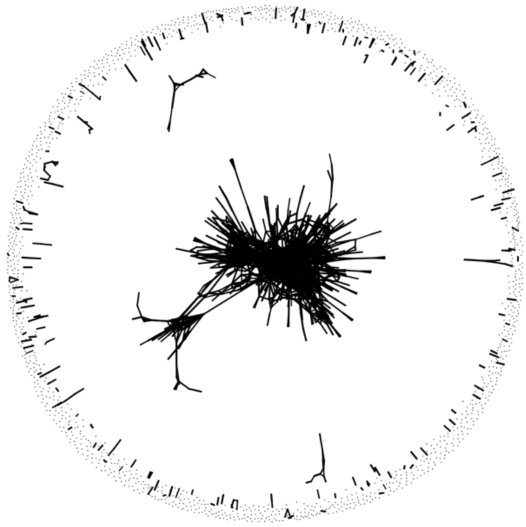
Giant component size is largest in the A1 lexicon (=44%), with an additional 25 islands (=10%) and 279 singleton nodes (=46%) in the network (see Figure 14a). Islands are small and range from 2 to 4 nodes with 1 to 3 edges. In the A2 network, the giant component size has shrunk to 42%, with 77 islands (=13%) and 664 (=45%) singleton nodes. Island size and connectivity have increased and range from 2 to 6 nodes, with 1 to 6 edges per island. The giant component of the B1 network equals 38%, with 162 islands (=15%) and 1371 (=47%) singleton nodes in the network. Islands range from 2 to 18 nodes, involving 1 to 7 edges (see Figure 14b). In the B2 network, the giant component makes up 35% of the network, and an additional 328 islands (=18%) and 2193 (=47%) singleton nodes can be found. Islands sizes range from 2 to 26 nodes, involving 1 to 37 edges. The C1 giant component consists of 31% of all nodes, and 431 islands (=21%) and 2751 (=48%) singleton nodes are included in the network. Average island size has markedly increased, ranging from 2 to 70 nodes with 1 to 88 edges. At the C2 level, the giant component proportion shrinks further and constitutes only 30% of the overall network. The proportion of islands and singletons has increased (527 islands =21%; 3295 singleton nodes =49%, see Figure 14c). Islands sizes and connectivity range from 2 to 45 nodes involving 1 to 52 edges. In the BNC network, 515 islands and 2791 singleton nodes (=52%) were detected; islands range from 2 to 175 nodes (see Figure 14d, additional graphs of the learner networks can be found in Appendix A.

a

A1: 606 words, 25 islands, 279 singletons

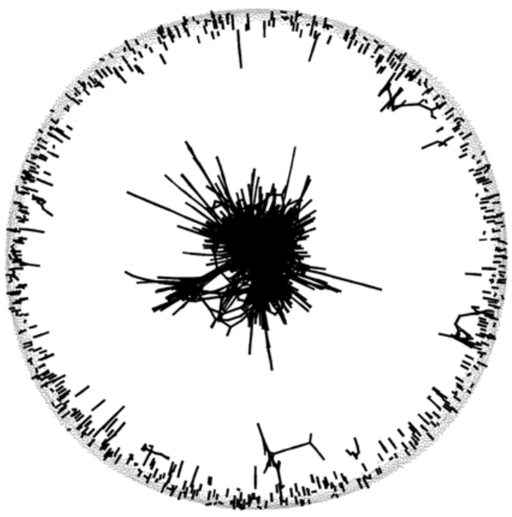
b

B1: 2901 words, 162 islands, 1371 singletons



c

C2: 6714 words, 527 islands, 3295 singletons



d

BNC: 5376 words, 515 islands, 2791 singletons

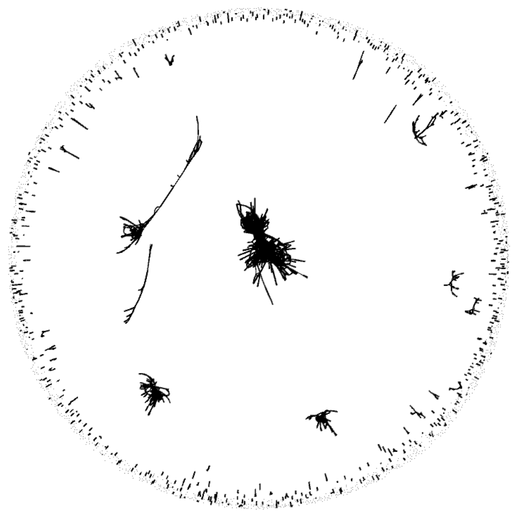


Figure 14a-d: Yifan Hu visualization (from the Gephi Network Project) of the A1 (a), B1 (b), C2 (c), and BNC (d) phonological networks. The giant component is central and surrounded by (lexical) islands and singleton nodes (i.e., lexical hermits).

The giant component proportions within the respective learner networks shrink as proficiency increases from its highest percentage of 44% in the A1 network to only 30% in the C2 network. A study on giant component sizes in various language by Arbesman and colleagues (2010) calculated the giant component of American English at approximately 34%, which was the smallest of all languages measured. For instance, Spanish and Basque had giant component sizes of 37% and 35%, respectively; Mandarin Chinese even reached 66% (see Arbesman et al., 2010). In the present study, the BNC giant component amounted to only 13.2% of the network. One important caveat complicates a comparison of the findings of the present study and those of Arbesman and colleagues. The Arbesman et al. analyses were not based on corpus data, but the authors constructed phonological networks from dictionaries of the languages, in the case of English the *Merriam-Webster Pocket Dictionary* from 1964, a database frequently used in psycholinguistics, containing 19,323 words. While their study design is not directly comparable to the usage-based account of the present study, it is interesting to note that the giant component sizes of the learner networks are quite similar to those measured by Arbesman and colleagues. The findings of the present study clearly support the hypothesis of small giant component sizes of English phonological networks, further underscored by the very small giant component of the BNC network of first-language users of British English.

Most social, technological, and biological networks can combine an extraordinarily large amount of nodes in a large giant component (up to 80-90%, see Newman, 2001). Several disadvantages are associated with large giant component sizes, chief among them low network robustness (Hu & Lee, 2020, see section 3.2.2.). Small giant component sizes predict better resilience against accidental failures and targeted attacks on a network, and the phonological networks may be particularly well structured to make them maximally robust (Siew & Vitevitch, 2016; Vitevitch, 2008). Arbesman and colleagues (2010) found the phonological network of several languages to be highly resistant to network failures, primarily caused by small giant component sizes paired with large numbers of islands and singletons. New words preferentially attaching to lexical islands and singletons, rather than contributing to the giant component, are predicted to help maintain network stability (Siew & Vitevitch, 2016; Zhao & Xu, 2009). Such a dynamic keeps constant the structure and size of the giant component, where the bulk of lexical processing occurs, all the while promoting growth in the overall network by supplementing smaller components (islands, singletons) with new words, without drastically influencing the overall functioning of the lexical network (Siew & Vitevitch, 2016).

The A1 level houses 44% of all words in the giant component, the highest rate of all learner networks. This indicates a stronger tendency for phonological clustering at the initial stage of word learning. Studies have shown that phonological similarity greatly facilitates word learning in children (Hoover, Storkel, & Hogan, 2010; Storkel & Lee, 2011) and this may be particularly pronounced when fewer words are involved and at beginning learning stages (be it child language acquisition or foreign language learning). While phonological clustering is most pronounced in the A1 network, already at the A2 level, the giant component of the learner networks is starting to shrink, while the islands grow at a continuous pace to the C2 level. From a macro level perspective, network growth takes place primarily in the islands, which grow at each level, starting at a proportion of 10% within the A1 network and increasing to 21% of the C2 network. This is in agreement with findings by Siew and Vitevitch (2020a) in relation to network growth during child language acquisition in American English and Dutch, which will be described in more detail in chapter 5.

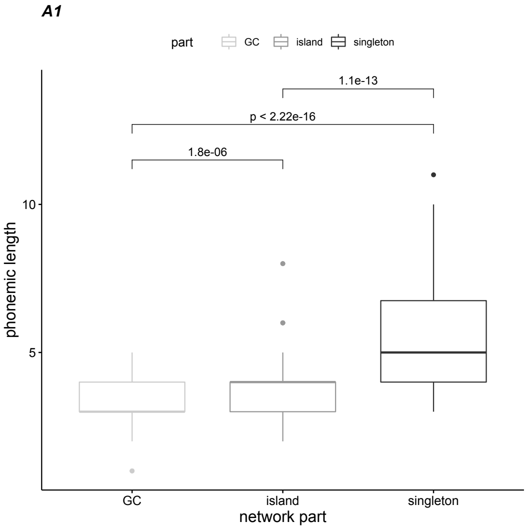
The proportion of singleton, non-connected words is rather stable across all learner levels, ranging between 45% and 49%. Almost half of the learner network at each respective level consists of singleton nodes, indicating a propensity to learn words even in the case of zero phonological neighbors. A phonological network of American English described by Vitevitch (2008) also showed a large proportion of singletons (lexical hermit) words: 53% of the investigated words (see Gerometta, 2015, for similar findings). In the BNC network the proportion of singleton nodes was 52% of the network.

Th network part to which a word belongs has consequences for its lexical access. The processing disadvantage of giant component words has been described in the diffusion activation context (see section 2.3.). The results of the present study demonstrate a shrinking of giant component proportion with each successive proficiency level, possibly reflecting a cognitive development geared toward increasing processing advantages and minimizing lexical activation competition (which is most pronounced in the giant component). During the build-up of the ESL lexicon, network resources (i.e., newly learned words) are gradually shifted more toward peripheral parts of the network, thereby avoiding concentration in the giant component. Due to activation diffusion, smaller giant component proportions could be an adaptation to improved lexical processing: with each successive proficiency level, a larger number of words needs to be navigated in lexical search, which places a burden on processing efficiency. Larger vocabularies need better organization for efficient navigation, and smaller giant component sizes may facilitate that.

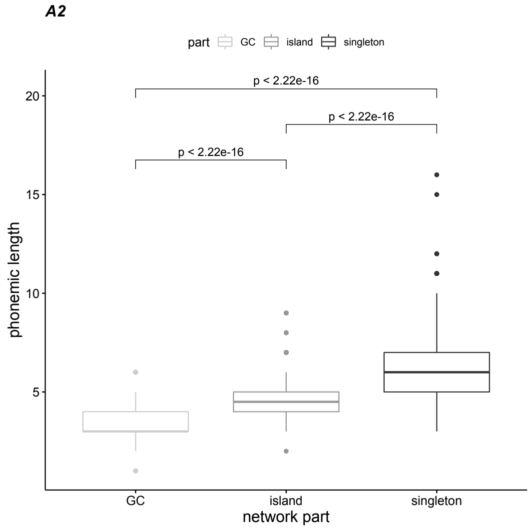
3.3.3. Lexical characteristics of network parts

The results of the present study confirmed a core finding by Siew (2013) for an L1 American English network: shorter words have a higher likelihood of residing in the giant component. In all learner networks, phonemic length of words was lowest in the giant components and longest in singleton nodes (see Figures 15a-f; tests reported for the learned networks: *Kruskal-Wallis one-way ANOVA on ranks with post-hoc Wilcoxon signed rank tests*; A1: $\chi^2(2)=313.6$, $p<0.001$; A2: $\chi^2(2)=786.8$, $p<0.001$, B1: $\chi^2(2)=1576.2$, $p<0.001$, B2: $\chi^2(2)=2428.5$, $p<0.001$, C1: $\chi^2(2)=2837.5$, $p<0.001$, C2: $\chi^2(2)=3413.4$, $p<0.001$).

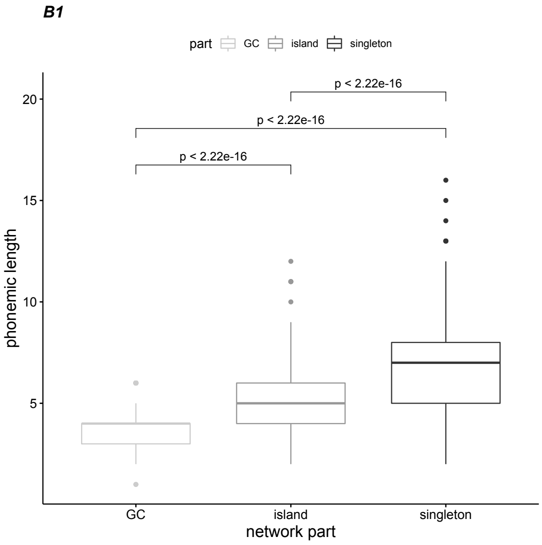
a



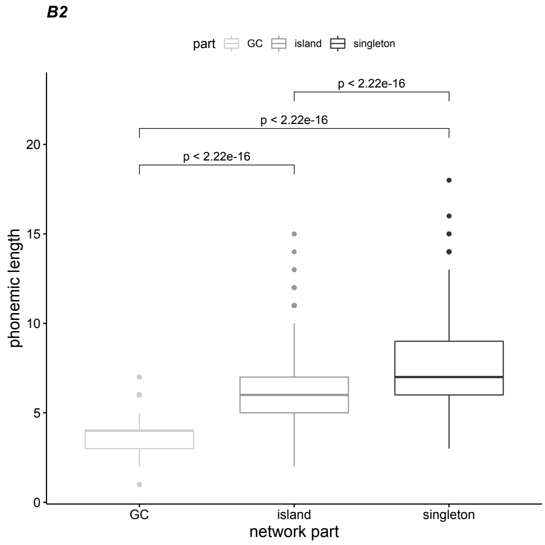
b



c

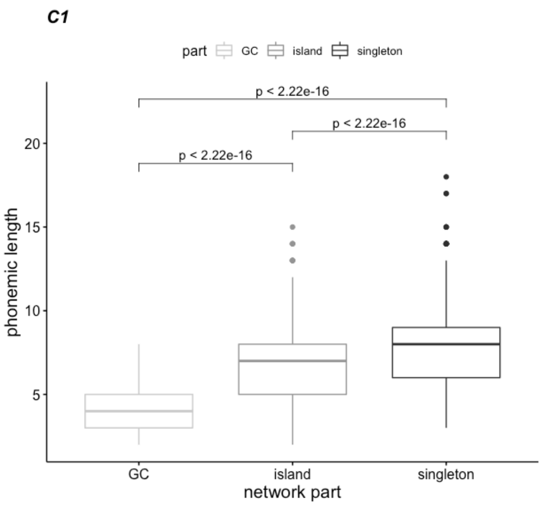


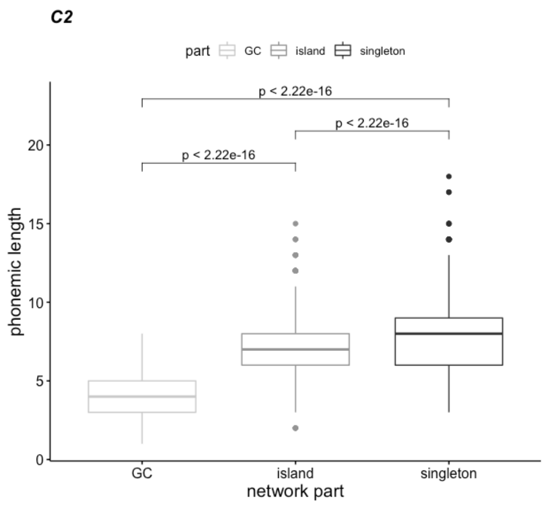
d



e

f

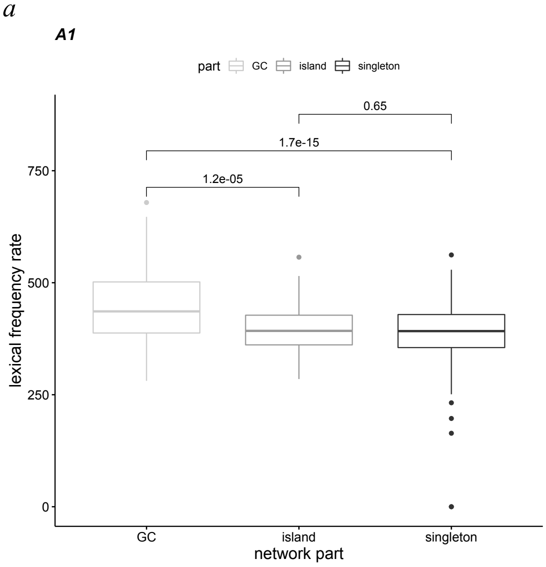




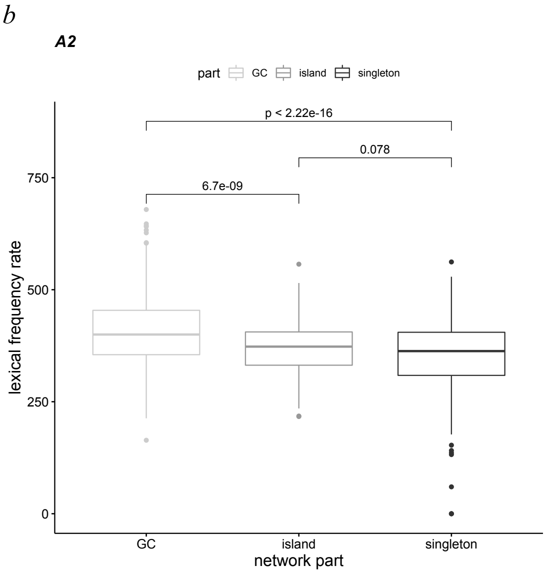
Figures 15a-f: Phonemic length of words belonging to different network parts (giant component, island, singleton) in the six learner networks. Post-hoc Wilcoxon signed rank test results indicated.

Additionally, words of higher lexical frequency tended to cluster together in the giant component of the learner networks (also see Siew, 2013; frequency log-transformed, A1: $\chi^2(2)=67.6$, $p<0.001$, A2: $\chi^2(2)=118.6$, $p<0.001$, B1: $\chi^2(2)=150.13$, $p<0.001$, B2: $\chi^2(2)=227.38$, $p<0.001$, C1: $\chi^2(2)=329.7$, $p<0.001$, C2: $\chi^2(2)=402.7$, $p<0.001$). Figures 16a-f show the lexical frequency differences between network parts across the six learner networks.

a



b



c

d

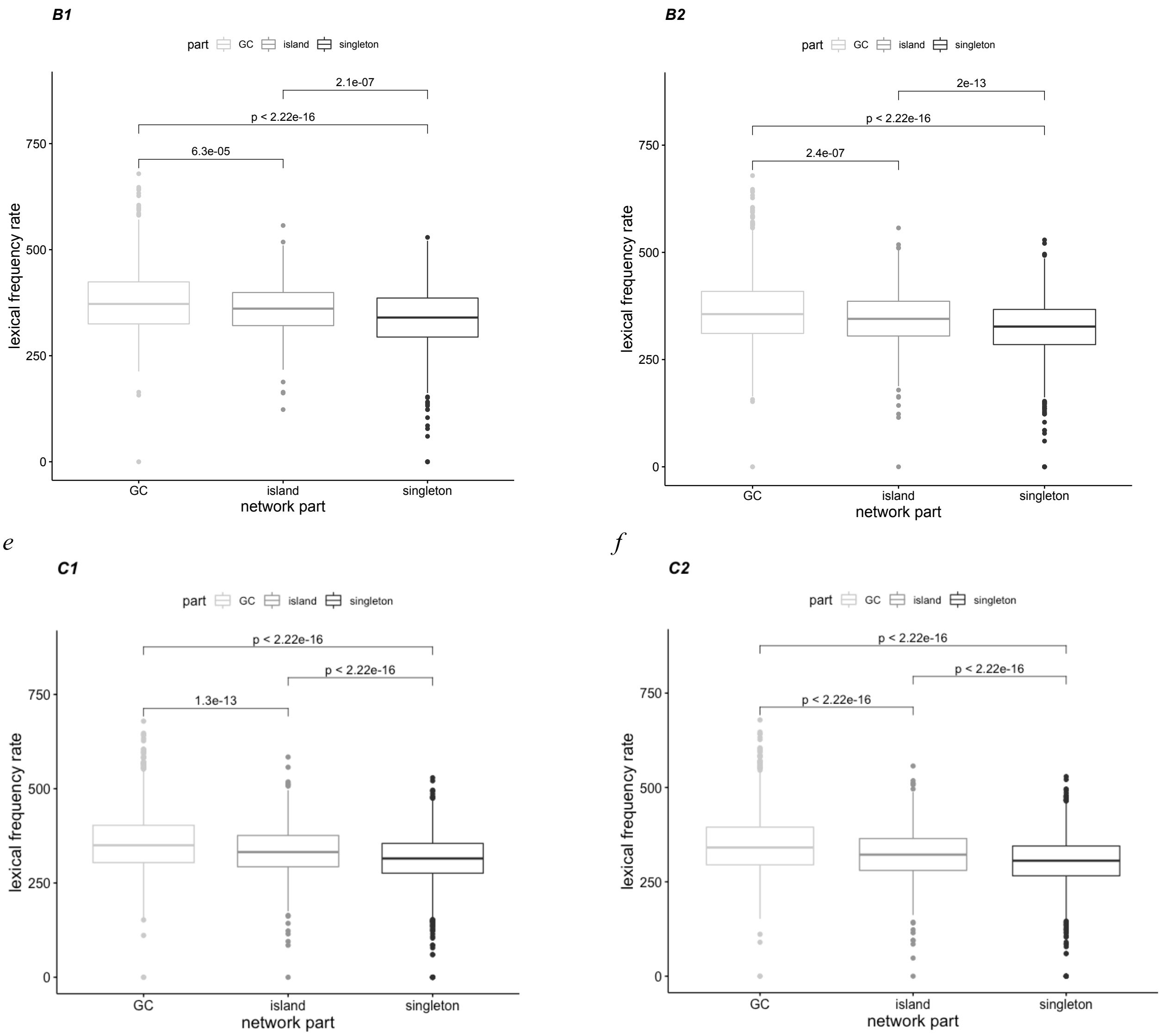


Figure 16a-f: Lexical frequency rates of words belonging to different network parts (giant component, island, singleton) in the six learner networks. Post-hoc Wilcoxon signed rank test results indicated.

Phonotactic probability was generally lowest in the giant component words, with islands and singletons containing higher-probability phonemic combinations on average (log-transformed biphone probability, A1: $\chi^2(2)=11.5$, $p=0.003$, A2: $\chi^2(2)=30.1$, $p<0.001$, B1: $\chi^2(2)=75.4$, $p<0.001$, B2: $\chi^2(2)=120.3$, $p<0.001$, C1: $\chi^2(2)=142.8$, $p<0.001$, C2: $\chi^2(2)=199.7$, $p<0.00$). See Figures 17a-f for details. In the A1 network, singleton nodes stand out at significantly higher phonotactic probability as compared to giant component and island words. Starting at the A2 level, the giant component develops a statistically outstanding role, with the lexical islands and singletons comprising the majority of high-probability words. In the C2 network, all three network parts differ from one another.

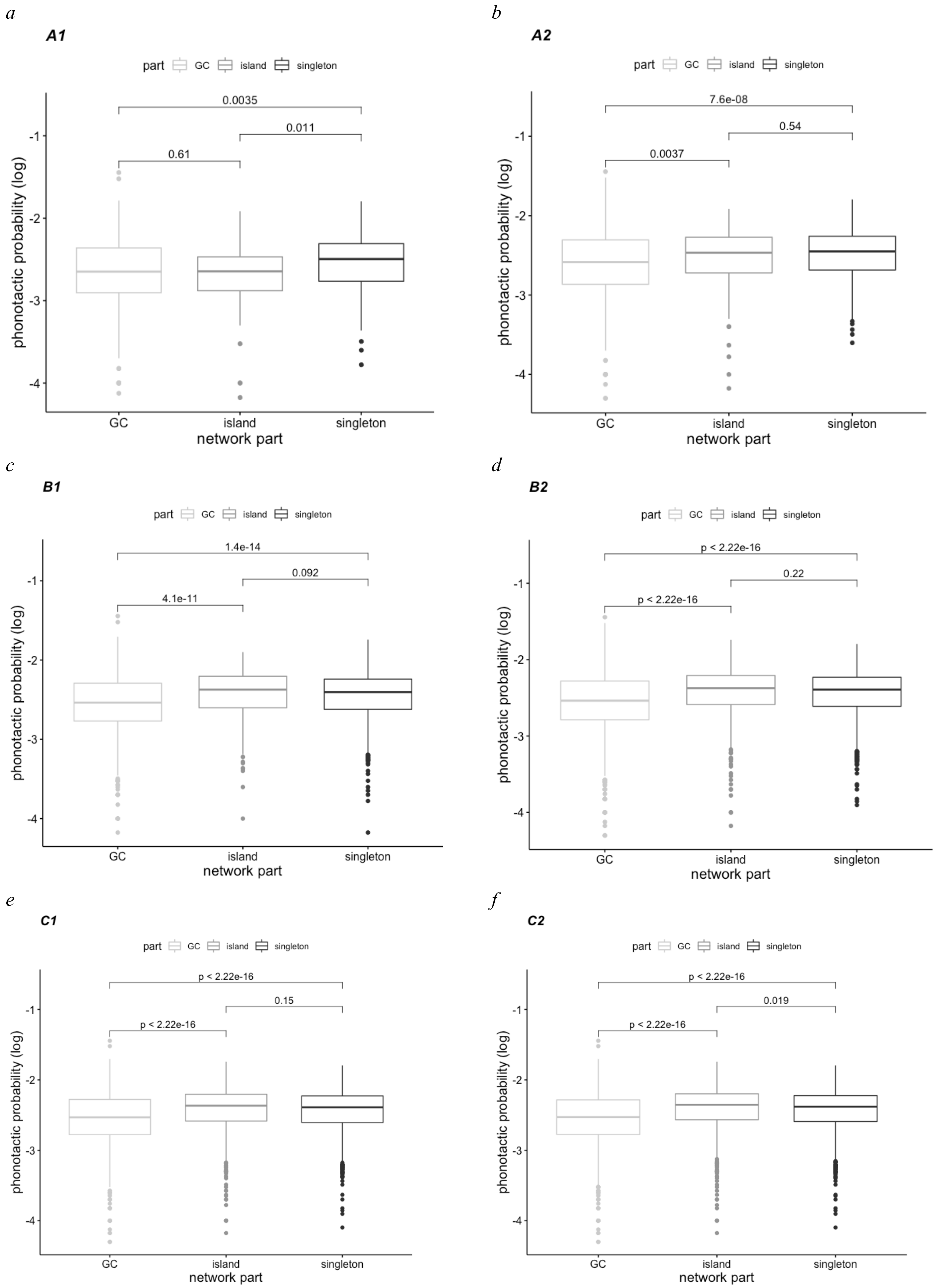


Figure 17a-f: Phonotactic probability of words belonging to different network parts (giant component, island, singleton) in the six learner networks. Post-hoc Wilcoxon signed rank test results indicated.

3.3.4. Network of L1 British English

Similar to the learner networks, phonemic length of words was highest in singleton nodes and lowest in the giant component (Kruskal-Wallis: $\chi^2(2)=1912.5$, $p<0.001$; see Figure 18).

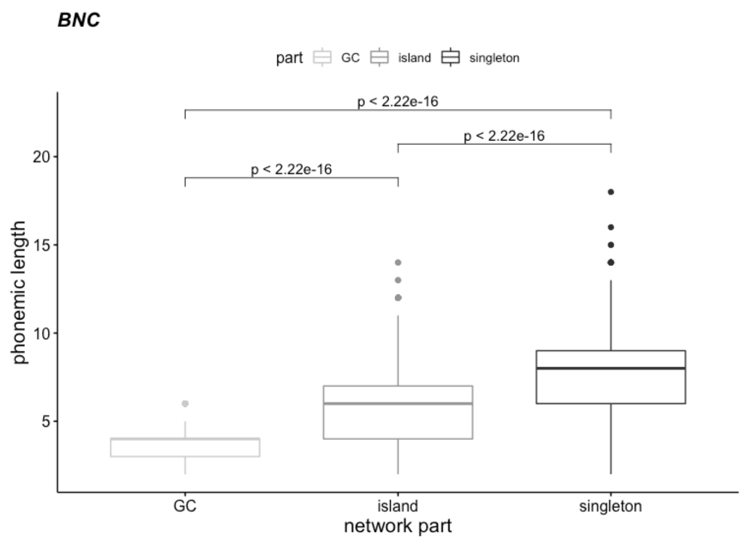


Figure 18: Phonemic length is shortest in giant component words and longest in singleton nodes.

The BNC giant component contained more high-frequency words than the island or singleton components (Kruskal-Wallis: $\chi^2(2)=156.4$, $p<0.001$), mirroring findings from the learner networks (see Figure 19).

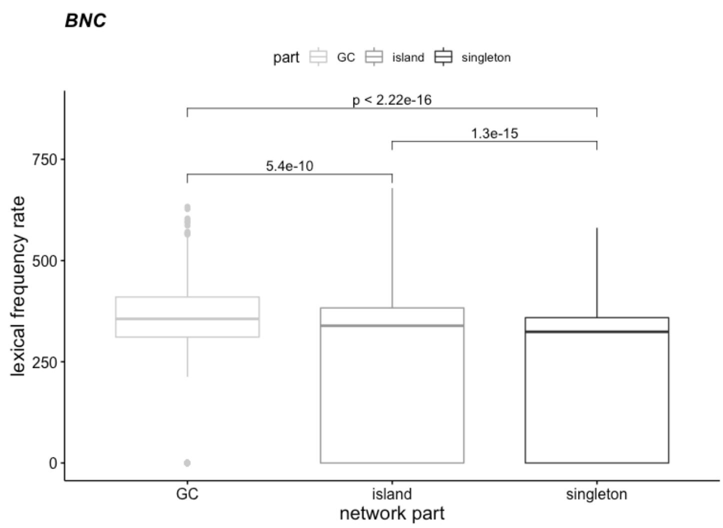


Figure 19: Lexical frequency rate in the different network parts of the BNC network.

Phonotactic probability was lowest in the giant component but highest in the singleton nodes (Kruskal-Wallis: $\chi^2(2)=812.3$, $p<0.001$; see Figure 20).

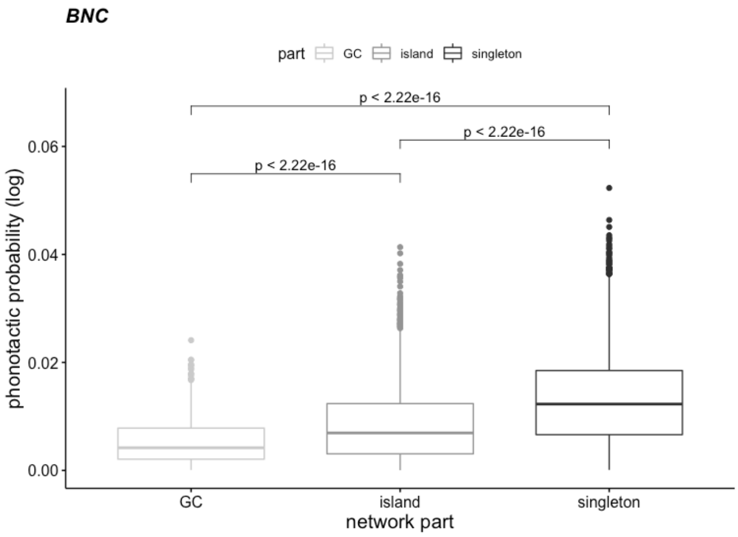


Figure 20: Phonotactic probability is lowest in giant component words and highest in singleton nodes.

In sum, the BNC network and the learner networks show identical dynamics with regard to network parts and the distribution of lexical characteristics. The findings are also in agreement with results reported by Siew (2013) for L1 American English: shorter words of high lexical frequency rate and low phonotactic probability possess a clear learning advantage and are acquired earlier in first and second language learning (also see, e.g., Ellis, 2002; Storkel et al., 2006). The fact that these words tend to be found in the giant component of all learner networks highlights the importance of giant component formation in word learning. As can be seen in the topological development of the learner networks, reliance on giant components decreases over time, while reliance on lexical islands increases.

3.3.5. Assortativity by degree

As described in section 2.2.1.2., assortativity by degree is the tendency of nodes to connect to nodes with similar degree centrality, thus highly connected nodes link up with other highly connected nodes (Newman, 2002a). Assortativity by degree increased as the learner networks grew from A1 to C2. Positive assortative mixing in phonological networks is recorded when $r>0.5$ and all learner networks exceeded that value. Preferential neighborhood formation by degree makes it less likely that sparse neighborhoods can become linked to dense ones, in other words the neighborhood distribution will inevitably be skewed. Arbesman et al. (2010) report r=0.65 for their L1 American English network, which represents a high positive degree correlation between neighboring nodes (as compared to social networks with r commonly ranging from 0.1 to 0.3). Their analysis of Spanish even yielded r=0.76. The ESL learner networks exhibited lower assortativity values (ranging from 0.57 to 0.68, see Table 2) but with a clear tendency to increase degree assortativity with network growth. The L1 British English network yielded a high r-value of 0.71, while the C1 and C2 proficiency levels reached r=0.68, thus exceeding the Arbesman estimation for English.

Figure 21 displays assortativity patterns in the A1 network, with Pearson's assortativity coefficient r equaling 0.57 and average node degree across the network being 1.9; average node degree in the giant component was 4.1. The A1 network is rather small and the giant component contained 44% of the 606 total nodes. Clusters of nodes with similar degrees are clearly visible.

Figure 21: Assortativity in the giant component of the A1 network (hub size and color by degree).

As r increased to 0.59 in the A2 network, average node degree rose to 2.4 and in the giant component to 5.3. The B1 and B2 networks showed average node degrees of 2.6 and 2.5 across the whole networks and 6.2 and 6.7 in the giant components; assortativity coefficients equalled 0.65 and 0.67, respectively. At the C1 level, r equalled 0.68, with the average node degree being 2.4 across the network and 6.7 in the giant component. The average degree per node was

2.4 across the C2 network and 7 in the giant component, while r amounted to 0.68 (see Figure 22).

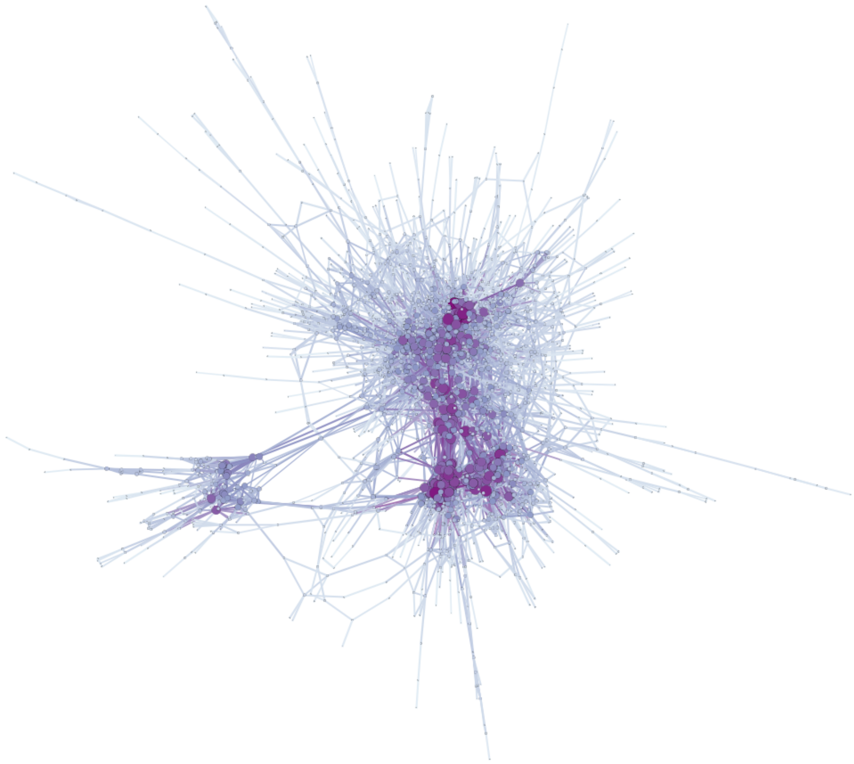


Figure 22: Assortativity in the giant component of the C2 network (hub size and color by degree).

The BNC network was characterized by r=0.71, with the average node degree being 1.8 across the whole network and 7.2 in the giant component (see Figure 23, rendering neighborhoods in the giant component denser than in the learner networks.

Figure 23: Assortativity in the BNC giant component (node size and color by degree).

As previous studies indicate (Chan & Vitevitch, 2009; Siew, 2013), denser and more tightly connected neighborhoods can impair lexical retrieval due to the higher level of activation diffusion and/or more activation competition. Continuous enrichment of already "rich" neighborhoods cannot be an open-ended process in a lexicon, as the ability of lexical recall starts to suffer at some point. On the other hand, network robustness increases with higher positive assortativity levels (Newman, 2002a). The gradual increase in assortativity in the learner networks points toward a word learning bias that favors robustness over lexical retrieval efficiency.

3.3.6. Phonological distance: Edge weight

Edge weight reflecting phonological/phonetic distances between the phonological neighbors served as indicators of connectivity in the learner and BNC networks. Words in the giant component tend to be more closely related in phonological terms, and the difference in phonological distance between the giant component and the islands increases as the learner networks grow (see Figure 24).

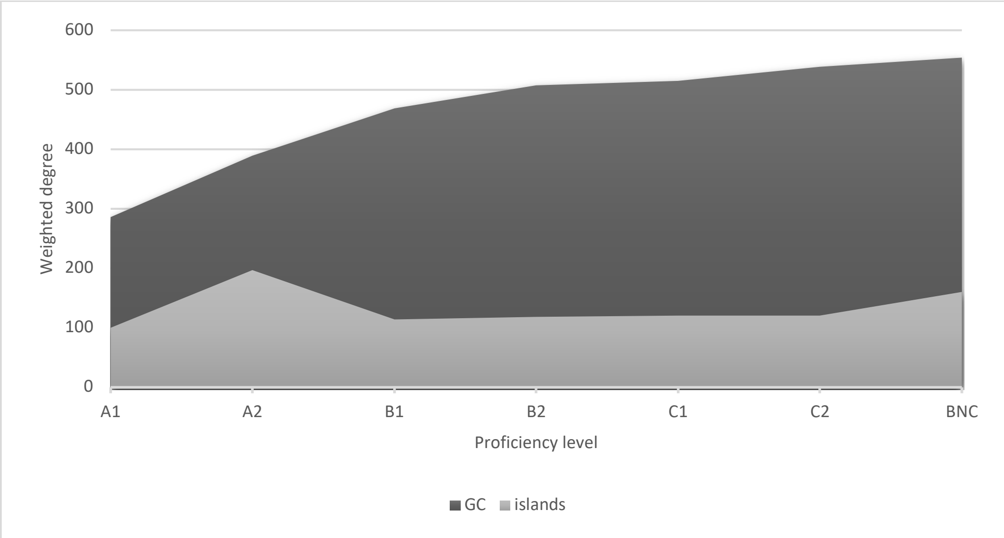


Figure 24: Edge strength (assessed through weighted degree centrality) increases in the giant component as the learner networks grow, while it stays relatively constant in the islands.

Since the giant component words were typically of higher lexical frequency rates and shorter phonemic length than words in other network parts (see above), correlations with edge weight were computed. Lexical frequency and weighted degree were not correlated at any significant degree (average Pearson's r across all learner networks <0.33). However, phonemic word length and weighted degree showed a moderate negative correlation (average across all proficiency levels r=-0.58), with shorter word length being positively correlated with higher weighted degrees. Shorter words are known to have more phonological neighbors (Bard & Shillcock, 1993; Charles-Luce & Luce, 1990; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985), with phonetic similarity playing a crucial role, as shown by the present findings. The probability of phonological neighbor creation in different segmental position – the P-metric (Johnson & Pugh, 1994) – is higher in shorter words, with the spread of the neighborhood being more widely dispersed (Vitevitch, 2007). Since short words tend to have more neighbors in general, they might be more flexible when it comes to neighbor-creation and the P-metric. In terms of activation spreading, more closely related phonological neighbors are assumed to take more of the spreading co-activation within a neighborhood. In the diffusion activation framework, this could mean that phonologically close neighbors receive more and also pass on more of the existing activation to their own neighbors, potentially strengthening the phonological representations of themselves and their connections in the network at the expense of phonologically distant neighbors of a target word. Future studies focussing on activation spreading along phonological-relatedness lines within the giant component may show that phonologically strong links share more of the overall activation among themselves, with the phonologically weak links losing out on activation sharing.

3.3.7. Diameters and average path lengths in the giant components

The learner networks showed some fluctuation in network diameter and average path length between nodes in their giant components (see Figure 25). Both measures were highest in the A1 network, and then decreased in the A2 and B1 levels, after which they continued to grow again. These data indicate a more loosely linked network at the beginning of learning which seems to develop into a more tighly linked structure as learning progresses. In the BNC network, the diameter of the giant component and average path length were comparable to the advanced learner proficiency levels.

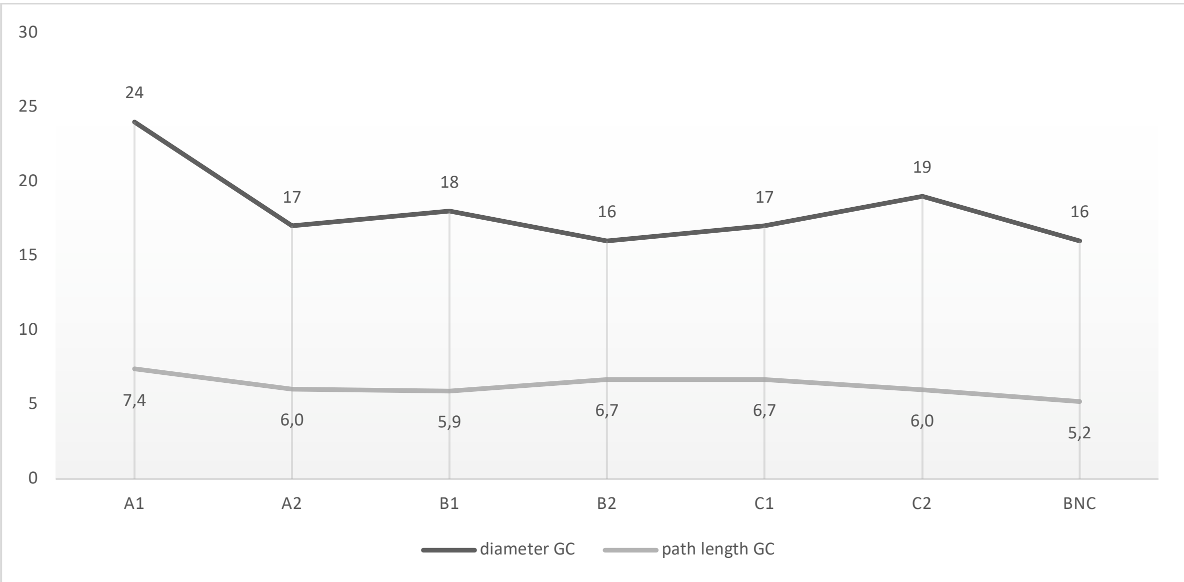


Figure 25: Average path length and diameter of giant components in the learner and BNC networks.

The longest average path length and diameter in the A1 network predisposes it to less efficient lexical processing, according to Vitevitch and colleagues (2016), underscoring the previous observation that lexical processing efficiency increases as the learner networks grow (see section 3.3.2. on findings reported on giant component size, for instance). Efficiency in terms of diffusion activation (shorter average path length, shorter diameter, smaller GC) continues to be shaped as learning progresses.

3.3.8. Centrality

The average centrality measurements yielded by the analyses are reported in Table 3. While overall rather similar, the BNC and learner networks differ markedly in betweenness centrality and closeness centrality, with the BNC network displaying lower average betweenness but higher average closeness values (see Appendix A for betweenness centrality graphs of the learner and BNC networks). This is most probably related to the unusual topology of the BNC network, where larger islands co-exist with a very small giant component.

Table 3: Mean network measures and their standard deviations across the networks.

	A1	**A2**	**B1**	**B2**	**C1**	**BNC**
Degree centrality	1.92±2.6	2.41±3.5	2.58±4.2	2.55±4.42	2.38±4.34	1.8±3.4
Weighted degree centrality	136.21±174.9	178.56±250.4	196.27±306.6	196.57±324	185.21±320.4	141.1±254
Closeness centrality	0.15±0.26	0.19±0.28	0.19±0.3	0.22±0.33	0.22±0.34	0.26±0.36
Betweenness centrality	380.35±961.7	670.6±1813.3	1040±3145	1362.19±4039.3	1391.7±4418.3	220.6±952
Clustering coefficient	0.16±0.29	0.14±0.26	0.13±0.24	0.12±0.23	0.11±0.23	0.1±0.23
Eigenvector centrality	0.046±0.13	0.047±0.12	0.043±0.13	0.032±0.11	0.029±0.09	0.03±0.09

Density analyses showed heavy-tail distributions for all network measures (see Figures 26a-f and Table 4).

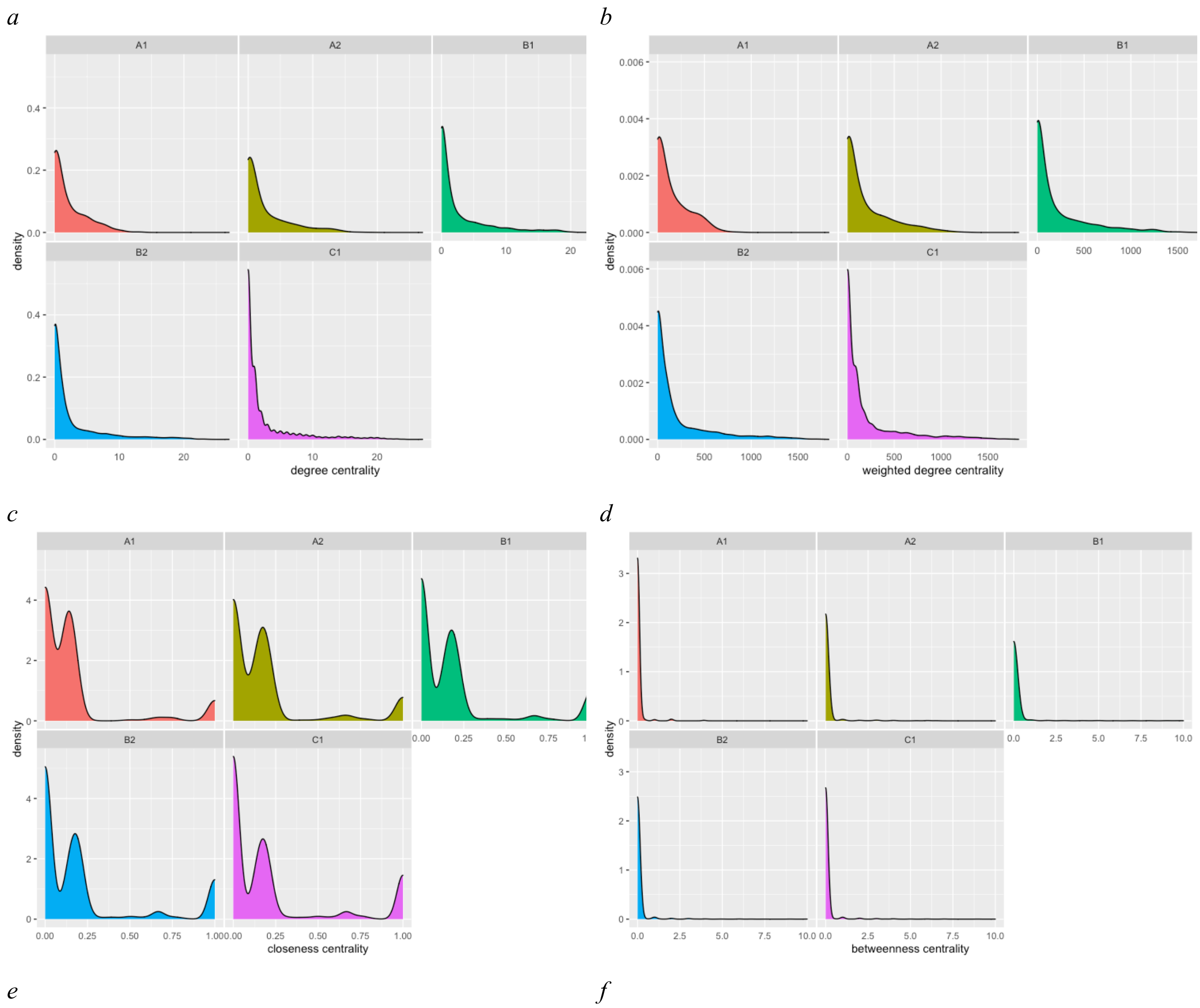
Table 4: Best-fitting distribution per network measure (across all proficiency levels).

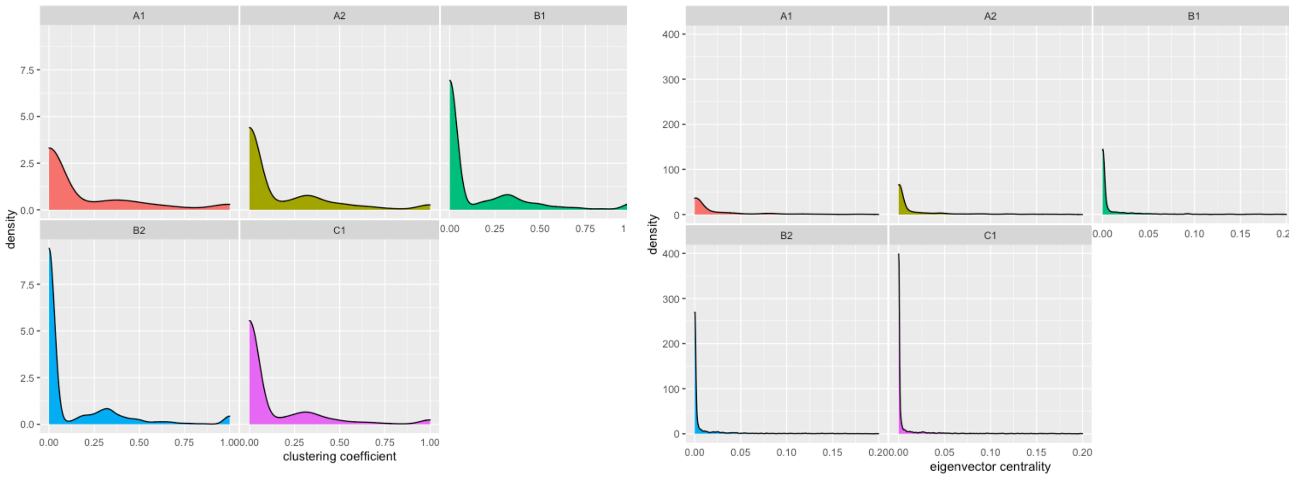
	Degree centrality	*Weighted degree c.*	*Closeness centrality*	*Betweenness centrality*	*Clustering coefficient*	*Eigenvector centrality*
Best fit	Log-normal	Power law (Pareto)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Log-logistic

As expected based on the controversy surrounding power laws in scale-free networks (see discussion in section 4.1.1.1.), a power law distribution was not observed for degree distribution in the phonological networks (also see Vitevitch, 2008, for a similar finding). Over the course of L2 development, there is an obvious shift of degree centrality towards large values in only a few prolific nodes, approximating an exponential distribution of node degree at the C1 proficiency level (see Figure 25a). The skewness of node degree gets more extreme with each proficiency level. In other words, few neighborhoods can increase their density during growth, with dense neighborhoods impeding lexical retrieval. Studies in child language acquisition found that children improved word learning when phonological neighborhoods start to emerge in their lexicon (Donnelly & Kidd, 2020). The developmental trajectory of node degree in the ESL learner networks could indicate a similar tendency, with denser neighborhoods emerging in more advanced stages of lexical learning.

A special case of neighborhood effect and language development is implied by Karimi and Diaz (2020) in their study of age-graded neighborhood density in child language acquisition. They compared lexical processing efficiency in words acquired at different acquisitional stages in L1 English-learning children and found that dense phonological neighborhoods facilitated retrieval of early-acquired words but inhibited retrieval of late-acquired words. The authors suggested that this may be related to weaker phonological representations in late-acquired words, as opposed to the more strongly ingrained early-acquired words. The level of activation in a target word in relation to the level of activation in its neighbors plays a role here: if an early-acquired target word with strong phonological representation resides in a neighborhood with late-acquired words with weak representations, the target word is hypothesized to take most of the activation in the neighborhood. While Karimi and Diaz defined phonological neighbors more broadly than the present study, their findings have implications for the learner networks studied here. L2 learning could incorporate similar age-of-acquisition features in neighborhood activation: early-learned words may have stronger phonological representations, setting them apart from their immediate neighbors to a greater extent, thereby heightening discriminability and reducing lexical competition, which could ultimately influence the neighborhood structures at different stages of language learning. On a more speculative level, lexical retrieval in the more advanced learner lexica, which include a multitude of both old and new words (early-learned and late-learned), may ultimately not be impeded by dense neighborhoods. Old and new words have phonological representations of different strengths, reducing activation competition to some extent, and allowing denser neighborhoods to be formed. At the beginning stages of language learning, the ages of learned words do not differ markedly, so that the A1 lexicon includes only newly-learned words. Here, phonological neighborhoods should conceivably trend toward more sparsity, in order to facilitate lexical retrieval by minimizing activation competition. Changing degree distributions over the course of learning (see Figure 25a) may hint toward such a dynamic in the learner networks, as they show a dominance of more numerous and moderately dense neighborhoods in the beginning networks, followed by the emergence of very few dense and many sparse neighborhoods in the higher proficiency networks.

For various network measures, it can be observed that inequality increases as the networks grow. Not only degree, but also weighted degree, clustering coefficient, and eigenvector centrality gain in skew during the course of language learning. This could be a function of increased network sizes and different mathematical probabilities relating to those differently sized networks. Alternatively, this could reflect an inherent property of complex systems: as they grow, properties change as a function of different rules emerging in the system. At the very least, these findings demonstrate that the beginner network is not just a smaller mirror image of the more advanced lexicon but is a different system operating according to its own set of rules.





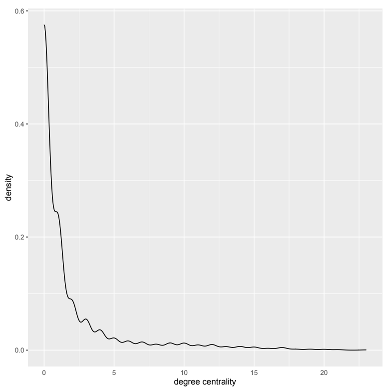
Figures 26a-f: Distributions of network measures in the learner networks.

The BNC network measures were distributed in a similar way to the learner networks and the results are displayed in Table 5 and Figures 27 a-f.

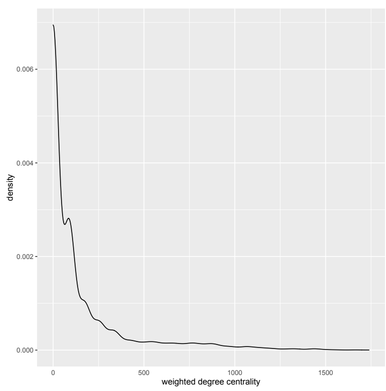
<u>Table 5</u>: Best-fitting distribution per network measure in the BNC network.

	Degree centrality	*Weighted degree c.*	*Closeness centrality*	*Betweenness centrality*	*Clustering coefficient*	*Eigenvector centrality*
Best fit	Multimodal (dip test p: <0.001)	Bimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Burr	Multimodal (dip test p: <0.001)	Log-normal

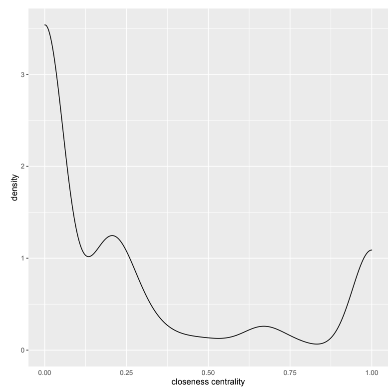
a



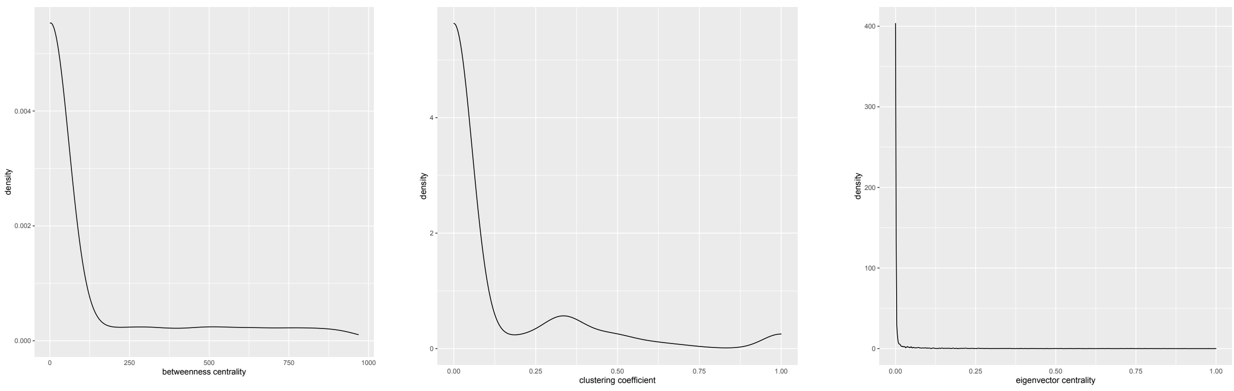
b



c



d *e* *f*



Figures 27a-f: Distributions of network measures in the BNC network.

In sum, all of the network measures in the learner and BNC networks showed heavily skewed distributions. Clustering coefficient and closeness centrality were the least skewed measures, where high values are spread to a larger number of nodes in the networks. High values in eigenvector and betweenness centrality were restricted to very few prolific nodes, as would be predicted from their roles in networks.

3.3.9. Small-worldness

All learner networks are characterized by high clustering coefficients in combination with relatively short average path lengths (i.e., shortest path length), indicating a tendency to incorporate small-world features. Even though small-worldness is notoriously hard to define (see, e.g., Telesford, Joyce, Hayasaka, Burdette, & Laurienti, 2011) and has been questioned in phonological networks (Gerometta, 2015), the concept serves some purpose in relation to lexical competition in phonological networks (Levy et al., 2021; Vitevitch, 2008). In the learner networks, a development toward more small-worldness is obvious (see Table 6). Sigma values, calculated by comparison of path length and transitivity to a random network (see section 3.2.) are commonly used to prove the existence of small-world networks when $\sigma>1$ (Humphries, Gurney, & Prescott, 2006). The learner networks start out with rather high σ values at the A1 level, which decrease in A2 and B1 but increase significantly starting at the B2 level (see Table 6). Small-worldness is rather pronounced in the upper proficiency levels, but reach even higher σ values in the BNC network. A major advantage of small-world networks is their improved searchability where inspection of a small number of interconnected nodes yields a high likelihood of success in locating a particular node in a system (Kleinberg, 2000). Lexical search in the adult and developing L1 lexicon can be greatly aided by such a structure (Arbesman et al., 2010; Carlson et al., 2014; Chan & Vitevitch, 2009).

Table 6: Small-world coefficients across the proficiency levels of the learner and BNC networks.

	A1	A2	B1	B2	C1	C2	BNC
σ	5.19	4.22	4.25	5.72	6.87	6.96	7.9

Tight interconnectedness of neighbors leads to higher rates of co-activation in a neighborhood and eventually results in discrimination problems of the target word, slowing down and making less accurate the lexical access and retrieval process (Chan & Vitevitch, 2009, 2010; Goldstein & Vitevitch, 2014; Yates, 2013). Word learning in the ESL networks could be maximally facilitated at the initial-intermediate stages (A2, B1) when small-worldness is lowest. Since the vocabulary is relatively small at these stages, lower degrees of small-worldness could have a relatively large effect on learning success. Following the Goldstein and Vitevitch proposal (2014), word learning should become more difficult when small-worldness increases in the advanced proficiency stages. At this point, the vocabularies are generally rather large and integrating new words increasingly crowds the phonological spaces in the network (especially the giant component), making lexical processing more arduous from a small-world perspective.

The learner networks showed a clear tendency to decrease transitivity with each progressive stage (see Table 7). This could be related to the increase in network size, with smaller networks and fewer nodes having an advantage in terms of interconnectedness. The larger a network gets, the more difficult it becomes to keep a larger proportion of nodes interconnected.

Table 7: Transitivity values across the learner and BNC networks.

	A1	A2	B1	B2	C1	C2	BNC
Transitivity	0.47	0.42	0.41	0.4	0.39	0.39	0.39

Transitivity values measured by Arbesman et al. (2010) for a variety of languages range from approximately 0.23 in Basque to 0.4 in Mandarin Chinese. American English scored 0.31, which is lower than what the present study found for British English or any of the learner networks. Higher transitivity in the beginning learner networks could indicate a more tightly knit phonological network during the initial stages of language learning where fewer words are present, which eases up at the later stages. The C1, C2, and the BNC networks are identical in terms of transitivity.

3.3.10. Phonemic inventories and neighbor formation

During the course of language learning, the phonological connectivity naturally increases per successive proficiency level. The limited (but not small) phonemic inventory of British English (see Roach, 2004) restricts phonotactic possibilities and formation of new words so that a small number of phonemes appears in a large number of words. Previous studies have shown that phonemic distributions within languages frequently exhibit a power law distribution (e.g., Zipfian or Yule), and that this is also the case for English (Tambovtsev & Martindale, 2007).

When only a small set of phonemes is used to assemble the majority of words in a language, the likelihood of phonological neighbor creation increases since there are fewer possibilities to add, delete, or substitute phonemes. In fact, biased word formation processes are commonly observed, where highly frequent and probable phonemes are more likely to be used for new word formation (Macklin-Cordes & Round, 2020). This "rich-gets-richer" dynamic on the phonemic scale could point toward a bias in phonological neighborhood creation, where a certain set of "preferred" phonemes is being awarded a main growth advantage at the expense of other phonemes. Figure 28 shows an overview of phonemic distributions per proficiency level of the ESL learners in comparison to the BNC data. Distributional skew is evident and different at each proficiency level. Density plots show that the A1 and A2 levels show the largest phonemic skew, while the C1 and C2 as well as the BNC distributions display the smallest skew and the longest tail.

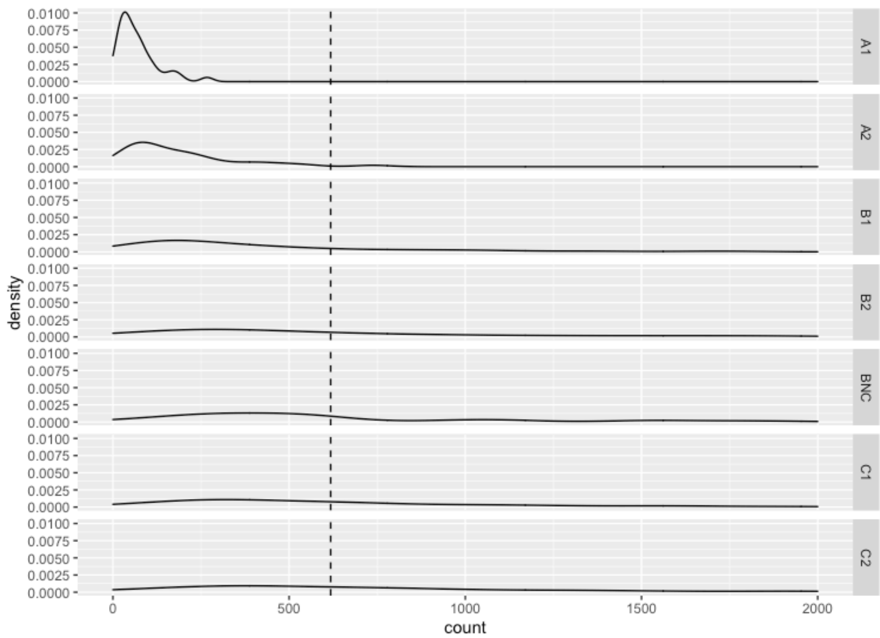


Figure 28: Phonemic distribution in the six proficiency levels of L2 speakers and the BNC data of L1 speakers.

Phonemic distributions were compared between the proficiency levels by calculating a series of independent-samples Kolmogorov-Smirnov tests to determine distribution differences. Results showed that the A1 and A2 levels were significantly different from all the other levels (A1: p values <0.001, D values range from 0.47 to 0.9, df=38; A2: p values <0.012, D range=0.37-0.66, df=38). The B1 distribution differs from C1, C2, and the BNC distributions (p values<0.045, D range: 0.32-0.34, df=38), while the rest of the proficiency levels do not show different phonemic distributions.

Power laws are intriguing for phonemic distributions in languages, as they suggest that growth perpetuates with "fitness" of particular phonemes, a reasonable and often-cited explanation for differences in cross-language phonology (see, e.g., Macklin-Cordes & Round, 2020; Martindale, Gusein-Zade, McKenzie, & Borodovsky, 1996; Tambovtsev & Martindale, 2007)/ This phonological rich-gets-richer mechanism would indicate a tendency to re-use particularly fit phonemes in word formation. The phonemic inventories of the EFL learners and the L1 British English users were tested for their underlying heavy-tail distributions (see Clauset et al., 2009; Gillespie, 2014), but none of the phoneme distributions in the learner networks follow a power law (see results of the distributional analyses in Table 8).

Table 8: Phonemic distributions in the learner and BNC networks.

	A1	*A2*	*B1*	*B2*	*C1*	*C2*	*BNC*
Best fitting distribution	Gamma (dip test p=0.8)	Gamma (dip test p=0.9)	Log-logistic	Log-normal	Log-normal	Log-normal	Log-normal

Following Piantadosi (2014), phonemic distributions have real-world implications of what they actually reflect about a language. Log-normally distributed phonemes per lexicon indicate that the increase in phonemes is proportionally and predictably related. Larger entities in log-normal distributions increase in proportional amounts (the so-called law of proportional effect), rather than through disproportional accumulation, as in power-law distributions. Log-normally distributed phonemes reflect greater predictability, as each phoneme increases its presence in the lexicon based on its proportional share.

The skewed phonemic distributions in the beginning levels suggest a more pronounced tendency for phoneme clustering to involve a few select phonemes at initial learning stages when the vocabulary is rather small. The vowels [ɪ] and [ə] and the consonants [t], [s], and [n] are the most prolific in the A1, while phonemic distribution is more equal in the other levels and the BNC data. One might assume that this distributional skew in A1 and A2 could reflect the fact that the A1 lexicon houses the most frequent words in general (see micro analysis of networks in section 5.1.1. and Figures 16a and 42), which are in turn characterized by frequent phonemes. However, the four mentioned phonemes are generally the most frequent ones across all investigated lexica, including the BNC lexicon (see list below), suggesting that the phonemic make-up of beginning stages of language learning follows statistical probabilities shown by the target language, with learners preferentially learning the higher-frequency phonemes. This is in line with findings from child language acquisition, showing that young infants are already finely attuned to statistical frequencies of phonemes in their native language and acquire high-frequency sounds first (de Boysson-Bardies & Vihman, 1991; Velleman & Vihman, 2007). There are indications that learning of second language phonology also follows this principle, and the present results could support this hypothesis.

Neighbor formation tended to involve different phonemes per proficiency level. The five most common phonemes forming neighbors (by deletion, addition, or substitution) are listed below (in order of frequency):

A1: / ɪ, t, a, n, k/
A2: / ɪ, t, l, e, k/
B1: /i, ɪ, t, n, l/
B2: /ɪ, t, n, s, e/
C1: /ɪ, t, l, n, s/
BNC: /ɪ, r, t, n, ə/

The phonemes /ɪ, t/ are by far the most frequent ones involved in neighbor formation at all proficiency levels, followed by /n, l/ and /k, s, e/. As mentioned above, /ɪ, t, n/ are among the most frequent phonemes to be found in all learner and BNC networks, and their role in neighbor formation underscores the special status of high-frequency phonemes in a phonological inventory. The fact that not all phonemes are equally frequent within and across languages has long been recognized and formalized in the hypothesis of a universal phonological complexity hierarchy that is based on the human articulatory and perceptual language apparatus (Greenberg, 1966; Jakobson, 1941/68). The simpler phonemes are acquired earlier by children and retained for longer in the face of language impairment (see Romani, Galuzzi, Guariglia, & Goslin, 2017, for an overview). Zipf (1935) recognized that the most frequent phonemes are simpler, and the least frequent ones tend to be more complex in articulatory terms (the "principle of least effort"). These findings have more recently been confirmed by showing that easier phonological forms tend to be used and re-used, resulting in shorter words being more frequent and using sequences with higher phonotactic probabilities, while at the same time being associated with higher degrees of homophony and polysemy (Piantadosi, Tily, & Gibson, 2012). Phonological measures of simplicity and complexity are subject to debate (see, e.g. Maddieson, 2009), but markedness theory has played a central role (see Trubetzkoy, 1939). A principle of least effort could apply to phonological neighborhood formation, where more complex ("marked") phonemes have less opportunity to form neighbors, possibly due to their complexity property and the fact that speakers try to minimize their articulatory or perceptual efforts. Preferential accumulation of more frequent and articulatorily easier phonemes in the phonological space of the mental lexicon would be the logical result of such a trend.

3.4. Community structure

Using the Louvain method of community detection in the giant component of the L1 American English phonological network (see Vitevitch, 2008), Siew (2013) identified numerous communities of various sizes with different characteristics pertaining to the phonological word form associations. She investigated a number of lexical characteristics, but her findings concerning phonemic word length, lexical frequency rate, and node degree are most relevant for the present study. Specifically, Siew found larger communities to consist of more frequent and shorter words. This is in agreement with Zipf's law (Zipf, 1935) that states that word length and lexical frequency are negatively correlated. The overall structure of a language is thus reflected on the mesoscopic community level of a phonological network. Ravasz and Barabási (2003) have proposed an intriguing hypothesis that communities of a network are simply sub-graphs of the original network, mirroring characteristics and dynamics of the global network on a local level. In this sense, the largest community is seen as analogous to the giant component of a network, with smaller communities mirroring the islands. In the case of the phonological network investigated by Siew (2013), this prediction holds true.

Generally, words of high lexical frequency are less common, as languages tend to consist of many low-frequency words (Zipf, 1935). The fact that high-frequency words are primarily found in large rather than small communities may indicate that not all areas of the giant component share an equal amount of cognitive processing (Siew, 2013). The high-frequency clusters presumably require more cognitive effort in lexical retrieval, as opposed to other parts of the phonological word form lexicon. Siew also found large communities to contain more high-degree nodes (i.e., denser phonological neighborhoods).

3.4.1. Learner networks

The community detection in the learner networks closely followed the method outlined by Siew (2013). First, modularity values *Q* were obtained for each of the learner networks in order to check the reliability of community detection. Next, the number of communities in each network (excluding singleton nodes) and in the giant component were calculated, and community sizes (including mean, standard deviation, and range) were computed. Average lexical frequency rate, phonemic length, node degree/neighborhood density, and clustering coefficients per giant component community and network were then calculated. The clustering coefficient was included in the present study to determine the effect of phonological neighborhood interlinking in network communities. Table 9 presents an overview of *Q*, communities and their sizes per learner and BNC network.

Table 9: Modularity (*Q*), number of communities, and their sizes in the learner and BNC networks.

	A1	*A2*	*B1*	*B2*	*C1*	*C2*	*BNC*
Q	0.83	0.76	0.74	0.74	0.76	0.75	0.84
Communities (no singletons)	37	87	171	338	439	538	520
GC communities	11	10	11	12	11	11	6
Community size: mean, sd, range	9.1±12; 38	9.4±23.1; 137	8.9±34.6; 315	7.2±34.9; 429	6.6±34.2; 406	6.4±36.4; 673	4.9±22.5; 375

In general, *Q* values are quite high, indicating robust community identification in the phonological networks. Number of communities increased with increasing network size, but community sizes decreased as the networks grew larger. Community sizes were rather large in the learner networks as opposed to the BNC networks. The C2 and the BNC networks were quite similar in terms of the number of communities they house: 538 in the C2 network and

520 in the BNC network. Community sizes (i.e., number of words per community) are largest in the beginning networks and decrease substantially as learning progresses. The fact that standard deviations increase indicates that the variability in community sizes becomes more pronounced in larger networks.

While community sizes were different between the networks (possibly related to network size), the number of communities per giant component was relatively constant. Only the BNC network of L1 users had fewer communities, presumably owed to the fact that the BNC giant component is very small (approximately 13% of the overall network). By comparison, the communities of the phonological network charted by Siew (2013) identified 17 giant component communities; average community size was measured as 383 (standard deviation of 250). Methodological differences between the networks presented by Siew for American English and the present study for British English may explain this discrepancy.

Giant component communities of the seven networks were further investigated and the graphs and results are presented in Tables 10 to 16 and figures 29 to 32 (see below, and see Appendix A for more graphs). Communities were classified as *large* when they contained a number of nodes exceeding the 75th percentile of all communities, as *medium* size when they contained nodes between the 25th and the 75th percentile, and as *small* when they contained less than the 25th percentile in number of nodes.

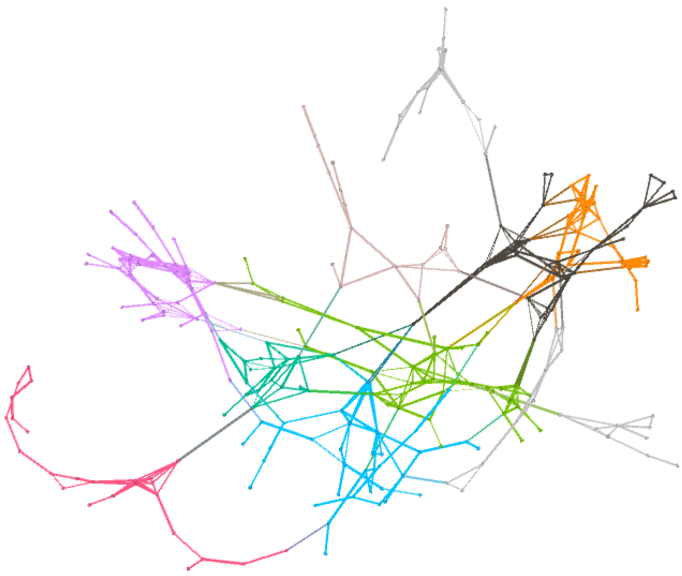


Figure 29: Community structure in the giant component of the A1 network.

In agreement with findings by Siew (2013), the larger communities of the A1 giant component contained words of higher lexical frequency, shorter phonemic length, and denser neighborhoods. Clustering was less pronounced in larger communities than in smaller.

Table 10: Lexical communities of the A1 giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Small (<25th percentile)	440.3	74.6	3.7	0.8	3.2	1.6	0.45	0.4
Medium (25th-75th percentile)	441.1	71.7	3.5	0.8	4.2	2.3	0.36	0.3
Large (>75th percentile)	450.3	86.4	3.0	0.8	4.2	2.9	0.33	0.3

In the A2 giant component, no small communities could be detected. Results of the remaining communities differed from Siew (2013) and larger communities contained fewer high-frequency and generally longer words, as well as sparser neighborhoods with less node clustering (see Table 11 for details and Appendix A for corresponding graph). This inverse trend to the L1 American English network detailed by Siew continued for all consecutive proficiency levels up to C1, and higher lexical frequency rates and shorter words were primarily found in smaller network communities. Node degree and clustering coefficient both increased with community sizes for all networks (see Tables 11 to 14).

Table 11: Lexical communities of the A2 giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Medium (25th-75th percentile)	414.9	81.8	3.4	0.8	5.8	3.9	0.33	0.3
Large (>75th percentile)	402.7	75.9	3.8	0.9	5	3.7	0.31	0.3

Table 12: Lexical communities of the B1 giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Small (<25th percentile)	391.5	48.8	3.5	0.7	0	0	0	0
Medium (25th-75th percentile)	374.3	78.7	3.8	0.9	5.6	4.3	0.31	0.3

Large (>75th percentile)	380.2	78.3	3.7	1	6.4	5	0.29	0.2

Table 13: Lexical communities of the B2 giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Small (<25th percentile)	391.5	48.8	3.5	0.7	0	0	0	0
Medium (25th-75th percentile)	357	72.2	4.5	0.8	4.5	3.3	0.31	0.33
Large (>75th percentile)	364.8	76.8	3.9	1	7.1	5.6	0.29	0.26

Table 14: Lexical communities of the C1 giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Small (<25th percentile)	391.5	48.8	3.5	0.7	0	0	0	0
Medium (25th-75th percentile)	361.7	71.7	4.4	1.1	7.2	6.3	0.38	0.33
Large (>75th percentile)	356.5	77	4.1	1.1	6.7	5.5	0.28	0.26

In the C2 giant component communities, only medium-sized and large communities could be detected, and here, both communities were equal in terms of lexical frequency, phonemic length, node degree, and clustering coefficient (see Table 15 and Figure 30).

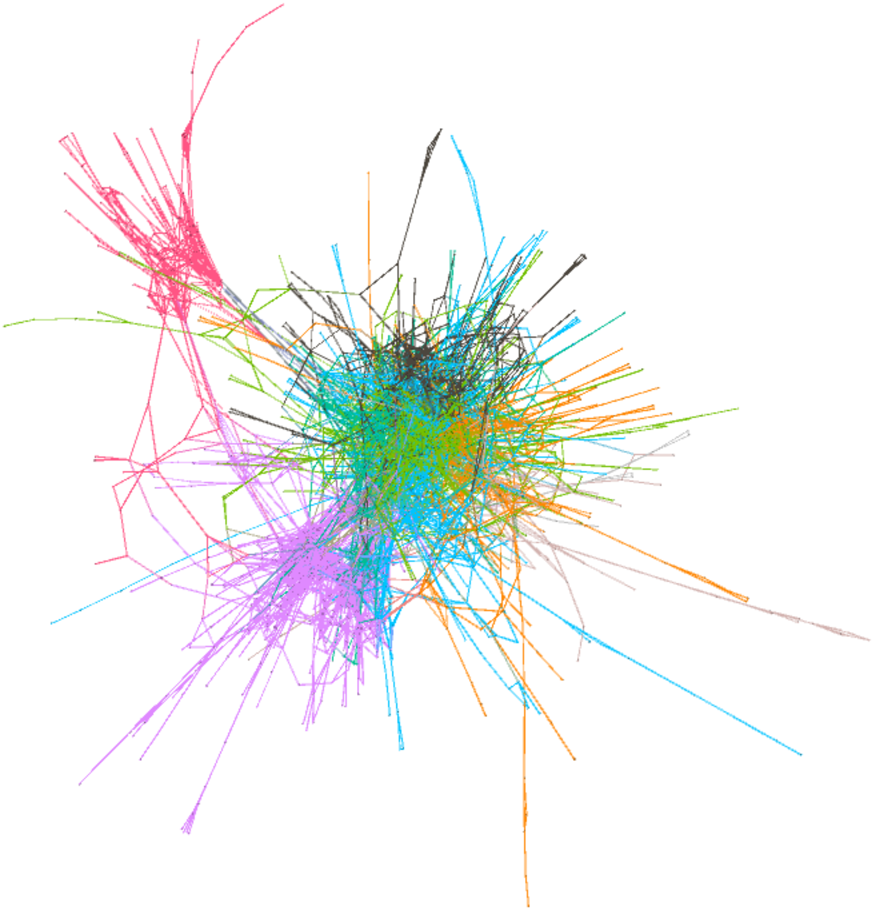


Figure 30: Communities of the C2 giant component.

Table 15: Lexical communities of the C2 giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Medium (25^{th}-75^{th} percentile)	350.3	76.8	4	1	7.1	6.8	0.29	0.27
Large (>75^{th} percentile)	348.8	77.5	4.2	1.1	7	5.8	0.3	0.3

3.4.2. BNC network

The BNC network showed an unusually small giant component and a number of slightly smaller islands/communities surrounding it (see Figure 31).

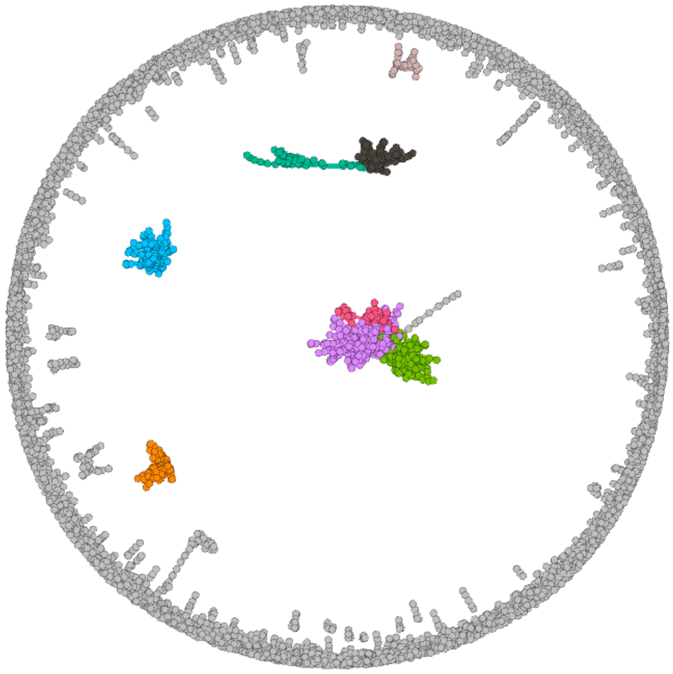


Figure 31: Communities of the BNC network: giant component in the middle, and other larger islands/communities nearby (in color), surrounded by small islands and singleton nodes (in grey).

Table 16: Lexical communities of the BNC giant component.

GC communities	*Lexical frequency*		*Phonemic length*		*Neighborhood density*		*Clustering coefficient*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Small (<25th percentile)	3.9	0.5	3.5	0.7	0	0	0	0
Medium (25th-75th percentile)	2.9	1.8	4.9	0.8	2.3	0.9	0.19	0.35
Large (>75th percentile)	3.1	1.7	3.9	0.8	7.3	5.1	0.32	0.24

The community dynamics as outlined for the intermediate and advanced learner networks are similar in the BNC network: high-frequency, short words are primarily found in smaller communities, while denser and more tightly clustered neighborhoods are found in larger communities. Denser neighborhoods in large communities were also reported by Siew (2013), and this seems to be a robust measurement in relation to phonological community size. The present findings on lexical frequency and phonemic length, however, are not in agreement with Siew. Different databases used for the phonological network constructions could explain these differences. Alternatively, the fact that the advanced learner communities differ from what was shown by Siew could be explained by different word learning dynamics in L1 and L2 speakers.

Growth patterns in the differently sized communities of the learner networks may be able to shed more light on the discrepancy recorded here (see sections 5.1.2. and 5.4.).

In the A1 and A2 learner networks, larger communities house more words of higher frequency, shorter phonemic length, and denser neighborhoods. As these words tend to be learned earlier in language acquisition, the initial build-up of the phonological network and its stratification into (giant component) communities could be geared towards greater robustness and efficiency. Early-acquired words form robust phonological communities, which may ultimately promote the growth of a cohesive language network necessary for rapid and efficient lexical processing (Siew, 2013). L1 and L2 word learning starts with an initial development of large phonological communities in the giant components of their respective networks. Once a small lexicon has been acquired, the statistics of word learning change so that larger communities receive new neighbors of lower frequency and longer duration, while at the same time forming denser and more tightly clustered neighborhoods. Through this dynamic, smaller communities become supplemented with new words which show characteristics of easier word learning (high frequency and short length). The community structure of the learner networks seems adjusted to fewer small communities and more medium-size to larger communities, ultimately leading to the presence of a multitude of similarly-sized communities in the giant component. This effect could be described in terms of inequality minimization in network communities, where the goal is to establish comparable communities. This is an optimal network design in terms of robustness (see section 4.2.2.).

4. Network growth algorithms

4.1. Accumulation of links in networks

Phonological networks grow as learners acquire new words. Growth in networks can follow various developmental paths, and growth patterns play a major role for network functionality and stability. How new nodes and links are added to a system and which changes are initiated to the overall structure of the network as it grows are fundamental questions for evolving networks.

4.1.1. Random and scale-free network growth

Early network models followed the assumption of random network design, where node size is fixed and the relationship between two nodes is seen as a random event that is independent of other links in the network (Janson, Luczak, & Rucinski, 2000; 2001). New nodes entering a system randomly select other nodes to which they can link. This paradigm had its roots in the Erdős–Rényi exponential network model (Erdős & Rényi, 1959) which was originally conceived to represent a static rather than a growing network. Pairs of nodes are connected at a specific probability, leading to a Poisson distribution of degrees. Thus, growth in random networks occurs in a uniform manner across the network (referred to as "uniform attachment model", see, e.g., Fotouhi & Rabbat, 2013; Lugosi & Pereira, 2019), with the expected degrees of nodes increasing over time (see Figure 32).

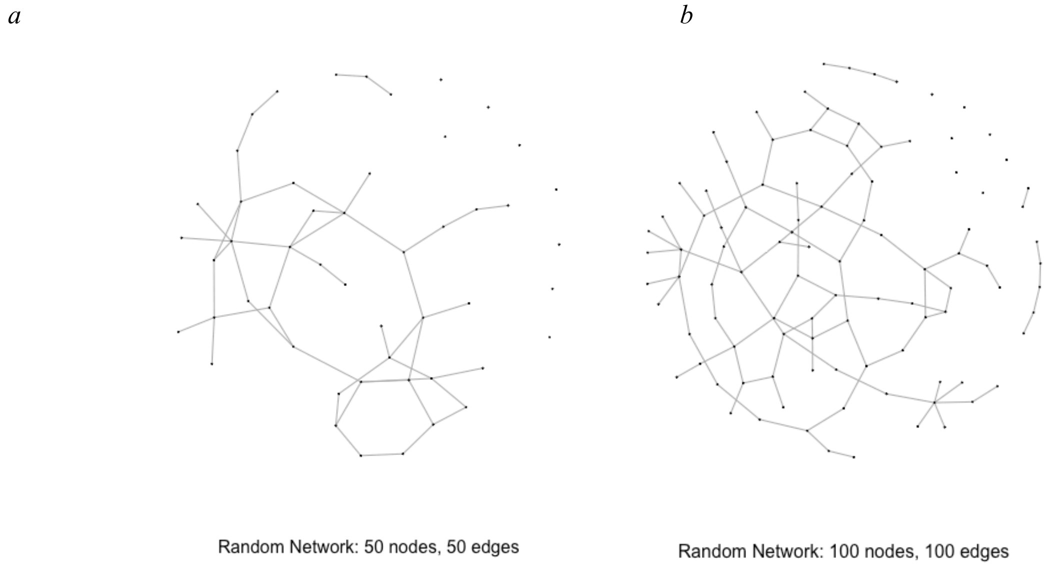


Figure 32: Random network growth (R package “igraph”, function “erdos.renyi.game”).

A variety of complex networks do not follow random distributions of growth but it is a common observation that some nodes have a large number of links, whereas many nodes have only a few links. Connections of world wide web pages, transportation systems, citation networks of academic works, scientific collaborations, among many other real-life networks, contain hubs (i.e., nodes with a very high number of links), with most nodes having just a few connections or none at all (Newman, 2001, 2003b). Such scale-free networks are not predicted to add new nodes randomly but according to degree. This process is called “preferential attachment” and it describes the phenomenon that high-degree nodes (i.e., those with many links) attract the majority of incoming new nodes (Barabási, 2009; Hébert-Dufresne, Allard, Marceau, Noël, & Dubé, 2011); low-degree nodes lose out and cannot grow. Preferential attachment forms the basis of scale-free networks and leads to highly-connected nodes (or hubs) becoming even more highly-connected over time, a rich-gets-richer dynamic or the so-called *Matthews effect* (Rigney, 2010). At the same time, nodes with few links become increasingly impoverished (“weak-gets-weaker” phenomenon). Figure 33 shows an example of network growth in a non-random scale-free network.

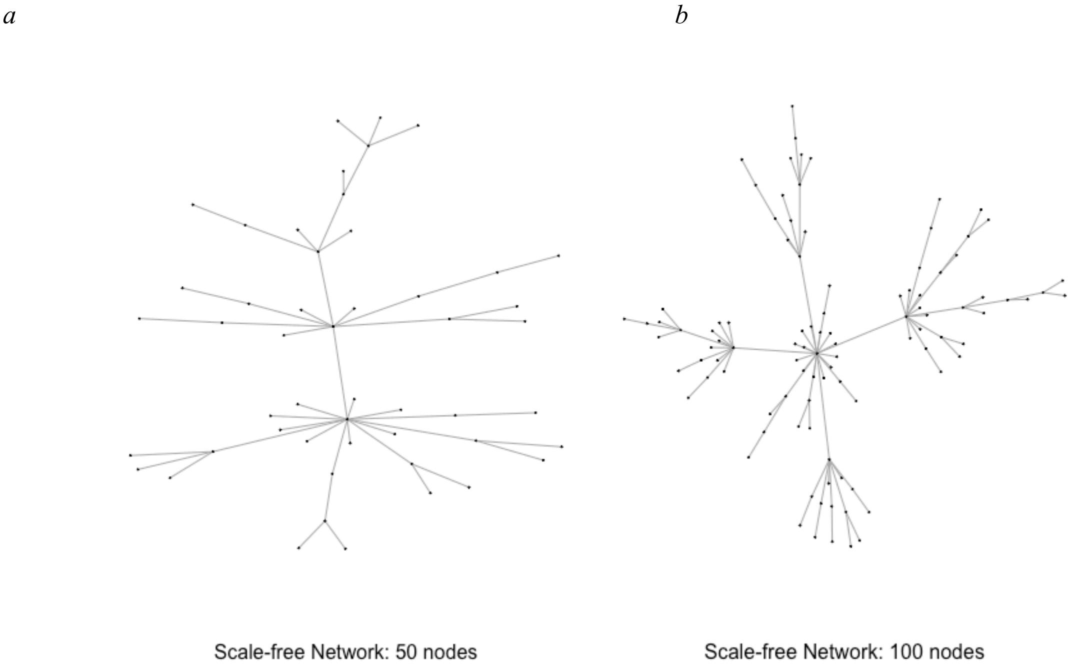


Figure 33: Scale-free network growth (R package "igraph", function "barabasi.game").

An inherent feature of scale-free networks is the assumption that not all nodes are created equal (Barabási, 2009; Barabási & Albert, 1999). Some nodes are more attractive to attach to by virtue of their higher degree: a creator of a new website may choose to hyperlink other popular, high-traffic websites, thereby exercising a bias toward those websites that are already well connected. Some nodes display higher fitness and thus attract the majority of incoming new nodes: for instance, a nice person attracts more friends through their personality in a social network. Preferential attachment in growth can lead to the existence of hubs in networks. As new nodes appear, they tend to attach to highly-connected nodes, driving inequality of degree in the network over time (Bauer & Kaiser, 2017). There is a chronological aspect to growth benefits, with early-acquired nodes being able to increase their connectivity at the expense of late-acquired nodes. Contrary to random networks, in scale-free networks a small number of edges and hubs becomes immensely important for network connectivity. Thus, betweenness centrality of nodes is a meaningful variable to be measured in scale-free networks.

4.1.1.1. Power laws in scale-free networks

Barabási and Albert (1999) have suggested that preferential attachment causes networks to become scale-free, and node distribution will follow a power law. Generally, a network is considered scale-free if a fraction of nodes follows a power law distribution $k^{-\alpha}$, with $\alpha > 1$ (Broido & Clauset, 2019; Newman, 2010). It has been proposed that power law distributions are the only scale-free distributions in networks (Newman, 2005), and thus the terms "scale-free" and "power law distribution" are often considered as synonymous (Milojević, 2010). However, this claim has been challenged by numerous researchers who posit that scale-freeness can arise from other distribution types (see Li, Alderson, Doyle, & Willinger, 2005, for a short review). Clauset, Shalizi and Newman (2009) found that statistical tests rarely yield a power law distribution in networks, and Stumpf and Porter (2012) point out the suboptimal mathematical fitting commonly employed to determine power laws in networks.

In general, a non-linear power-law relationship between node size and growth is not the norm, as there can be other mechanisms of preferential attachment (Redner, 2005; Sheridan & Onodera, 2018). Linear preferential attachment can be observed when a new node is twice as likely to attach to a node that has twice as many neighbors as compared to a competing node. Non-linear preferential attachment mechanisms, on the other hand, are less predictable and can lead to scenarios where a new node may be ten times as likely to attach to a node with twice as many neighbors. Over time, non-linear growth leads to a star-shaped network typology (also referred to as "hub-and-spoke" typology), where the "winner-takes-all" mechanism is prevalent and all incoming nodes will primarily attach to the large hub (Vlachos, Parousis-Orthodoxou, & Simos, 2008; see Figure 34).

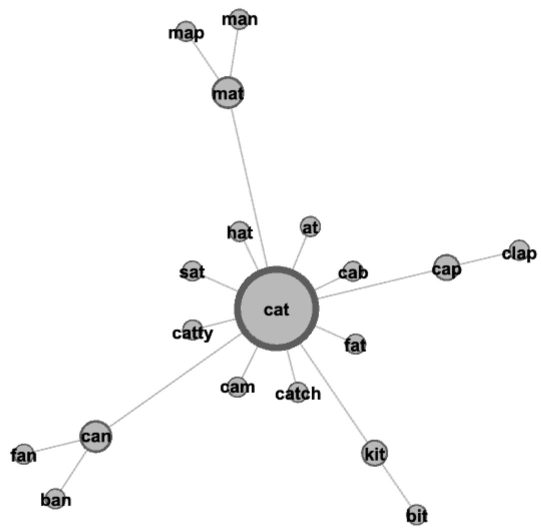


Figure 34: Hub-and-spoke network typology resulting from a close approximation of the "winner-takes-all" mechanism of node growth.

It has been suggested that costs associated with forming new links make a power law distribution less likely to occur (Amaral, Scala, Barthélémy, & Stanley, 2000). If nodes are only able to accommodate a fixed number of overall neighbors, each new neighbor adds an overall cost. Such a restriction may well exist in phonological networks, where word length, phonemic repertoire, and phonotactic constraints could limit growth potential per node (Vitevitch, 2008). A word cannot grow phonological neighbors ad infinitum, as phonotactically legal segments are limited and the cognitive ability to discriminate words must be upheld by not clustering too many homophones and phonological neighbors together. In fact, power law degree distributions are not commonly found in phonological networks (including the learner networks constructed for the present study, see chapters 3 and 5). Cost-conscious growth can have consequences for assortativity patterns and/or the presence of lexical isolates (singleton nodes) in a phonological network (Vitevitch, 2008). A striking conclusion from scale-free networks is that many nodes interact with a limited number of other nodes, whereas a small subset of nodes interacts with a high number of other nodes. This places a high processing burden on the hub (or hubs). In the case of phonological networks, this interaction pattern leads to few words being characterized by high prominence and influence in the network, whereas the large majority of words have limited ability to exert influence on (or diffuse activation to) others.

Two concepts are important in the context of network growth: evolvability and robustness of networks. The former refers to the ability of a network to change and adapt over time (including the incorporation of new nodes), while the latter concept refers to the ability to withstand failures.

4.2. Evolvability and robustness of networks

Per common definition, 'evolvable networks' are able to accommodate changes and adaptations, while 'robust networks' show a topological shape that makes them resilient to node and edge disturbances by keeping a stable connectivity even in the face of changes (see, e.g., Payne & Wagner, 2014). There is an intricate relationship between evolvability and robustness: a resilient structure may present challenges to evolvability of a network (Pigliucci, 2008). Robustness and evolvability are generally opposing mechanisms – while the former promotes stability, the latter requires some degree of instability (Wagner, 2007). However, robustness must also be seen as a result of the ability to adapt to changes through evolution (Chen & Lin, 2011), as only an adaptable system can maintain its own long-term survival.

4.2.1. Evolving networks

A debated question in the network sciences concerns the node locations that are most evolvable within a scale-free network. Hubs with many neighbors take a central position and their behaviors are influential across large parts of the network. This has led to the assumption that large hubs should be better insulated against changes, as the multitude of links exert conformity pressures within the word neighborhood and thus suppress change (Kim, Korbel, & Gerstein, 2007). The more payoff there is for each individual node, the more likely it is to conform to the conservative standard upheld by its neighbors. In phonological network, payoff can be measured in activation – the more co-activation a phoneme/word receives, the more likely it is to keep its current (phonological) form and resist changes. Indeed, the degree of a node can correlate with the rate of change or evolvability (Alvarez-Ponce, Feyertag, & Chakraborty, 2017). Therefore, changes (or mutations) are frequently expected to show up in more peripheral nodes. But network dynamics can also be different, depending on the type of network that is investigated. In some scale-free networks evolvability is actually facilitated by the existence of "perturbations" or "mutations" in hubs (Helsen, Frickel, Jelier, & Verstrepen, 2019). In genetic networks, a higher rate of mutations in central and highly connected nodes (i.e., hubs) does not lead to a breakdown of the network but actually offers evolutionary advantages by allowing a fitter phenotype to gain a foothold and spread through the network (Koubkova-Yu, Chao, & Leu, 2018). The essential and central role of hubs should theoretically lead to slower evolutionary rates (Alvarez-Ponce et al., 2017; Kim et al., 2007), however certain characteristics may predispose hubs more to being evolvable. In a phonological network of L1 German, hubs were more likely to show one type of sound change, while non-hubs were more likely to incorporate a different type of sound change (Luef, in preparation). Characteristics of the specific hubs as well as of the specific changes that are introduced (type of adaptation) may both be relevant for the question of evolvability in phonological networks. Since hubs are more highly connected and regulate information flow to a greater degree within a network, their behavior has major implications for network connectivity.

4.2.2. Robustness measures

Network robustness is commonly investigated with percolation theory, which stems from statistical physics and mathematics and can be applied to the question of how networks can break up into smaller, non-connected parts (Li, Zhang, Zio, Havlin, & Kang, 2015). It essentially tracks the gradual fragmentation of a network by removing nodes at random, all the while analyzing how their absence influences the integrity of the network (essentially the inverse of a percolation process, Barabási, 2016). Removal of few nodes does not have much of an impact; removal of a larger number of nodes can lead to a break-up of the giant component. A critical threshold value of the node removal variable f_c is defined that cannot be exceeded or the giant component disappears (Barabási, 2016). Random node removal can induce a phase transition from a connected to a fragmented network; in fact, after a finite fraction of nodes (i.e., f_c) are removed, a network break-up is imminent (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006). In scale-free networks, however, a large number of nodes can be removed with the network remaining largely intact. This unusual robustness if furthered by the fact that in those networks large hubs and giant components can be found (Barabási, 2009; Cohen & Havlin, 2010). Scale-free networks have unusually large f_c values (which is related to their degree distribution, Barabási, 2016), and are able to withstand an arbitrary level of node removal without suffering major failures. As random node removal does not consider node degree, the likelihood of a low-degree node being removed in a scale-free network is much higher than the likelihood that one of the very few high-degree nodes (hubs) is affected. Low-degree nodes contribute little to network connectedness and their removal has little impact on the overall network. As seen in Figure 35, in a network of 30 nodes (constructed with data from the B2 learner network of the present study), each node has a $1/30^{th}$ chance of being selected for removal in the percolation theory context. The chances that one of the 26 small-degree nodes is selected for removal is far higher than the chance that one of the three larger-degree or the one hub node are affected.

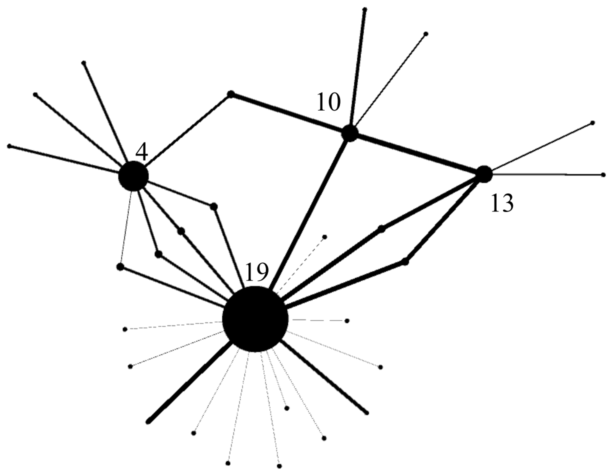


Figure 35: According to percolation theory, each node has an equal 0.03 change of being selected for removal. This leads to a higher likelihood of small-degree nodes being selected as they are more numerous.

Using R ("runif" function for uniform distributions), 1000 runs of probability simulations of the above network produced the following choices for random node removal (see Table 17): the probability of node 19 failing is rather low and the node is selected for random removal only 29 out of 1000 times.

Table 17: Probability of node removal in network of 30 nodes in 1000 simulations.

	Node 19	**Node 4**	**Node 10**	**Node 13**	**Average other nodes**
Hits	29	24	45	41	33
Probability of event	0.03	0.02	0.05	0.04	0.03

In general, scale-free networks are very robust against accidental failure or random node/edge removal but may fall apart if targeted attacks against the large hubs are performed (Albert, Jeong, & Barabási, 2000; Newman, 2002b; in phonological/semantic networks: Stella, 2020). Network failures, due to a loss of connection between nodes, may lead to (substantial) increases in average path length, and this is an indication that the network functioning is compromised (Albert et al., 2000). Networks become more robust when node clustering is more evenly distributed and less tightly clustered in one giant component (Hu & Lee, 2020). What is known as 'cascading failure' in technological and biological networks is a chain reaction caused by an error (or failure) in one node that leads to larger, more far-reaching failures in connected nodes (Ren, Song, Yang, Baptista, & Grebogi, 2016). Large giant component sizes combined with high clustering coefficients and short average path lengths heighten the risk of cascading failures (Hu & Lee, 2020). When a highly connected node in the giant component fails, a large number of links may be terminated and giant component diameter, as well as average path lengths, increase.

A number of network centrality measures – chief among them degree centrality – are predictors for robustness, and networks with higher average degrees are more robust (Martin & Niemeyer, 2020). As network centrality measures are inherently related to network size, robustness is predictably different at different stages of network growth. Zhao and Xu (2009) investigated network growth (in theoretical mathematical models) from the viewpoint of strengthening network robustness by comparing three different mechanisms of node addition: (1) add a new neighbor to an isolated node, (2) add a new neighbor to a low-degree node (i.e., one with few neighbors), and (3) add a new neighbor to a high-degree node (i.e., one with many neighbors). Results proved that the strategy to supplement low-degree nodes in a network provides the greatest robustness benefits for growing scale-free networks. Through this kind of growth, focussed clustering of nodes in a giant component is markedly reduced and more numerous but smaller, unconnected clusters appear, protecting the network better against failure.

Phonological networks are prone to failure – diseases, cognitive decline, or temporary memory loss (e.g., forgetfulness, word-finding difficulties) can diminish phonological neighborhoods, with potential ensuing systemic failures in connectivity of the mental lexicon (Stella, 2020). Non-pathological temporary phonological processing impairments, such as speech errors or perception errors (i.e., 'slips of the ear'), are also indicative of lexical processing failures, and studies on speech errors have produced evidence that phonological neighborhood effects are involved, most frequently dense neighborhoods (Harley & Bown, 1998; James & Burke, 2000; Vitevitch, 2002c; Vitevitch & Sommers, 2003). Vitevitch (1997) found that overly dense and overly sparse phonological neighborhoods can both impede lexical processing and lead to more malapropism-type speech errors (depending on the neighborhood frequency), indicating that lexical cognition functions best when phonological neighborhoods consist of a moderate number of neighbors. Thus, a general mechanism by which word growth in phonological networks is primarily concentrated in low-density neighborhoods (i.e., low-degree nodes) best fits the stability model outlined by Zhao and Xu (2009) and the predictions stemming from Vitevitch (1997). A more equal distribution of (phonological) resources among the giant component, islands, and singleton nodes provides best protection against failures (Estrada, 2006). Therefore, a prevalence of more numerous smaller components in a network indicates improved robustness and higher stability.

4.3. Growing scale-free networks

4.3.1. Barabási–Albert model

Since the inception of preferential attachment, a variety of network growth models have been introduced that aim to supplement as well as challenge the concept of preferential attachment in aiming to better explain scale-free network growth. In the classical preferential attachment Barabási–Albert model, each node gets to increase its degree according to the equation

$$p_i = k_i / \sum_j k_j$$

where the probability that a new node links to node *i* (p_i) is the degree of the node *i* (k_i), divided by the sum calculated overall pre-existing nodes *j* (i.e., the denominator results in the overall number of edges in the network). Preferential attachment can be seen as highly probabilistic, with the probability of node attachment increasing with degree of a node (linear growth). For instance, a new node is twice as likely to connect to a six-degree node in the network than it is to a three-degree node. This is a simple degree-based attachment model that exemplifies the rich-gets-richer mechanisms in the network and incorporates a correlation of degree with age of nodes (Barabási, 2009). The fact that older nodes capture the majority of links is referred to as "first mover advantage" (Bianconi & Barabási, 2001).

In their study of phonological network growth in Dutch- and English-speaking children, Siew and Vitevitch (2020a) identified preferential attachment (termed PATT) as the main growth mechanisms for the early stages of language acquisition (ages 3 to 6 years). However, at more advanced acquisition stages (over 6 years of age), the inverse variant of preferential attachment (termed iPATT) was shown to be the driving force behind phonological neighbor acquisition.[2] As their results showed, young children grew words by adding phonological neighbors to high-density neighborhoods (high-degree nodes), but this growth mechanism was reversed when the

[2] It has to be noted that the authors did not calculate preferential attachment following the above network equation but used node degree as a proxy variable (see chapter 6 for more details).

children had already acquired a larger vocabulary at later stages of language acquisition. At this point, new words became primarily linked to words in lexical islands, which had few neighbors. To further explore these findings, Siew and Vitevitch (2020b) conducted a series of computational simulations of vocabulary growth algorithms, and their results confirmed that initial PATT followed by iPATT results in a network typology that resembles known phonological networks. Specifically, such a mechanism leads to a degree distribution that does not adhere to a power law and shows small network diameter combined with short average path length (Vitevitch, 2008). The iPATT strategy is analogous to a process known from theoretical network robustness studies – adding new neighbors to low-degree nodes helps strengthen a network's resilience (Zhao & Xu, 2009). Since a large giant component makes a network more vulnerable, one strategy to minimize attack risk is to grow in a way that supplements low-degree nodes. Phonological networks follow this strategy, proving that robustness plays a role in network construction. Siew and Vitevitch (2020a) only found this dynamic in advanced stages of language acquisition, and this suggests that the lexicon of adult/advanced first-language learners is set up in a way to strengthen its resilience and counter vulnerabilities. At the initial stages of network construction (in child language acquisition), preferential attachment causes a "rich-club" topology where high-degree nodes gather most new links. These different strategies for vocabulary growth in beginning and advanced first-language learners offer intriguing perspectives into second-language learning strategies. Here, too, the initial vocabulary build-up may be guided by preferential attachment, followed by a "poor-gets-richer" mechanism at higher proficiency stages.

The original Barabási-Albert model was extended to include more mechanisms (besides from preferential attachment), resulting in a family of evolving network models that can better account for different phenomena that shape topology of real networks (Newman, 2003b). One extension of the Barabási-Albert model is the initial attractiveness model (Dorogovtsev, Mendes, & Samukhin, 2000a). Barabási (2016) gives the example of research paper citations: each new research paper has a finite probability to be cited at least once (see, e.g., Ying, 2011). In social networks, a new student arriving at a school makes at least one friend. Initial attractiveness can better model real-life situations where most unconnected nodes do acquire some first links (albeit few and only initially). In general, networks become more homogeneous if initial attractiveness plays a role. This mechanism reduces the size of hubs, and increases the probability of no-degree or low-degree nodes receiving a small share of the new nodes as neighbors. Initial attractiveness weakens preferential attachment to some degree, since it favors growth in small-degree nodes. Its effect on high-degree nodes is negligible (Barabási, 2016). By raising the possibility of small-degree attachment, initial attractiveness leads to a small-degree cutoff in the power law distribution.

4.3.2. Fitness model

While the preferential attachment model and its inverse variant can explain some phenomena observed in phonological networks, there may be other factors – besides from degree and node age – that influence the ability of a word to grow its neighborhood. A number of network growth models consider node-intrinsic qualities, with the internet search engine Google being a famous example: even though Google came late, it quickly became the biggest hub of internet search engines (Bell et al., 2017)). Here, young age and few previous links did not pose obstacles to Google acquiring the majority of future links. Thus, a consideration of node-intrinsic properties that make a node in a network particularly prone to neighbor acquisition can add considerably to growth models. Node fitness refers to an accumulation of attributes that make a node prone to growth: a fit node acquires a larger number of new neighbors than a less fit one (Bedogne' & Rodgers, 2006; Caldarelli, Capocci, De Los Rios, & Munoz, 2002). In principle, node fitness can be modelled after any quantifiable property or properties of individual nodes (Ferretti, Cortelezzi, Yang, Marmorini, & Bianconi, 2012) or depend on the properties of the neighboring node (Ferretti et al., 2012; Papadopoulos, Kitsak, Serrano, Boguna, & Krioukov, 2012), with the exact fitness qualities depending on the type of network that is charted. In phonological networks, factors that contribute to better retention or retrieval of a word could serve as aspects of node fitness. It is well-known that word length and lexical frequency are major contributors to better retention and faster and more accurate word learning, in addition to low phonotactic probability (e.g., Crossley, Skalicky, Kyle, & Monteiro, 2019; Ellis, 2002; Goodman, Dale, & Li, 2008; Storkel et al., 2006; Storkel, 2001).

Fitness-based growth models attribute more growth to fitter nodes and assume that scale-free topologies arise from the unequal distribution of node fitness across networks (Garlaschelli & Loffredo, 2004). This conceptual approach is often referred to as the “fit-gets-richer” mechanism (Caldarelli et al., 2002) and shows that power-law distributions can arise from factors other than preferential attachment (Mendes & Da Silva, 2009). While the majority of fitness growth models focus on higher growth in fitter nodes, a related mechanism has been suggested by Bell and colleagues (2017), the “avoidance of the weakest links” dynamic. Such growth occurs in a way that minimizes “maximum exposure to node unfitness” (p. 1), essentially providing each node in a network with a minimum of fitness. A bias of new nodes to attach to the least-fit nodes in the system could serve as evidence for that.

Hybrid models combine node fitness and preferential attachment to model network growth. A prominent example is the Bianconi-Barabási model, following this equation:

$$p_i = \eta_i k_i / \sum_j \eta_j k_j$$

where the probability of a new neighbor at node *i* can be calculated by taking into account the fitness of the node (η_i) and the fitness of the other nodes in the network (η_j). If two nodes have the same fitness, the one with the higher degree will be more likely selected. If two nodes have the same degree, the one with the higher fitness score will excel (Bell et al., 2017). In the Bianconi-Barabási model, a young node can grow faster than an older one, overcoming the age disadvantage (induced by the first-mover-advantage) through higher fitness.

As mentioned above, measurements of fitness can vary. Node-intrinsic qualities are different depending on the type of network under investigation. One can also investigate growth rate itself as a fitness measure. Here, neighborhood growth of words is tracked over a period of word learning, and those words that acquire the most links overall are classified as the fittest nodes. This second approach captures a network's "collective perception of a node's importance relative to the other nodes" (Barabási, 2016: p. 208) and provides a measure of evolvability by comparing a node's evolution to that of other nodes in the network. The fittest nodes can then be investigated in terms of their intrinsic qualities, and conclusions can be drawn as to what constitutes node fitness in a specific network.

A fitness-related issue in network growth is the Bose-Einstein condensation, a specific behavior of subatomic particles in quantum gas that can be mapped to networks (Bianconi & Barabási, 2001). The original Bose-Einstein condensation (Bose, 1924; Einstein, 1924) describes a specific energy state of gas particles, where a group of atoms is cooled close to absolute zero, at which point the atoms begin to clump together and enter the same low energy states. They are not able to take on more energy at this point, regardless of how much Bose liquid is being added (see Glazer & Wark, 2001). The only particles taking on the incoming energy are the few ones with high energy levels. This principle can be useful as it describes the transition process from a fit-gets-richer to a winner-takes-all growth mechanism, a phenomenon in certain networks (Barabási, 2016). By mapping network nodes to energy levels and network links to particles, Bianconi and Barabási (2001) demonstrated that networks can undergo Bose-Einstein condensation. The process predicts two distinct phases: a scale-free phase of fit-gets-richer dynamic followed by the Bose-Einstein condensation where all network nodes crowd to the lowest growth level, with a few prolific (fittest) nodes capturing a finite fraction of the new links and turning into super-hubs. At this point, scale-freeness is lost and the dominant dynamic can be described as winner-takes-all (Barabási, 2016). Whether a network can undergo a Bose-Einstein condensation depends on the fitness distribution, where a few fittest nodes and a majority of less fit nodes compose the network. Developments of fitness distribution over stages of language learning can give insights into a possible Bose-Einstein condensation in phonological networks of ESL learners. Figure 36 shows a Bose-Einstein condensation in a hypothetical network.

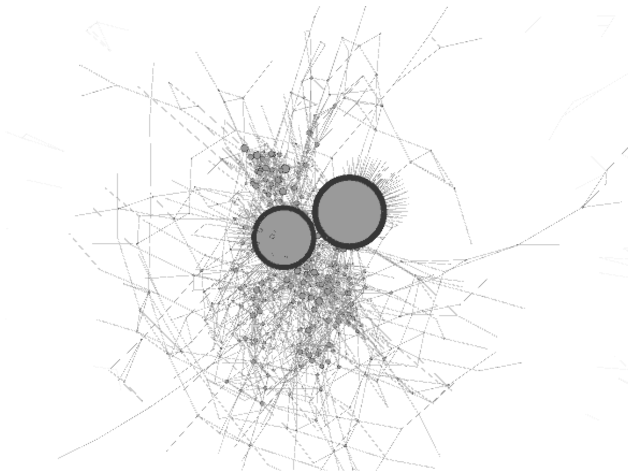
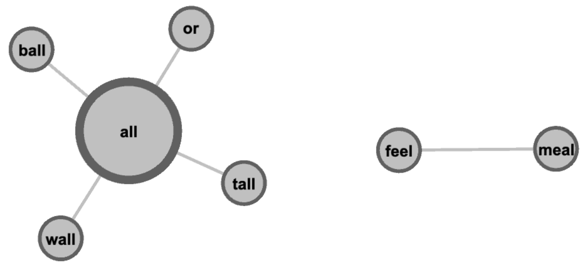


Figure 36: Bose-Einstein condensation in a network with two super-hubs taking all incoming links.

4.3.3. Internal linking: "double preferential attachment"

Another special network dynamic can result when new neighbors have the effect of linking together two pre-existing nodes in a network that were not neighbors of each other previously. In such a case, there are two (interlinked) mechanisms by which new links are established in the network: (1) arrival of new nodes and (2) linking of existing nodes. In phonological networks, this commonly happens during word learning (see Figure 37). In English beginners (A1 proficiency level), "all" and "feel" are not linked neighborhoods, whereas in slightly advanced learners (A2 proficiency level), the two neighborhoods are linked by the introduction of the new word “fall”. Different predictions concerning co-activation ensue (see section 2.3.) on the effects of the clustering coefficient, average path length, and separate network components on lexical processing).

A1



A2

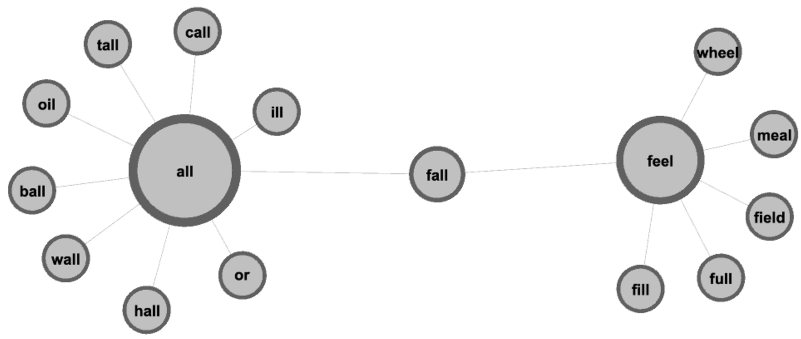


Figure 37: The introduction of new words can lead to new neighbor formation in pre-existing words in a phonological network (data from the A1, above, and A2, below, networks of the present study).

The internal linking mechanism can lead to so-called “double preferential attachment” (Ghoshal, Chi, & Barabási, 2013), which is the probability that an internal link connects two high-degree nodes network-internally. In the A2 network, neighborhood density is quite high for “all” and “feel”, making them more likely to become linked by the introduction of a new word if double preferential attachment is the driving force. Double preferential attachment generally leads to more heterogeneity in a network by making larger hubs larger (through linking them), which results in small-degree nodes losing out in terms of growth potential (Barabási, 2016). It is also possible that internal links are not governed by preferential attachment but by “random attachment”, where internal links are established on a random basis (Ghoshal et al., 2013). In such a case, networks would become more homogeneous in terms of node-degree distribution.

4.3.4. Aging effects and declining networks

Real networks are often characterized by node removal; in the case of phonological networks, that is essentially forgetting a word due to cognitive decline or speech errors (Stella, 2020). If node disappearance outpaces node arrival, the network naturally shrinks (for instance in neurodegenerative disease or when a learner stops using a language). So-called ‘declining networks’ lose nodes at a higher rate than they add new ones. A mechanism called “preferential survival” has been suggested to account for node loss, with high-degree nodes being better equipped to survive a round of node deletions (Kong & Roychowdhury, 2008). In some networks, new nodes may be added at the same rate as old ones are removed, in which case the network may lose its scale-free nature (Saavedra, Reed-Tsochas, & Uzzi, 2008). Here, the question is also whether nodes are removed at random, or whether there is a system to node removal. For instance, low-degree nodes may be more prone to failure or loss, as shown in lexical networks of aphasics (Stella, 2020). Co-existence of node removal with other network-internal processes can lead to complex interactions, such as was shown for the declining network of the garment industry in New York City, where the giant component shrunk considerably but the degree distribution was not affected by this (Saavedra et al., 2008). In the Saavedra et al. study, the declining network preferentially lost low-degree nodes (weak-gets-weaker phenomenon), but the new nodes that occasionally arrived followed the rules of preferential attachment. This led to a tight clustering of the shrinking number of nodes in a giant component, which overall shrank because node removal exceeded node arrival.

The so-called aging effect in networks relates to node disappearance when nodes have a limited lifetime (Sun, Michaels, & Mahadevan, 2020). This is the case for collaboration networks of scientists, when the professional life span of the scientist is limited by the length of their employment or active publishing period. The nodes here are not removed suddenly but gradually fade away through an aging process: scientists start collaborations early in their academic life, then go through a period of higher productivity that may be characterized by a multitude of collaborations, and then toward the end of their career they slowly end collaborations, until they retire and stop collaborating altogether, at which point they exit the network. Different dynamics may be at play in networks showing aging effects: new nodes may preferentially connect with older ones in the network (thereby accelerating preferential attachment), new nodes may show a tendency to connect with young nodes in the network, or each new node may attach to the previously introduced one, turning the network into a chain-like structure (Zhu, Wang, & Zhu, 2003).

4.3.5. Accelerated growth

Networks may also show accelerated growth or even hyper-accelerated growth, in addition to decelerating growth (Gagen & Mattick, 2005; Mendes, 2003). Accelerated growth is defined as a super-linear increase in nodes in a network over time (Liu et al., 2019) and it occurs for instance with scientific collaboration networks or the world wide web (Dorogovtsev & Mendes, 2003). Popular network growth theories calculate the probability that a new node attaches to an existing one by taking into account the degree of the existing one (see, e.g., Albert & Barabási, 2002; Barabási & Albert, 1999), essentially defining growth as a linear process. Many real networks, however, do not follow this rule but show growth rates exceeding the linear predictions. Here, the accelerated growth over a period of time eventually leads to the creation of super-hubs.

4.3.6. Local-world networks

The phenomenon of local worlds captures the notion that networks are divided into sub-groups (i.e., the communities), where members show more feature similarity to one another than to other nodes in the network (Shai, Stanley, Granell, Taylor, & Mucha, 2017). Local-world evolving network models have been developed to account for the fact that in many networks new nodes can only attach to a specific part of the network (Li & Chen, 2003; Li, Jin, & Chen, 2003). Each new node can access a sub-part of the network where the group of pre-existing nodes show certain similar characteristics to the new node. For instance, only certain countries trade with one another, thus a new shipping company can only establish routes between a set of predefined countries, rather than between any two random countries globally that participate in shipping. Local-world models are suited for phonological network investigations: a new word form can only attach to phonologically similar word forms, i.e., to a group of words sharing similar phonotactic combinations but not to any random node in the whole network. There are different ways to conceptualize phonological sub-groups: either through community detection in the phonological networks, where certain communities are the natural attachment points for new words (see Cong & Haitao, 2014; Siew, 2013), or by finding all possible neighbors to which a new node can attach and group them as the "potential attachment points" (the so-called lure-of-the-associates mechanism, Hills et al., 2009b). The first approach defines a larger phonological neighborhood, where a new node can find connection points. The second approach zooms in on specific connection points but disregards the larger phonological neighborhood to which those points belong. This is schematized in Figure 38. The different colors indicate phonological communities, created through greater phonological similarity among its member (i.e., community detection, see Yang, Algesheimer, & Tessone, 2016), and they can serve as local-world communities in local-world networks. Alternatively, the actual attachment points for a new node can serve as the local-world community, since they represent all the attachment possibilities that exist for a new node. If the new word "hit" were to enter the system, its community membership (orange) or its link to individual nodes ("hat", "hot") could become the decisive feature.

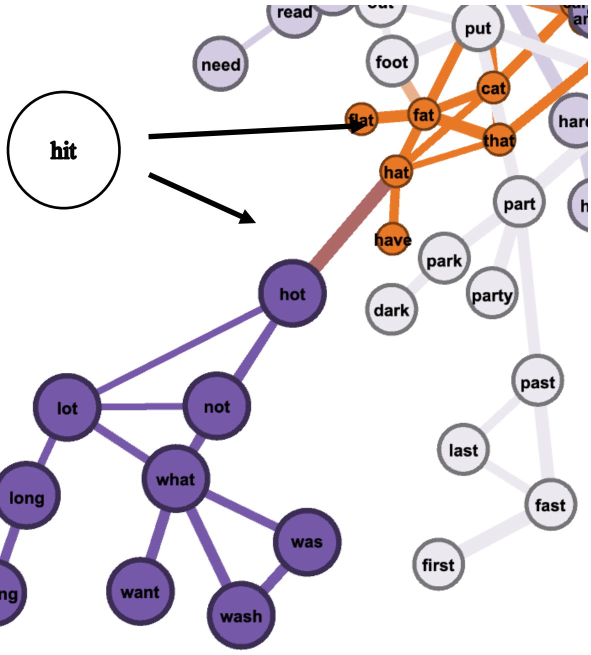


Figure 38: In the A1 learner network, the new word "hit" has one network community (orange) but another community of individual connection points ("hot", "hat", i.e., the lure-of-the-associates dynamic).

The following equation, adapted from Li and Chen (2003), calculates the preferential attachment of node *i* by taking into consideration the local sub-group of nodes (LOCAL) to which an incoming node can potentially attach.

$$p_i = k_i / \sum_{j \in \text{LOCAL}} k_j$$

Local-world connectivity has major implications for phonological networks. First, it recognizes that a node has a limited ability to attach in the network, since only a small number of the overall nodes are phonologically predisposed for anchoring. Second, the local-world effect generally improves robustness of scale-free networks (Tang & Liu, 2014), an important criterion for cognitive network construction (Walsh & Gluck, 2015) and paramount for phonological networks (Siew & Vitevitch, 2016; Stella, 2020). Local-world effects likely occur in combination with other network-relevant mechanisms, such as fitness dynamics (including the Bose-Einstein condensation), internal linking, or accelerated growth, among others. These mechanisms may act out locally in node communities rather than across the entire network.

4.3.7. Uniform and preferential attachment

Another group of network growth models, the hybrid random networks (HRN), combines preferential attachment with uniform attachment (as known from random network growth, see above). Here, the probability of new node creation is taken into account when assigning it to the preferential or uniform attachment mode (Wang & Zhang, 2021), as well as the age of nodes (with younger nodes becoming attached uniformly, see A. Pachon, L. Sacerdote, & S. Yang, 2018). Classical preferential attachment models predict all nodes in a network to behave in the same way and attach to the highest-degree node. Models where this strict assumption is relaxed has arisen recently, and they allow new nodes to attach to a network following a combination of uniform and preferential attachment (Cooper & Frieze, 2003; L. Pachon, L. Sacerdote, & S. Yang, 2018). While there are different ways to calculate this combination mathematically (see, e.g., De Ambroggio, Polito, & Sacerdote, 2020), the following equation represents a simple compound variable

$$p_i = UA_i k_i / \sum_j UA_j k_j$$

with uniform attachment "UA" denoting the proportion of a node within its network (for instance, each node in the A1 network has a proportion of $1/606^{th}$). UA and degree of a node (k) are computed proportionally to the sum of all nodes, with their UA values and degree yielding the hybrid uniform/preferential attachment variable. Preferential attachment of node *i* and random attachment probability – the proportion of node *i* in the network – are combined.

The introduction of uniform attachment leads to lighter tail distributions in the networks than what would be expected if preferential attachment was the main mechanism. An important consideration for node growth in networks is the possibility that a new node does not have to attach to a pre-existing one, as implied by scale-free and most other network models (e.g., Barabási & Albert, 1999). Rather, a new node may enter the network as a singleton node without any immediate neighbors. Phonological networks are characterized by a large proportion of singleton nodes (approximately half of the network, see Gerometta, 2015; Vitevitch, 2008, and the results of the present study), rendering a mechanism to add nodes without attachment an important theoretical consideration. Some random network growth models can account for singleton nodes entering the network (Callaway, Hopcroft, Kleinberg, Newman, & Strogatz, 2001). New emerging types of hybrid growth models combine elements from random network growth with preferential attachment and can potentially explain how a large number of singleton nodes can enter a network, while at the same time accounting for preferential attachment playing a role (e.g., Anwar, Yousuf, & Abid, 2021; Banerjee & Bhamidi, 2021; Callaway et al., 2001; Shao, Zou, Tan, & Jin, 2006; Wang & Zhang, 2021; Weaver, 2015).

In general, most networks are characterized by a complex combination of growth processes, including preferential attachment, random attachment, and local-world effects, among others. The majority of research in network science recognizes this fact and new models combining various of the above-mentioned network growth mechanisms are constantly being developed (e.g., Bauer & Kaiser, 2017; Bedogne’ & Rodgers, 2006; Bell et al., 2017; Callaway et al., 2001; Dorogovtsev et al., 2000a; Ferretti et al., 2012; Pham, Sheridan, & Shimodaira, 2015).

4.4. Growing phonological networks

A small but growing body of literature has focussed on network growth theories regarding semantic and phonological word forms in the mental lexicon of first-language users. The pioneering study in the field of semantic network growth was conducted by Steyvers and Tenenbaum (2005) who established a correlation between word learning probability and connectivity of words (i.e., degree) in a semantic network. They identified a process akin to preferential attachment responsible for word growth in children, where words were more likely learned if they connected to already known words in a child’s current lexicon, leading to the scale-free nature observed in the semantic network (Beckage & Colunga, 2019). The preferential attachment dynamic in semantic growth was also shown in paired-associate-learning tasks (memorizing word pairs) where well-connected words (=high-degree words) in the semantic network were recalled/recognized faster and more accurately (Mak & Twitchell, 2020). Later research on semantic network growth conducted by Hills and colleagues (Hills, Maouene, Riordan, & Smith, 2010; Hills, Maouene, Maouene, Sheya, & Smith, 2009a; Hills et al., 2009b) did not confirm these initial findings but identified a growth mechanism called "preferential acquisition" as the main driver for word learning in L1 English-learning children. As opposed to preferential attachment, preferential acquisition takes into account the connectivity of a semantic network in the fully-developed adult lexicon (see Hills et al., 2009b). When an evolving network represents an earlier stage of a known, fully-grown network, conclusions can be drawn as to the developmental trajectory from the earlier to the final network. In language learning this is naturally the case. The child's growing lexicon can be compared to the adult lexicon, and in fact, growth in the child's lexicon is strictly channelled toward the adult lexicon. The two linguistic growth algorithms of preferential attachment and preferential acquisition reflect fundamentally different concepts of lexical growth: preferential attachment suggests that future word knowledge is shaped by the learner's current knowledge, whereas preferential acquisition suggests that it is shaped by the input. In the case of preferential attachment, a new word links to a known word with many neighbors in the learner's current lexicon; in the case of preferential acquisition, a new word links to a highly-connected word in the adult lexicon of a language. As shown by the works of Hills and colleagues, children growing up in that linguistic environment are more likely to learn words that have more neighbors in the adult L1 English lexicon than words with low connectivity. In addition, Hills et al. (2009b) tested the network growth measure called lure-of-the-associates, which predicts that new words that can find many semantic neighbors to which they can potentially attach will be learned more efficiently than new words which have few semantic neighbors in a lexicon. This measure essentially counts the number of neighbors a new word can find in the existing lexicon of the learner but disregards information about how many neighbors these words have (which is the core feature of preferential attachment). The conclusion reached by Hills et al. (2009) that preferential acquisition is the most influential mechanism by which English-speaking children learn words in their first language was not supported by a later study by Beckage and Colunga (2019) investigating semantic and phonological network growth in first language acquisition of English-learning children. Rather, they found lure-of-the-associates to be the most influential factor for network growth. In general, their study was focussed on individual word learning of children with the result that learners were found to use different strategies for word learning, encompassing all three network growth algorithms (preferential attachment, preferential acquisition, lure-of-the-associates) in idiosyncratic ways.

Siew and Vitevitch (2020a) adapted the three growth measures from Hills et al. (2009) in the following way for their study growth of the L1 American English phonological network: “lure of the associates” came to represent the phonological docking words that new words encounter in a network, thus the phonological word forms to which new words can potentially attach. “Preferential attachment” took into account the phonological neighbors (=degree in the known lexicon of the learner) of each potential anchor word, while “preferential acquisition” counted the number of neighbors of words in the adult lexicon. Phonological networks were constructed for each vocabulary acquisition stage of children aged 3 to 9 years, and successive networks were compared in order to detect the node locations of greatest growth during the different vocabulary growth stages. Results showed that preferential attachment was the dominant mechanism in the early stages of word learning, with a process called “inverse preferential attachment” taking over in the later stages. The authors suggested that this could reflect an organizational structure of the mental lexicon where crowded phonological spaces cause delay in word retrieval and are thus avoided when the lexicon has reached a certain threshold word count (Siew & Vitevitch, 2020a).

In a similar vein, Fourtassi, Bian and Frank (2020) constructed phonological (and semantic) networks of children’s language acquisition in ten languages, including English. The study focussed exclusively on nouns and dealt with rather small networks (word counts ranged between 180 and 339 words). Due to the rarity of one-segment noun neighbors in the networks, the authors opted to include a 2-segment phonological distance score. These significant methodological differences prevent direct comparison with Siew and Vitevitch (2020b) and the present study. Fourtassi and colleagues calculated comparable measures of “preferential acquisition” and “preferential attachment” as Siew and Vitevitch, but they found preferential acquisition to be the main predictor for word learning across all ten studied languages. Specifically, well-connected words (=high-degree nodes) in the adult lexicon led to better word learning in children exposed to a language. Contrary to Siew and Vitevitch (2020b), preferential attachment had no effect. Fourtassi and colleagues suggest that the ambient learning environment has a larger impact on word learning in children than statistical algorithms based on the children’s current vocabulary, a finding in agreement with Hills et al. (2009). An explanation given by Fourtassi and colleagues is that caregivers put more emphasis on words which are connected in their own language, thereby increasing saliency and furthering the learning process in the child (also see Carlson et al., 2014). An interesting addendum provided by Fourtassi and colleagues (2020) was a cross-linguistic correlation of degree density in the ten languages. For instance, if the word *dog* is highly connected in the English lexicon, is the French word for dog, *chien*, also highly connected in the French lexicon? The answer was clearly 'no', indicating that phonological connectivity is different across languages and learners have to acquire knowledge about phonological neighborhood density in each new language they learn. This carries major implications for second language learners who cannot rely on neighborhood statistics from their first language. While it is conceivable that semantic networks show more overlap in node degree across different languages (a question not explored by Fourtassi et al., 2020), phonological networks can differ quite drastically in phonological neighborhood density. Language typology may play a role, since segmental probabilities and frequencies are important determiners of phonological relationships.

Recently, some researchers have opted to include a combination of semantic and phonological networks in so-called multiplex networks (Levy et al., 2021; Stella, 2020; Stella, Beckage, Brede, & de Domenico, 2018). Here, two (or more) layers of networks are linked with one another, and information spreading between the different networks can be tracked. While new insights can potentially be gained into word representations in the mental lexicon, this approach is not comparable to the single-layer networks of semantic or phonological networks. Multiplex networks address specific research questions that combine information on phonological and semantic word representations but remain limited by the fact that (in the English language) phonology and semantics are rarely related. Cases of polysemy where one phonological word form can have multiple meanings, such as [nəʊ] (i.e. "no", "know"), are special instances of activation spreading between phonological and semantic representations that can be observed in speech production (in aphasia, see Castro & Stella, 2018; Castro, Stella, & Siew, 2019). The role of homophones in language learning becomes elevated by these multiplex studies, and the implications of that for lexical access theories are unclear at the moment. Current theories of lexical processing can hardly account for a combinatorial effect of semantic and phonological similarity. An intriguing application of multiplex networks to second language learning would be the plotting of different languages of a speaker (first and other learned languages) on different layers of either semantic or phonological networks. Such an investigation could provide insights into the question of how the first language can prime or facilitate word learning in a second language, without the limitation of having to overcome conceptual differences between semantics and phonology. It is a well-established fact that meaning-based assumptions about words derived from the first language have an impact on another, learned language (e.g., Moreira & Hamilton, 2010). Similarly, phonological similarity between the first and the second language (or third language) have an influence on one another and can predict word learning (Marecka et al., 2021). Making visible and analyzable interactions between L1 and L2 in a multiplex network could yield interesting dynamics in relation to phonological or semantic transfer effects.

Chapter 5 will address the role of network growth algorithms in the measured vocabulary growth of the learner networks outlined in chapter 2. Growth rates of different network parts and communities and their distributions will be presented, followed by regression analysis to test the statistical significance of the investigated growth algorithms.

5. Growth of learner networks

5.1. Growing nodes in the learner networks

This chapter assesses growth in the ESL networks, its relation to network parts, communities, and individual nodes, and additionally, the contributions of various network growth algorithms and network centrality measures to growth patterns at different proficiency stages.[3]

5.1.1. Network parts

Growth was shown to take place primarily in the giant component of all proficiency levels, with singleton nodes displaying the slowest growth rate in all networks (see Figures 39a-e; reported tests are significant *Kruskal-Wallis ANOVAs* at $p>0.05$, with post-hoc *Wilcoxon signed rank tests*). In the beginning stages of language learning, giant component and island growth was rather pronounced but weakened as proficiency increased. This is a reflection of the macro-analytical trend to reduce giant component size over the course of language learning (see section 3.3.2.).

[3] Growth levels will refer to proficiency levels from which growth starts, for instance "A1 growth" references the vocabulary growth taking place to get from the A1 to the A2 level.

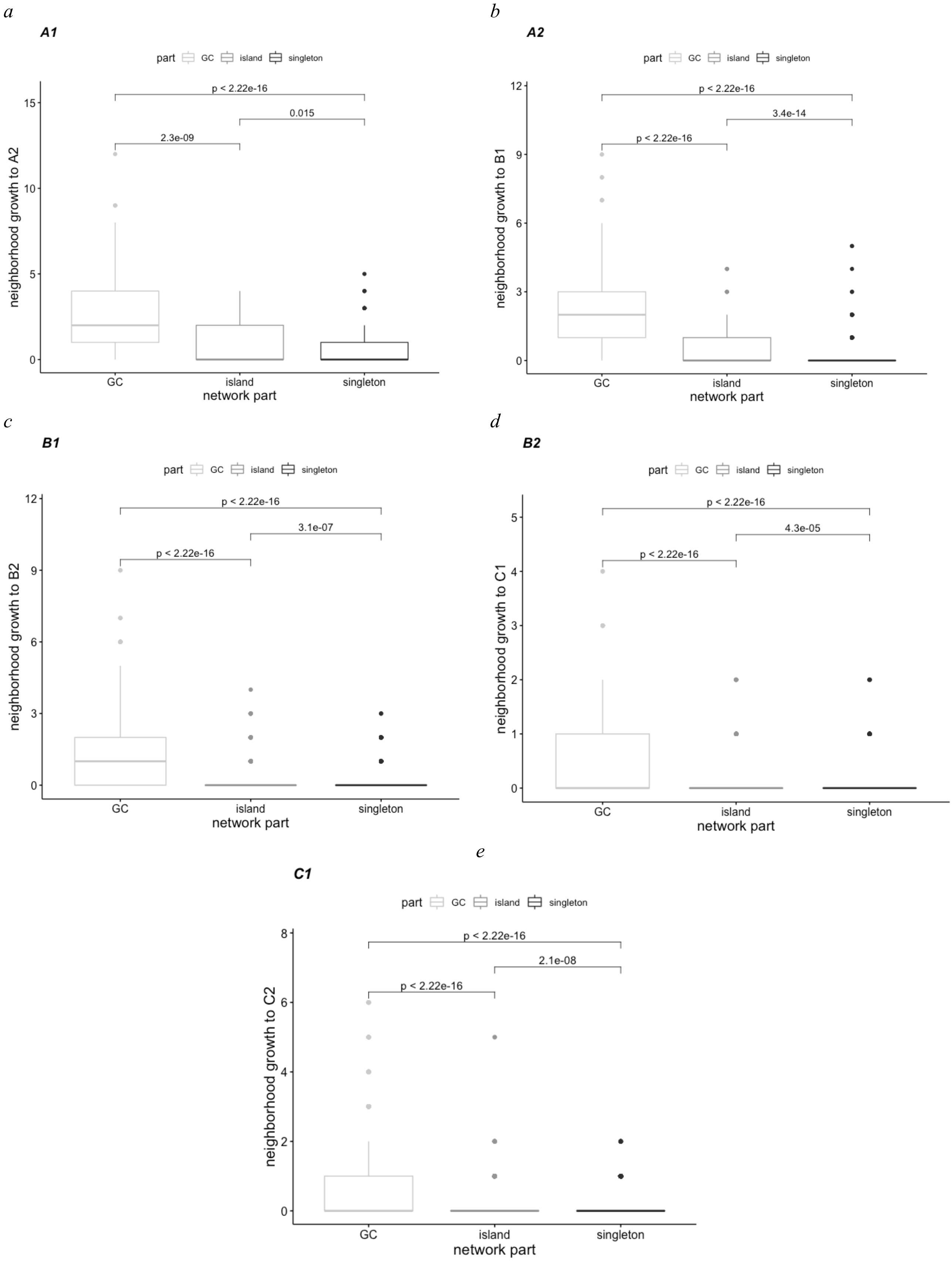


Figure 39a-e: Word growth (i.e., addition of new phonological neighbors) according to network parts in the learner networks.

Distributions of growth rates across the proficiency levels followed more egalitarian tendencies in the earlier networks (see Figure 40). Increasing inequality of growth opportunities became apparent starting at the A2 network, with the advanced learner networks being characterized by a large bias in the growth distribution, where a small number of nodes continued to grow, while the large majority of the nodes did not.

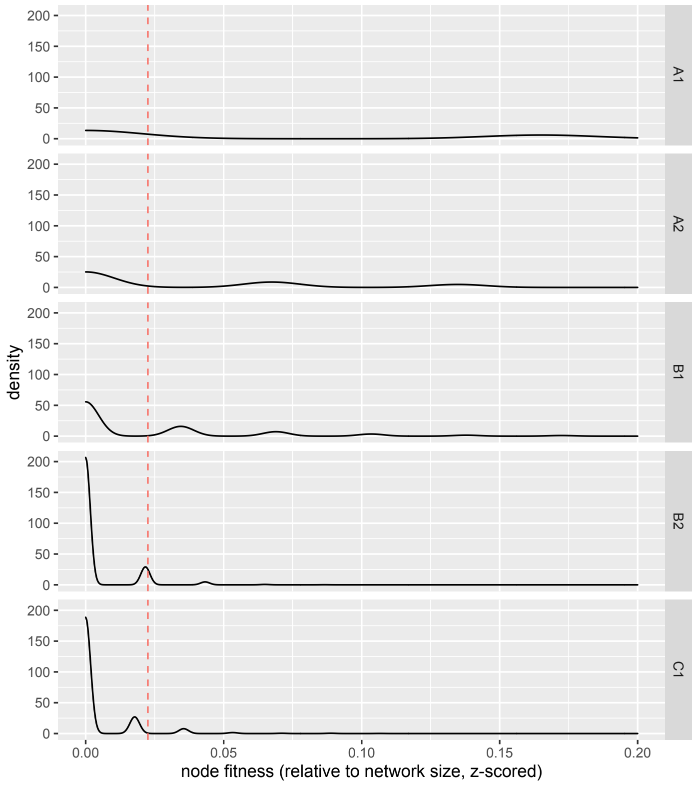


Figure 40: Growth distribution across the proficiency networks (heavy tails were shortened). Proficiency levels indicate starting point of growth development, i.e., 'A1 growth' refers to growth occurring between the A1 and A2 level.

Table 18 lists the results for the best approximations of growth distributions across the proficiency levels.

Table 18: Distribution types of growth spurts in the learner networks.

	A1	*A2*	*B1*	*B2*	*C1*
Best-fitting distribution	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)

The multimodal distributions indicate that each proficiency has more than one growth spurt mode and growth is not exclusively focussed on specific nodes. Table 19 lists the highest-growth nodes for each proficiency level and overall, during the course of ESL learning (from A1 to C1).

Table 19: Raw node growth rates of the highest-growth nodes per proficiency level (and overall).

A1	**A2**	**B1**	**B2**	**C1**	**Overall (A1-C1)**
WILL: 12	SEA/ SEE: 9	SEEM: 9	HOLE/ WHOLE: 4	DEEP: 6	FOR/ FOUR: 23
EIGHT: 9	FALL: 8	SCENE: 7	SIGN: 4	FOUR/ FOR: 6	RAIN: 19
KICK: 9	FOUR/ FOR: 8	SEAT: 7		KEEP: 6	EIGHT: 18
SIT: 9	FALL: 8	SELL: 7		DOOR: 5	WEAR/ WHERE: 18
WHY: 9	LOW: 8	DEAR: 6		LAW: 5	WILL: 18
HAIR: 8	BEAN/ BEEN: 7	FACE: 6		LEAD: 5	FAIL: 17
THEIR/ THERE: 8	MIGHT: 7	HEAD: 6		MORE: 5	FEEL: 17
WEAR/ WHERE: 8	RIGHT/ WRITE: 7	LIP: 6		NOR: 5	SEA/ SEE: 17
	SHORT: 7	RAIL: 6		PAW: 5	WHY: 17
	WAR: 7	RAISE: 6		POUR: 5	PAIR: 16
		SALE/ SAIL: 6		RED: 5	PAY: 16
		SET: 6		SHEEP: 5	SIT: 16
		TELL: 6		SHORE: 5	WHITE: 16
				WAR: 5	BEER: 15
				YOUR: 5	BUT: 15

To better understand which node characteristics contribute the most to node growth, a linear mixed effects model using the R package "lme4" and the function "lmer" (Bates, Maechler, Bolker, & Walker, 2014) was calculated to analyze how node growth (relative to network size) is affected by a number of fixed effect variables, namely phonemic length, lexical frequency rate, phonotactic probability, network centrality measures (degree centrality, closeness centrality, betweenness centrality, clustering coefficient, eigenvector centrality) and their weighted counterparts (weighted degree centrality). As random effects, "word" and "proficiency level" were entered into the model. Full models were compared with a corresponding null model lacking the fixed effect under investigation using a likelihood ratio test and the significance of the individual fixed effect was determined (Dobson, 2002; Forstmeier & Schielzeth, 2011). All variables were z-scored before being entered into the analysis.

The variables “degree” and “weighted degree” were correlated at r=-0.9 and combined as an interaction variable. Therefore, the first principal component (PC1) was computed via principal components analysis in order to combine the two variables into one that can account for the majority of the variance (see, e.g., Salem & Hussein, 2019). The first principal component (PC1) of “degree” and “weighted degree” was correlated at -0.71, with PC1 explaining 99% of the variance in the data. The computations were performed with R and the function “prcomp”. Due to high correlations with PC1, the two variables “betweenness centrality” (r=0.64) and “eigencentrality” (r=0.73) were removed from the models. The rest of the variables showed variance inflation factors of <1.8, as calculated with the R package “car” and the function “vif”), and collinearity did not appear to be an issue (Field, 2005; Quinn & Keough, 2002). The following pseudo code was used for the model:

```
growth rate ~ lexical frequency(log) + phonotactic probability + closeness centrality + clustering
coefficient+ PC1:degree & weighted degree+(1|word)+(1|proficiency level)
```

Results of the linear mixed models showed that node growth was primarily shaped by phonemic length, lexical frequency rate, phonotactic probability, clustering coefficients, and degree/weighted degree of a node (see Table 20).

Table 20: Results of the linear mixed model analysis.

Predictors	Estimate	SE	t	χ^2	p
(Intercept)	0.062	0.035	1.77		
Phonemic length	-0.0055	0.0008	-7.26	50.36	<0.001***
Lexical frequency	0.0034	0.0007	5.27	27.69	<0.001***
Phonotactic probability	0.0019	0.0006	3,17	10.02	0.002***
Closeness centrality	-0.0003	0.0006	-0.46	0.21	0.65
Clustering coeff.	0.0075	0.0007	10.45	107.85	<0.001***
PC1: Degree, weighted degree	-0.0051	0.0006	-9.09	51.04	<0.001***

For better illustration of the results (see figures below), growth rates were grouped according to how many neighbors each node grew from one proficiency level to the next. Number of neighbors were scaled to ensure comparability across the networks and different growth rates: words which grew a number of phonological neighbors above the 75th percentile were defined as “high” growth rate nodes, words that grew neighbors between the 25th and 75th percentile were defined as “mid” growth rate, and those that grew below the 25th percentile were defined as “low” growth nodes. Nodes that grew no new neighbors at all were classified as “zero” growth nodes.

Per linear mixed model analysis, the highest-growing nodes were of shorter phonemic length in all proficiency levels (see Figure 41). As the lexica of the beginning stages (A1, A2) generally contained shorter words, the effect of phonemic length on growth was less pronounced at those levels than in the advanced learning stages (C1). It has to be kept in mind that the raw numbers of phonological neighbor acquisition were the highest in A1 and A2 (up to 12 neighbors per word or twice as much as at the C1 level, see Table 19 above).

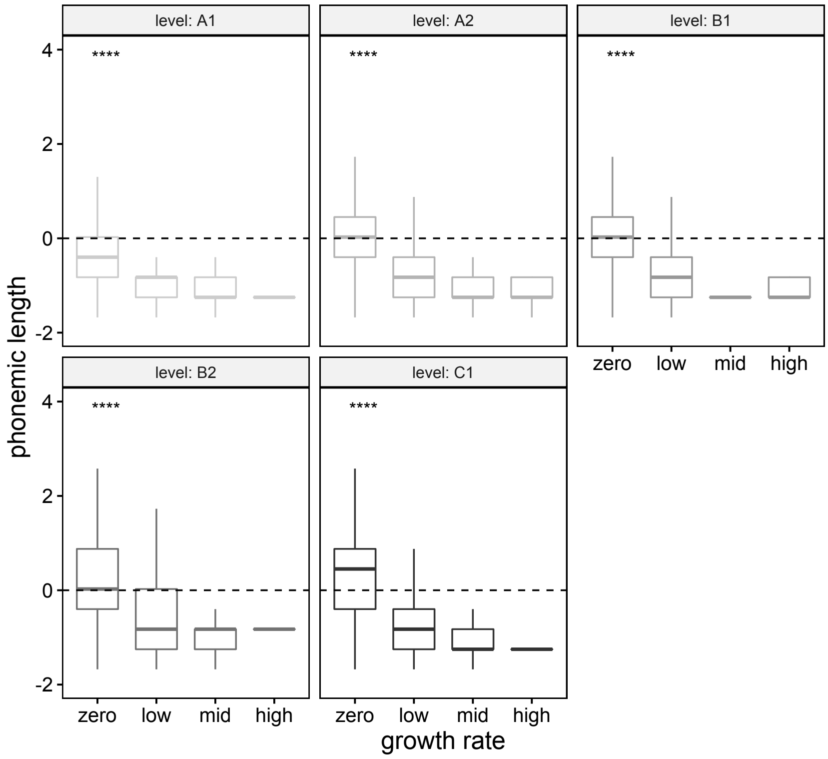


Figure 41: Standardized node growth in relation to phonemic length and per proficiency level. Kruskal-Wallis test statistics indicated by asterisks (within-proficiency-group comparisons).

Concerning lexical frequency rate, the highest-growing nodes came from the highest lexical frequency class (see Figure 42), and this was true for each proficiency level. While the A1 vocabulary showed above average lexical frequency rates, there was a general proficiency-wise progression toward lower lexical frequency rates for low-growth nodes and higher lexical frequency for high-growth nodes (which was most pronounced at the C1 level).

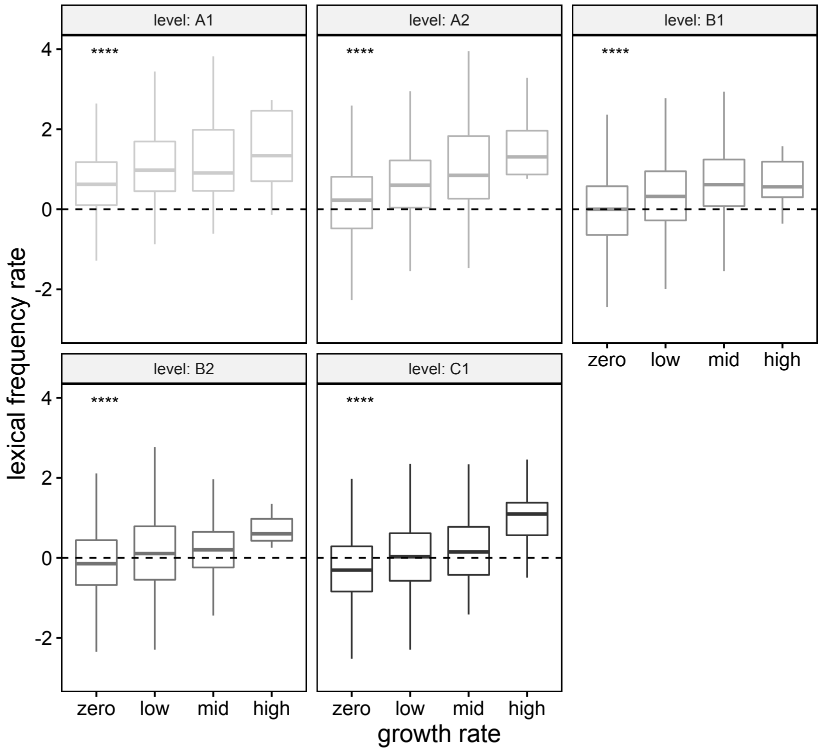


Figure 42: Standardized node growth and lexical frequency rate of words. Kruskal-Wallis test statistics indicated for each proficiency level.

Phonotactic probability had different effects on node growth at different proficiency levels (see Figure 43). High phonotactic probability aided phonological neighbor acquisition at the beginning A1 level but had the opposite effect at later proficiency stages. Node growth and phonotactic probability were only statistically significant in the levels A2 and C1, with the higher growth classes being characterized by low phonotactic probability. When few words exist in the lexicon, highly probable phonotactic combinations can amass without impairing word discriminability; at advanced stages of word learning, when the lexicon contains a large number of words, the "high probability disadvantage" (see Storkel et al., 2006, for L1 English child language acquisition) sets in and increases confusability of similar word forms. There could be a type of saturation effect where the lexicon has exhausted high-probability segments to such a degree that discriminability of words suffers. At such a point, low-probability phonotactics develop an advantage, as they aid discrimination of word forms.

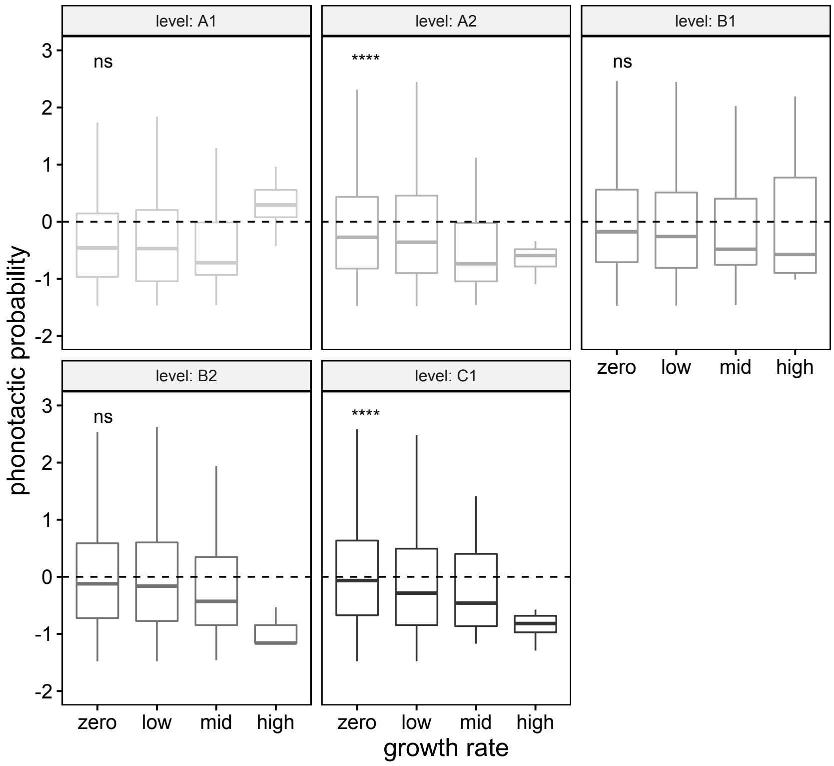


Figure 43: Standardized node growth and phonotactic probability of words. Kruskal-Wallis test statistics indicated for each proficiency level.

Growing nodes were generally characterized by high clustering coefficients at all proficiency levels, (see Figure 44), indicating a tendency to tightly interlink phonological neighborhoods through growth (i.e., phonological neighbors of target words becoming neighbors of one another). This finding supports previous studies of a word learning advantage when a new word becomes a member of a dense phonological neighborhood (see Gaskell & Dumay, 2003; Storkel et al., 2006).

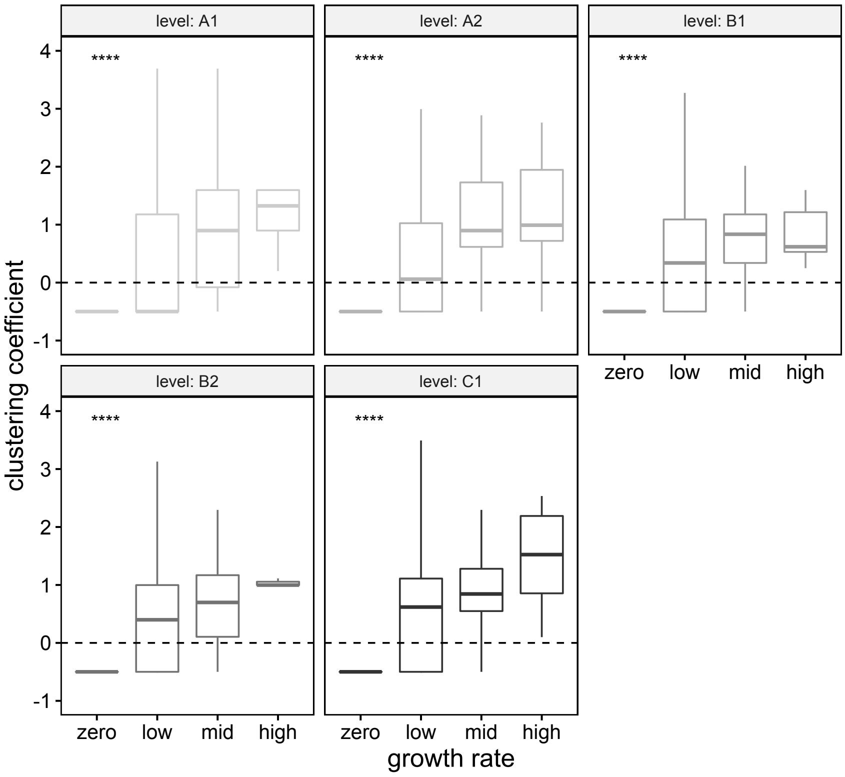


Figure 44: Standardized node growth and clustering coefficients of words. Kruskal-Wallis test statistics indicated for each proficiency level.

The interaction between node degree and weighted degree was influential for node growth, with high degree/weighted degree being characteristic of higher-growing nodes (see Figure 45).

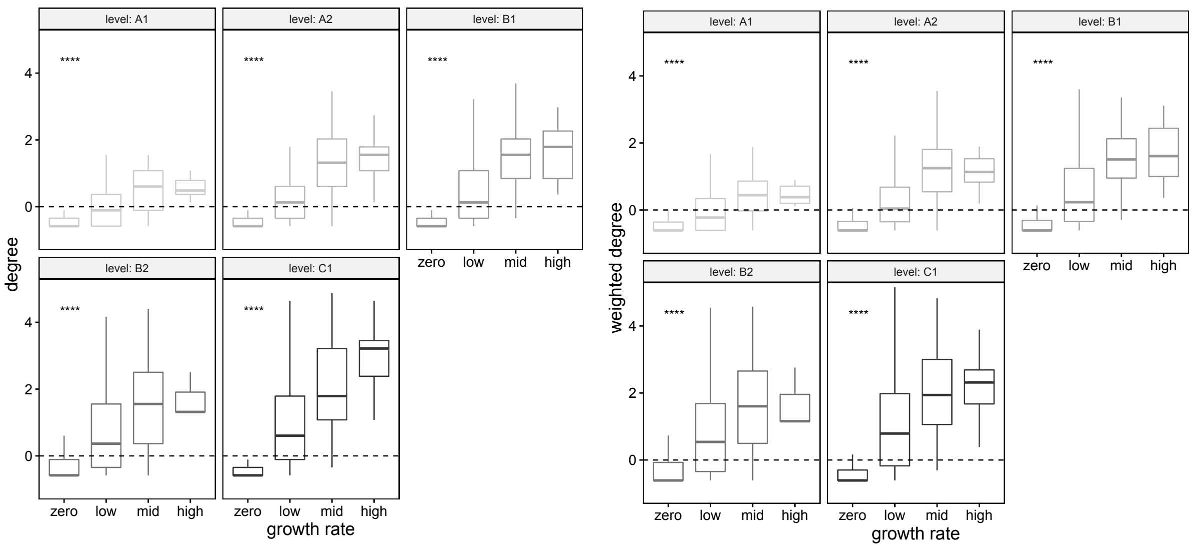


Figure 45: Standardized node growth and degree/weighted degree. Kruskal-Wallis test statistics indicated per proficiency level.

These results of the regression model closely correspond to the main factors contributing to word learning efficiency as derived from the literature: high lexical frequency, low phonotactic probability, high neighborhood density, and highly clustered neighborhoods of words are good indicators of their learning probability.

By virtue of their well-connectedness, the highest-growing nodes tend to reside in the giant component, with close to 100% of high-growth nodes being part of the giant component in all proficiency levels (see Figures 46a-e). In general, the giant component is characterized by a wider variety of node growth, encompassing nodes of all growth classes, whereas the islands and singletons mainly contain nodes of mid, low, and zero growth.

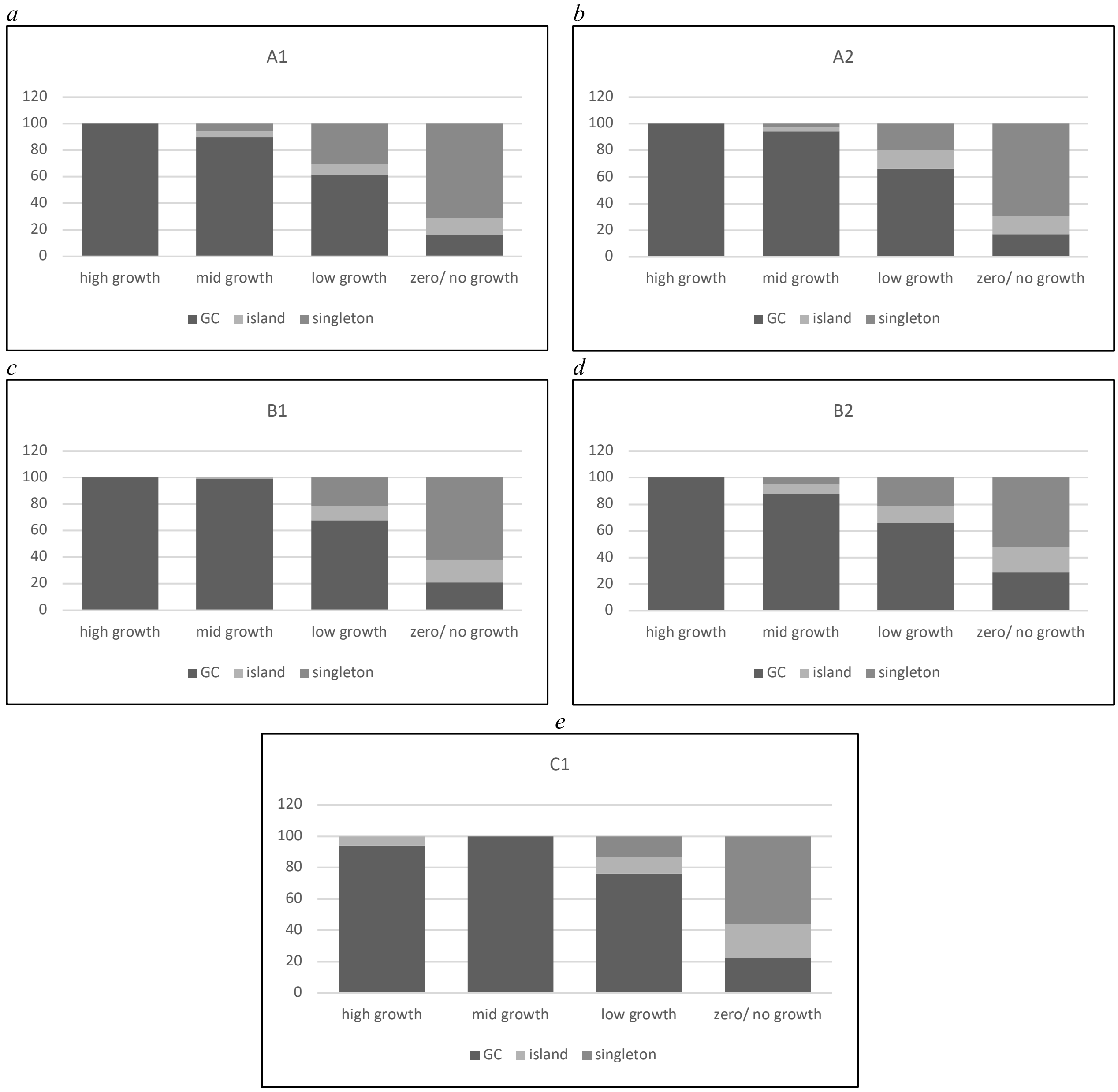


Figure 46: Node growth and network parts (giant component, islands, singletons).

While high growth rates are exclusively concentrated in the giant component in the beginning proficiency levels, high growth rates can also be observed in the C1 islands (see Figure 46e). The distribution of zero growth generally widens to encompass all three network parts to a larger extent as the networks grow.

5.1.2. Community size and growth

At the beginning stages of ESL learning, the majority of growth takes place in the mid-sized communities (see Figure 47), whereas in the advanced stages of language learning, growth is most prolific in the larger communities. Small communities tend to grow less (in relative terms) than larger communities. This fact could arise from the tendency of the larger communities to house words of different lexical characteristics (see section 3.4.). As mentioned in the community description in chapter 3, the beginning A1 network tends to put words of higher lexical frequency rate and shorter phonemic length in larger clusters (similar to what Siew, 2013, found for a phonological network of American English). As learning progresses, the larger communities start to receive fewer of those high-frequency, short words but become supplemented with phonemically longer and lower-frequency words. At this point of lexical development, community growth follows a different trajectory than initially. As can be seen in Figure 46, growth supports the formation of mid-sized and larger communities at first, but starting at level B2, it becomes more equally distributed across the differently-sized communities.

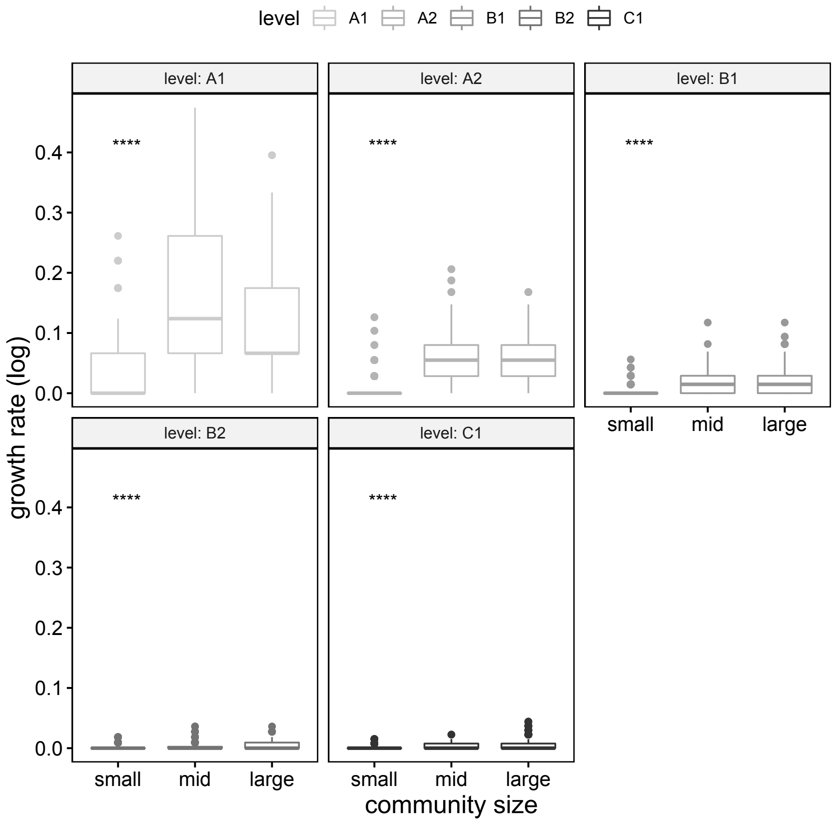


Figure 47: Standardized growth rate and community sizes across the proficiency levels. Kruskal-Wallis test statistics indicated per proficiency level.

Growth patterns of the beginning proficiency stages may lay the groundwork for mid-sized and larger community formation, with the later stages not contributing as much to community stratification any more. This finding is in line with what has been suggested by Siew (2013): larger communities are developed earlier in language learning and are a prerequisite for robust language development. They may provide the necessary scaffolding for lexical retrieval, which is largely dependent on similarity in phonological word forms (Vitevitch & Sommers, 2003).

5.2. Growth probability algorithms

5.2.1. Preferential attachment probability

The probability of a node to grow a new neighbor per growth cycle is proportional to the sum and degree of all nodes in the network, i.e.,

$$p_i = k_i / \sum_j k_j$$

Applied to the learner networks, this resulted in widely different distributions of the preferential attachment ('PA') value across the proficiency levels (see Figure 48). Preferential attachment probability is more evenly spread in the beginner networks (A1, A2), but develops into a highly competitive resource over time. At the C1 level, only few very prolific nodes have high preferential attachment probabilities. With advancing proficiency, few nodes stand out and accumulate growth opportunities through this network growth algorithm, while the vast majority of nodes lose out.

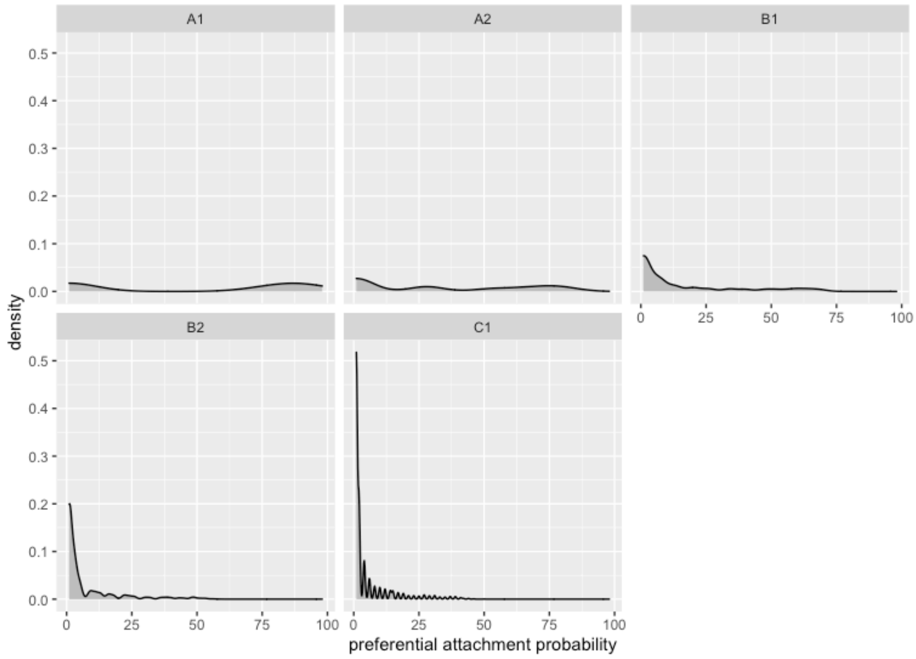


Figure 48: Probability of preferential node attachment in the different learner networks. PA values were normalized by 100.000.

The distributions of preferential attachment probability in the learner networks are presented in Table 21.

Table 21: Distributions of preferential attachment probabilities in the learner networks.

	A1	*A2*	*B1*	*B2*	*C1*
Best-fitting distribution	Bimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)

These results model the evolution of a bimodal distribution with two peaks in a distribution with multiple modes (A2) until the emergence of a comb-like multimodal distribution at the C1 level. While initially higher preferential attachment probabilities exist for two clusters of nodes in the A1 network, multiple peaks appear in the A2 level. At the B1 level, one peak accelerates growth (rich-gets-richer dynamic), reaching extreme preferential attachment probabilities at the C1 level. The fact that multiple smaller peaks develop in the advanced lexicon (B2, C1) may reflect the tendency of the phonological network to maintain at least a few more inclusive growth opportunities and avoid concentrating growth in an extremely limited number of nodes.

5.2.2. Fitness

Node fitness can be defined as the "inherent competitive ability (or the 'attractiveness') that a node has, which influences the rate at which it acquires links from other nodes as the network evolves over time" (Bell et al., 2017, p. 1). This growth-related fitness measure stems from node attributes that raise the likelihood of new neighbor attachment. In the case of phonological networks, the learnability of a word can be defined as its fitness. Based on previous psycholinguistic literature, attributes that make a word more likely to be acquired and retained by adult learners include

- high lexical frequency (e.g., Chen & Truscott, 2010)
- low phonotactic probability (e.g. Storkel et al., 2006)
- high neighborhood density or high node degree in network terms (e.g. Stamer & Vitevitch, 2012; Storkel et al., 2006)
- high clustering coefficient (Goldstein & Vitevitch, 2014)

Other well-known factors contributing to word learning, such as closeness to L1 (Poltrock, Chen, Kwok, Cheung, & Nazzi, 2018), context of word learning (Elgort, Beliaeva, & Boers, 2020), instructional setting (Luef, Ghebru, & Ilon, 2018), or phonological familiarity (Kaushanskaya, Yoo, & Van Hecke, 2014), among others, cannot be explored with the present dataset, as information pertaining to these variables is not available on EVP. In order to arrive at a measurement of node fitness for the present study, the variables "lexical frequency rate", "phonotactic probability (biphone-based)", "node degree", and "clustering coefficient" were combined as an interaction variable termed "fitness" by using principal components analysis for dimensionality reduction (Salem & Hussein, 2019). Composite z-scored principal components of the four variables were computed, and the first principal component was defined as the "fitness" variable. Results showed that the first principal component was correlated with degree at 0.64, while lexical frequency rate and clustering coefficient showed correlations of 0.42 and 0.62, respectively. Phonotactic probability showed a correlation coefficient of -0.17 (indicating a negative correlative relationship to the other variables, as predicted by Storkel et al., 2006). The first principal component accounted for 44% of the variance in the data. Figure 49 shows its distributional pattern at each proficiency level.

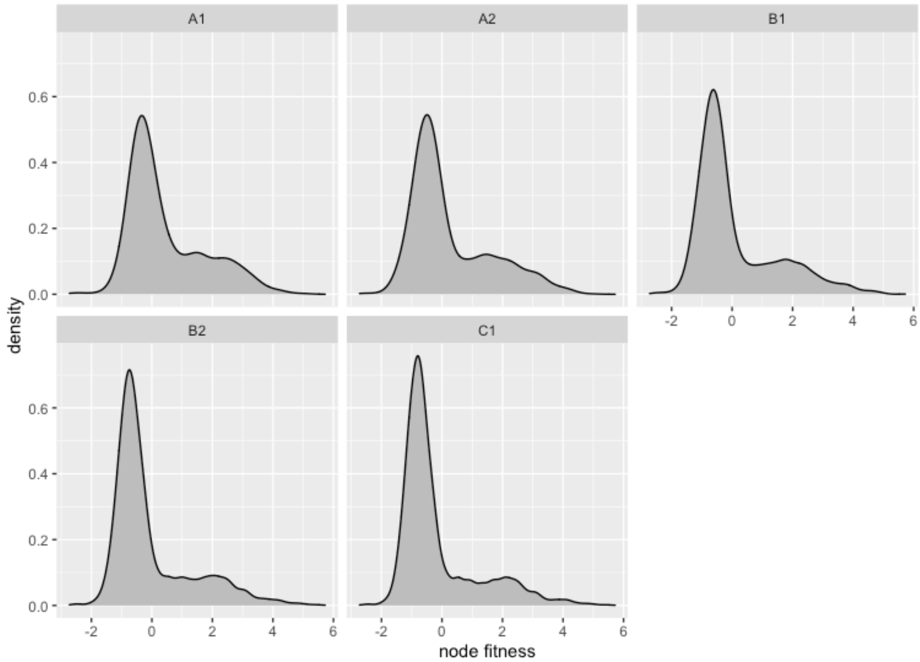


Figure 49: Density general node fitness across network (PC 1).

The distributions of node fitness in the learner networks are presented in Table 22. In order to be able to calculate the distributions, a value of "10" was added to each data point in order to shift the data to positive numbers. Hartigan's dip test yielded p>0.9, excluding the possibility that a distribution may be bimodal.

Table 22: Distribution types of node fitness of ESL learner network growth.

	A1	*A2*	*B1*	*B2*	*C1*
Best-fitting distribution	Power law	Power law	Power law	Power law	Power law

The distribution of fitness across all networks reflects the fact that few words possess all characteristics of improved word learning but that many words possess some of them. At the beginning stages of language learning, a larger percentage of words are of high fitness, and while their fitness is higher than that of the other words, the difference between high- and low-fitness words is less extreme. During the course of word learning, fewer high-fitness words gain higher fitness scores. Since all proficiency levels followed a power law distribution, a Bose-Einstein condensation was not tested.

5.2.3. Preferential attachment and fitness: PAFit

PAFit was calculated using the following equation:

$$p_i = \eta_i k_i / \sum_j \eta_j k_j$$

where the probability of node *i* to grow new links is the proportion of *i*'s fitness and degree (z-scored for the calculation) of the overall sum of nodes per network multiplied by those nodes' fitness values and degrees (both z-scored). The distribution of PAFit in the networks is exclusively focussed on few nodes, which gain scores over the course of learning (see Figure 50).

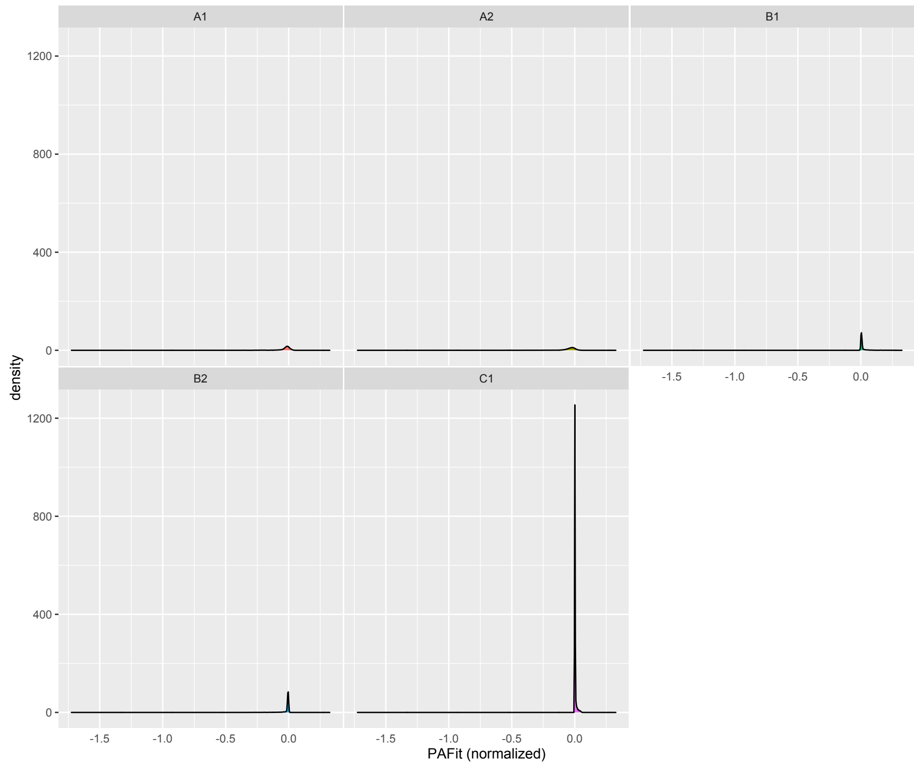


Figure 50: PAFit distributions across the learner networks. PAFit normalized by the factor of 1 mio.

The distributions of PAFit in the learner networks are presented in Table 23. A value of "1" was added to each data point in order to shift the data to positive numbers for distribution fitting.

Table 23: Distribution types of PAFit in the learner networks.

	A1	*A2*	*B1*	*B2*	*C1*
Best-fitting distribution	Burr	Log-logistic	Log-logistic	Weibull	Weibull

PAFit values are visibly highly concentrated on specific nodes in all of the proficiency levels. The advantage of the few prolific nodes increases substantially over the course of word learning, so that PAFit values increase exclusively in the nodes already showing high PAFit values from the beginning (i.e., A1 level).

5.2.4. Hybrid model of uniform and preferential attachment

The probability of random attachment in combination with preferential attachment is calculated by combining preferential attachment values per node with the proportion of the network the node represents (as the uniform attachment probability 'UA')

$$p_i = UA_i k_i / \sum_j UA_j k_j$$

The distribution of this hybrid variable is shown in Figure 51 (and table 24).

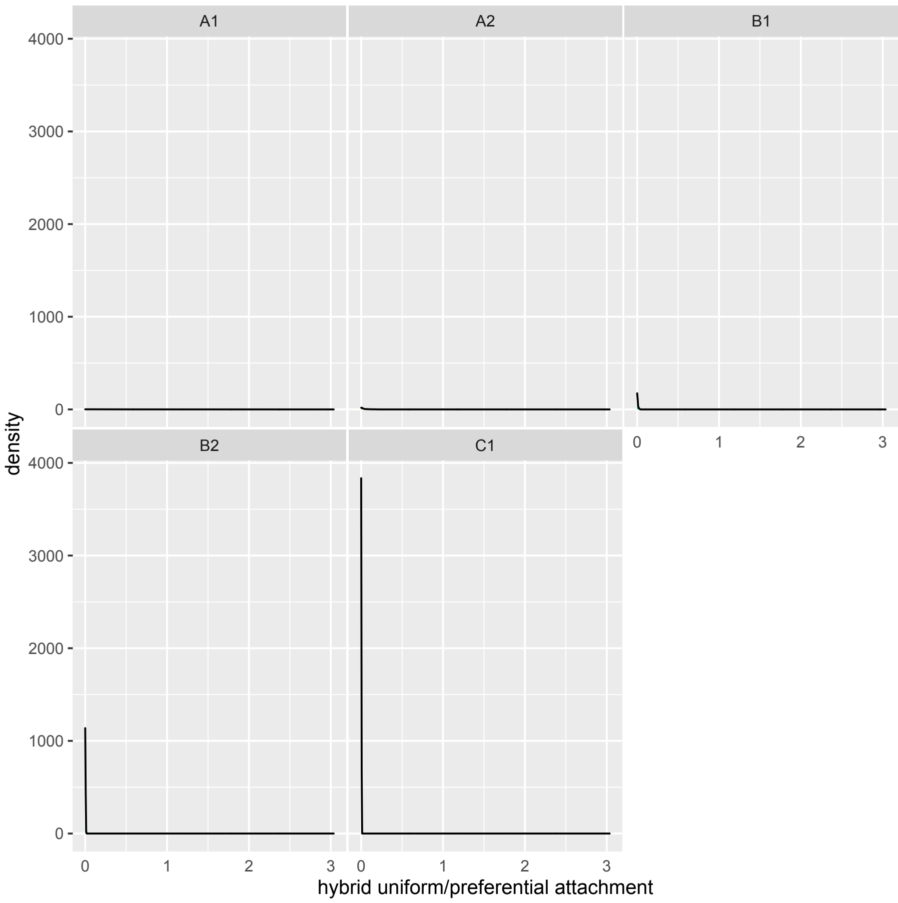


Figure 51: Distributions of the hybrid uniform/preferential attachment variable.

Table 24: Distributions of the hybrid uniform/preferential attachment variable.

	A1	*A2*	*B1*	*B2*	*C1*

Best-fitting distribution	Power law	Power law	Power law	Power law	Power law

Similar to PAFit, hybrid UA/PA values develop into an exclusive resource in all of the learner networks. This is due to their calculations where the effect of one attachment mechanism by itself is not accounted for but only their interaction (i.e., only non-zero values can become multiplied). While relatively low in the beginning stages, the hybrid UA/PA scores develop quite high values in later proficiency stages.

5.2.5. Statistical assessment

Rarely do singular explanations suffice to explain complex phenomena, and this certainly applies to linguistics and network science. Many network measures are conceptually or mathematically related to some degree (consider eigenvector and betweenness centrality or preferential attachment and PAFit), with changes in one often affecting changes in the other. It can therefore be reasonable to assume that more than one of the calculated network growth algorithms may be a relevant determiner for network growth dynamics or that multiple growth algorithms exert a concerted influence over a network, and this has indeed been implied by previous research (Fourtassi et al., 2020; Siew & Vitevitch, 2020a). Additionally, idiosyncratic principles of language learning may play a role, with different learners relying on different strategies for lexical network growth (Beckage & Colunga, 2019). Psycholinguistic investigations of lexical processing and word learning, as well, have shown that a myriad of factors – and their interactions – are responsible for observed effects. In order to account for a word learning effect of multiple of the network-related and psycholinguistic variables measured in the context of the present study, statistical analyses were performed to determine the most influential factors on growth.

Chapter 2 outlines a number of network centrality measures which have been shown to be influential in phonological networks, including degree centrality, weighted degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. For some of them, a role in network growth or word learning has been documented, such as degree and weighted degree, which can be viewed as proxy variables of neighborhood density. Denser neighborhoods, characterized by phonologically closer word forms, are known to become supplemented during word learning in first and second languages (Stamer & Vitevitch, 2012; Storkel, 2001, 2002; Storkel et al., 2006). High clustering coefficient is a measure of the interconnectedness of the phonological neighborhood and has been implicated in facilitated word learning (Goldstein & Vitevitch, 2014). Thus, words with high degree and weighted degree centralities, as well as high clustering coefficients are expected to possess higher growth potential in all networks. Closeness centrality, being an indicator of the phonological closeness of a target word to all others in the network (a fact known to incur a recognition advantage, see Goldstein & Vitevitch, 2014), can convey information about the extended phonological neighborhood going far beyond the one-segment metric. High closeness centrality words have many extended neighbors, and it can be hypothesized that in such cases word learning is facilitated by the dynamic of a "dense extended neighborhood" becoming supplemented in word learning, similar to the effect of degree centrality but extending beyond one degree of separation. In a similar vein, words of high eigenvector centrality may attract new neighbors by virtue of their denser extended neighborhoods of high degree words. Eigenvector centrality can account for the closer extended neighborhood, taking into account primarily the importance of the nearest neighbors of a target word; closeness centrality more loosely connects the farther reaches of the extended neighborhood. Degree (and weighted degree), closeness, and eigenvector centralities may all be contributing pieces of a rich-gets-richer dynamic in a network, where higher centrality measures predict more growth. If only the closest (one-segment) neighbors matter, degree and possibly weighted degree play an outsized role. If the two-hop neighbors are influential, this would show in eigenvector centrality playing a role in growth. If the larger three-and-more hop neighborhood plays a role, this would become evident in closeness centrality having an effect on growth.

Betweenness centrality can be viewed as conceptually related to internal linking of nodes (see section 4.3.3.) and the question is whether "bridge words" linking different neighborhoods have a tendency to strengthen their outstanding position through neighbor acquisition. As indicated by previous research (Vitevitch & Goldstein, 2014), such strategically placed key words in phonological networks tend to accrue activation benefits that result in lexical retrieval advantages, and it is conceivable that through their strengthened representation, words of high betweenness centrality are able to attract new neighbors.

In a first step, Pearson's correlation coefficients (r) were calculated to determine whether the network growth algorithms and the network centrality measures, together with the well-known variables for word learning (lexical frequency, phonemic length, phonotactic probability), are predictive for actual network growth. They are listed in Table 25 (see also Figure 52).

Table 25: Pearson's correlation coefficients of variables with growth rate per proficiency level (dark red cells r>0.6; light red cells r>0.5).

	A1	A2	B1	B2	C1	Mean
PA/ hybrid UA/PA*	0.49	0.69	0.56	0.36	0.51	0.52
PAFit	-0.21	-0.52	0.39	-0.24	0.39	-0.04
Fitness	0.41	0.63	0.51	0.32	0.46	0.47
Degree	0.49	0.69	0.56	0.36	0.51	0.52
Lexical frequency	0.23	0.3	0.22	0.11	0.16	0.2
Phonotactic prob.	0.06	-0.08	-0.02	-0.003	-0.06	-0.02
Phonemic length	-0.43	-0.51	-0.44	-0.26	-0.37	-0.4
Weighted degree	0.51	0.68	0.55	0.37	0.5	0.52
Closeness centr.	0.07	0.06	0.03	0.01	-0.01	0.03
Betweenness centr.	0.43	0.42	0.32	0.32	0.34	0.37
Clustering coeff.	0.32	0.48	0.39	0.22	0.31	0.34
Eigenvector centr.	0.19	0.5	0.42	0.2	0.34	0.33
Mean	***0.21***	***0.28***	***0.29***	***0.15***	***0.26***	
Highest r	***0.51***	***0.69***	***0.56***	***0.37***	***0.51***	
Lowest r	***0.06***	***0.06***	***-0.02***	***-0.003***	***-0.01***	

*The two variables are correlated at r>0.9 at each proficiency level.

The strongest overall correlations with growth rates were calculated for preferential attachment (highly correlated with hybrid UA/PA: r=0.9), degree, and weighted degree (r>0.5), and this indicates that rich nodes may get richer in phonological neighborhood growth. As graphically displayed in Figure 52, all correlations are largely positive, with the exception of phonemic length, where shorter words tend to grow more neighbors, and PAFit, where the A1, A2, and B2 levels grow by lower PAFit values, while higher values prevail in the other levels.

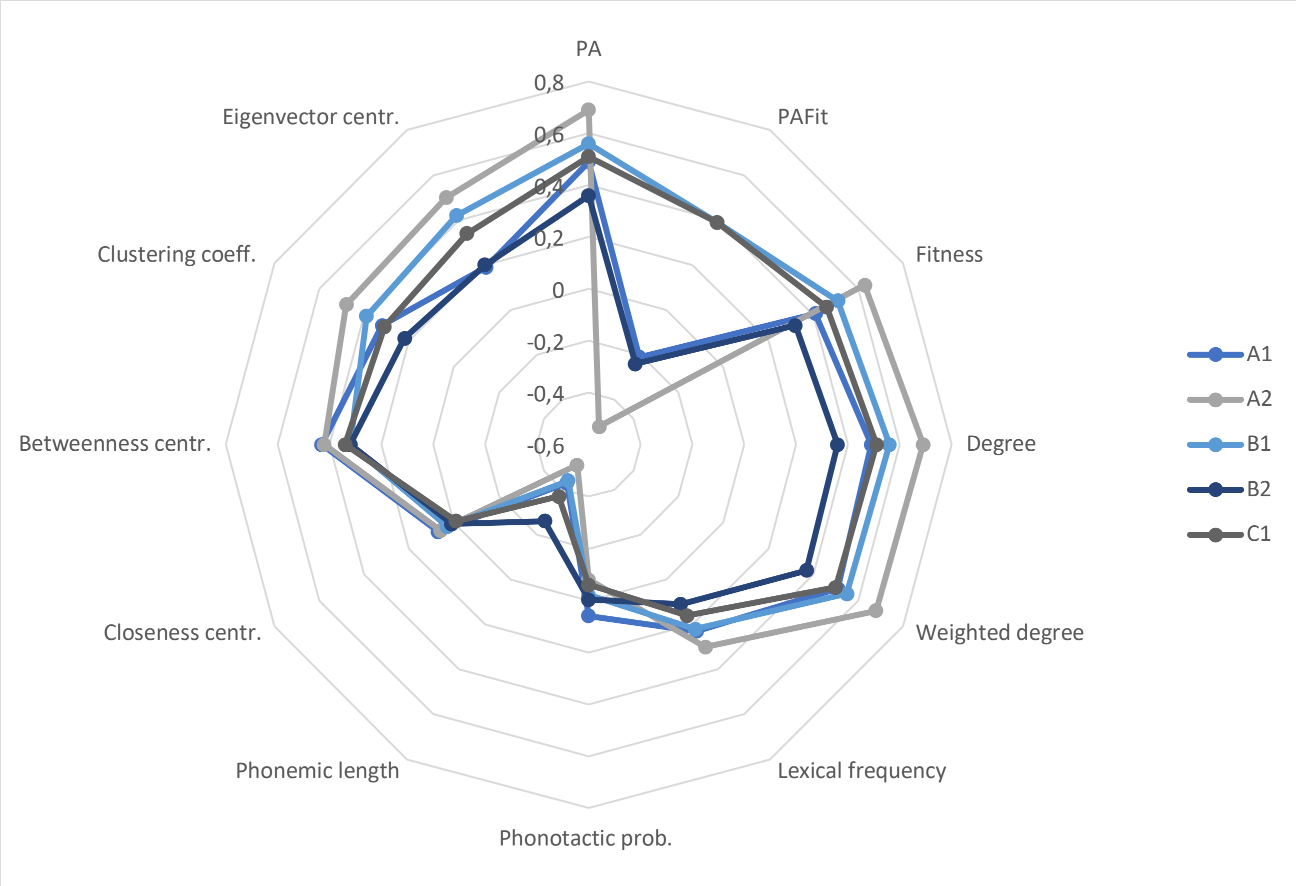


Figure 52: Correlation coefficients (Pearson's r) of network growth variables with growth rate per proficiency level.

It is evident that the correlations of some network variables with growth differ by proficiency level. The A2 level leads in terms of the magnitude of the correlations involving the majority of the variables, while the B2 level lags on many of them (see Table 25 and Figure 52). Specific learning dynamics may take place at various proficiency levels that may be responsible for the observed correlation dynamics (e.g., lower PAFit scores in lower proficiency levels). In addition, the utility of certain network growth measures may depend on the size of the vocabulary at a given point in relation to the learning environment of yet-to-be-learned words, which shrinks as the networks grow (see chapter 6 for an analysis of this issue).

To determine which network variable has the largest impact on node growth at any given proficiency level, series of linear mixed effects regression models with growth rate as the outcome variable (i.e., number of new words that each lexicon entry grows to the next proficiency level) were fitted. The following independent variables (fixed effects) were z-scored and included:

- preferential attachment probability

- node fitness (first principal component)
- PAFit
- phonemic length
- lexical frequency
- phonotactic probability
- degree*weighted degree centrality
- closeness centrality
- betweenness centrality
- eigenvector centrality
- clustering coefficient

As random effect, the variable 'network part' was included, accounting for unexpected influences that may be related to the part of the network a word belongs to (i.e., giant component, islands, singletons). Random slopes could not be included due to resulting convergence issues. The following pseudo code was used:

```
growth ~ PA+fitness+PAFit+phonemic length+lexical frequency+phonotactic
probability+degree*weighted degree+closeness+betweenness+eigenvector+clustering coefficient
+(1|network part)
```

The R package "lme4", with the function "lmer" (Bates et al., 2014), was used to perform linear mixed effects analysis models of the relationship between the fixed effects and the growth rate per word from one proficiency level to the next one up. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question (Dobson, 2002; Forstmeier & Schielzeth, 2011). Multicollinearity testing was performed with the function "vif" of the R package "car" to ensure variance inflation factors remained low (Field, 2005; Quinn & Keough, 2002). In case of correlations of >0.6 between variables, one of the variables was removed in a model. Due to its high correlation with preferential attachment in all models, the variable hybrid UA/PA was removed from each model. Results of the mixed model analyses are presented separately for each ESL proficiency level.

In the A1 network, lexical frequency rate and clustering coefficient were removed, as both were highly correlated with one another (r=0.8) and node fitness (r=-0.9). In addition, phonotactic probability was correlated with fitness (r=-0.72). Of the remaining variables, preferential attachment, phonemic length, and closeness and betweenness centralities were significant contributors to network growth (see Table 26). Specifically, high preferential attachment probability and high betweenness centrality, as well as short phonemic length and low eigenvector centrality, led to higher growth rate in A1 nodes.

Table 26: Results of the linear mixed effects regression model of the A1 network growth variables.

	Estimate	Std. Error	t value	χ2	p
(Intercept)	1.20284	0.22949	5.241		
PA probability	1.40931	0.54620	2.580	18.02	<0.001***
PAFit	0.42792	0.23253	1.840	3.38	0.07
Fitness	0.24336	0.14270	1.705	2.9	0.09
Phonemic length	-0.36294	0.11568	-3.138	9.77	0.002**
Degree*weighted degree	0.05806	0.28030	0.207	0.41	0.81
Closeness centrality	-0.08952	0.08098	-1.106	1.22	0.27
Betweenness centrality	1.06728	0.33410	3.194	10.12	0.001**
Eigenvector centrality	-0.34313	0.14608	-2.349	5.49	0.02*

A comparison of standardized regression coefficients showed that preferential attachment and eigenvector centrality were the most influential independent variables in the regression model (see Figure 53).

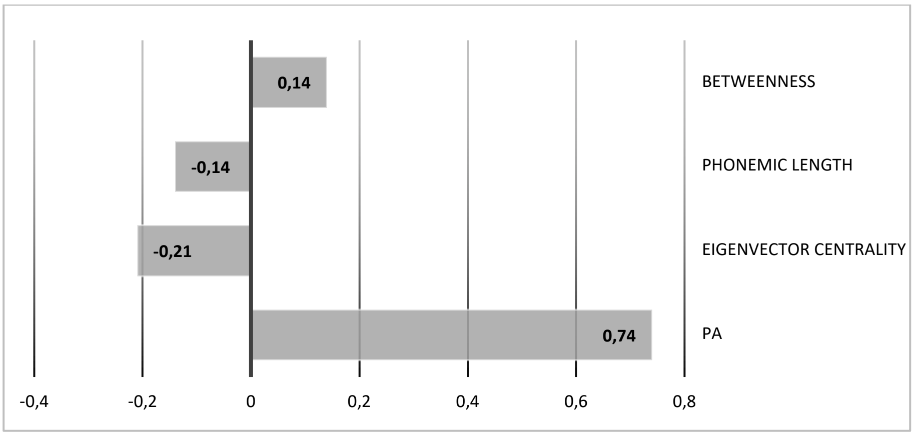


Figure 53: Standardized regression coefficients of significant network growth variables in the A1 network.

In the A2 network, phonotactic probability and lexical frequency rate were removed, as they correlated with one another (r=0.6) and with fitness (r=-0.8). PAFit correlated with preferential attachment (r=0.73) and the interaction degree*weighted degree (r=0.88), thus PAFit was removed from the model. In the A2 network, the majority of network growth variables were highly significant predictors for lexical growth to the B1 level (see Table 27, Figure 54).

Table 27: Results of the linear mixed effects regression model of the A2 network growth variables.

	Estimate	Std. Error	t value	χ2	p
(Intercept)	1.07333	0.05668	18.937		

PA probability	18.937	0.25134	6.761	175.4	0.001***
Fitness	0.01024	0.07902	0.130	0.017	0.89
Phonemic length	-0.21353	0.04652	-4.590	20.92	<0.001***
Degree*weighted degree	-0.18185	0.05562	-3.270	13.1	0.001**
Closeness centrality	-0.10529	0.03412	-3.086	9.5	0.002**
Betweenness centrality	-0.21632	0.09911	-2.183	4.76	0.03*
Eigenvector centrality	-0.15096	0.06423	-2.350	5.51	0.018*
Clustering coefficient	0.01503	0.04285	0.351	0.12	0.73

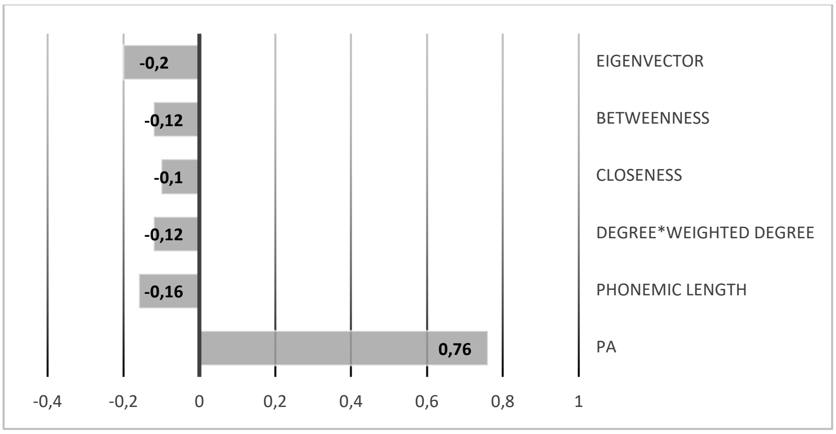


Figure 54: Standardized regression coefficients of significant network growth variables in the A2 network.

Variables removed from the B1 network included phonotactic probability and lexical frequency rate (correlated at r=-0.8 with fitness). Again, preferential attachment was a very significant predictor for lexical growth, together with the majority of variables with the exception of betweenness centrality (see Table 28, Figure 55).

Table 28: Results of the linear mixed effects regression model of the B1 network growth variables.

	Estimate	**Std. Error**	**t value**	**χ2**	**p**
(Intercept)	0.75578	0.03209	23.550		
PA probability	1.05182	0.19826	5.305	183.5	<0.001***
PAFit	-0.34913	0.07729	-4.517	20.33	0.001***
Fitness	0.19270	0.04657	4.138	17.1	0.001***
Phonemic length	-0.08894	0.02470	-3.600	12.93	<0.001
Degree*weighted degree	-0.16823	0.02888	-5.825	48.4	<0.001***
Closeness centrality	-0.10242	0.01859	-5.508	6.003	0.014*
Betweenness centrality	-0.02920	0.03033	-0.963	0.9	0.34
Eigenvector centrality	0.25415	0.04492	5.658	31.8	<0.001***
Clustering coefficient	-0.09076	0.02568	-3.534	12.5	<0.001***

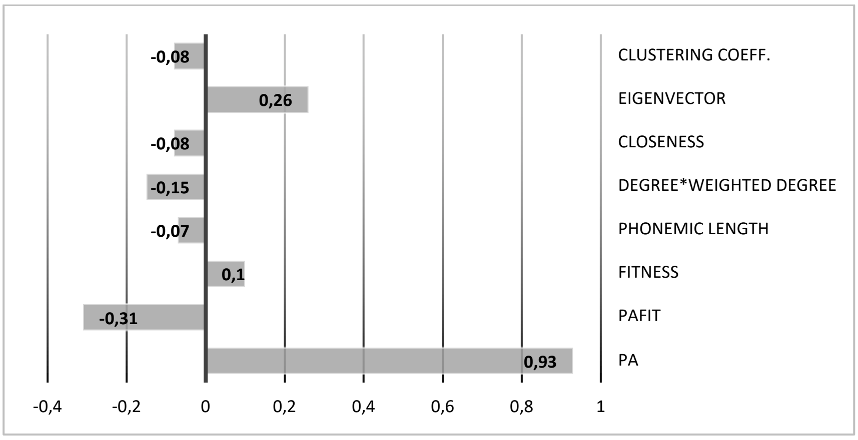


Figure 55: Standardized regression coefficients of significant network growth variables in the B1 network.

In the B2 network, degree was correlated with PAFit (r=0.72), and phonotactic probability and lexical frequency were correlated with fitness (r>0.7). Degree*weighted degree, phonotactic probability, and lexical frequency were removed (see Table 29, Figure 56).

Table 29: Results of the linear mixed effects regression model of the B2 network growth variables.

	Estimate	**Std. Error**	**t value**	**χ2**	**p**
(Intercept)	1.1683662	0.0061300	190.599		
PA probability	0.2661017	0.0216774	12.276	126.04	<0.001***
PAFit	0.0783564	0.0227093	3.450	11.89	>0.001***
Fitness	0.0320535	0.0178999	1.791	3.2	0.07
Phonemic length	-0.0067843	0.0080066	-0.847	0.72	0.4
Closeness centrality	-0.0156975	0.0061455	-2.554	5.76	0.02*
Betweenness centrality	0.0236926	0.0086342	2.744	7.52	0.006**
Eigenvector centrality	-0.0837386	0.0168937	-4.957	24.5	>0.001***
Clustering coefficient	0.0009425	0.0090544	0.104	0.01	0.9

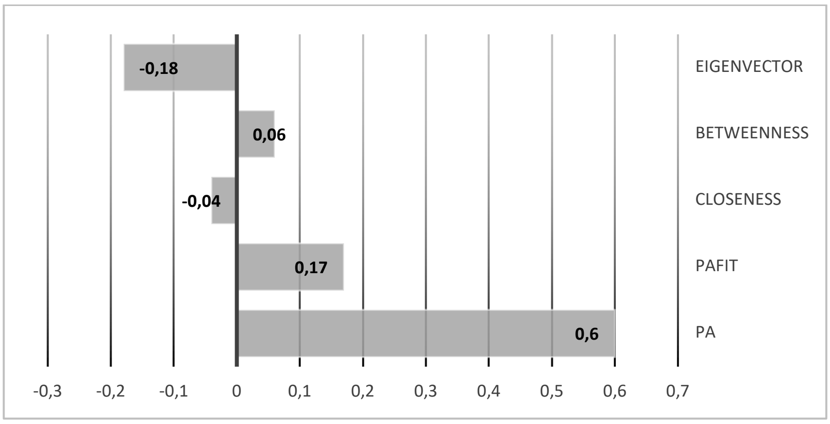


Figure 56: Standardized regression coefficients of significant network growth variables in the B2 network.

Variables removed from the C1 analysis included lexical frequency rate and phonotactic probability (both correlated with fitness at r>0.72), as well as degree, which correlated with PAFit at r=0.72 (see Table 30, Figure 57).

Table 30: Results of the linear mixed effects regression model of the C1 network growth variables.

	Estimate	Std. Error	t value	χ2	p
(Intercept)	-2.417e-02	1.588e-02	-1.522		
PA probability	8.213e+06	4.241e+05	19.366	260.9	<0.001***
PAFit	8.817e-03	2.580e-02	0.342	0.12	0.73
Fitness	8.768e-03	2.052e-02	0.427	0.18	0.67
Phonemic length	-4.124e-02	8.651e-03	-4.767	22.68	<0.001***
Closeness	-3.774e-02	6.787e-03	-5.561	6.82	0.009**
Betweenness	-4.327e-02	8.867e-03	-4.880	23.76	<0.001***
Eigenvector centrality	-1.873e-01	2.074e-02	-9.029	80.95	<0.001***
Clustering coefficient	-6.781e-03	1.043e-02	-0.650	0.42	0.52

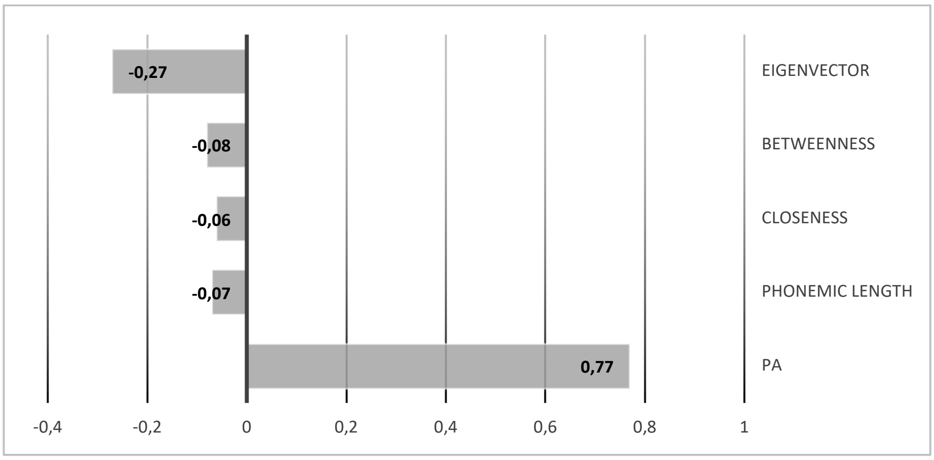


Figure 57: Standardized regression coefficients of significant network growth variables in the C1 network.

In all networks, the effect of preferential attachment was ubiquitous and overwhelming. Eigenvector centrality was also significant at all proficiency levels, whereas the effects of the other variables were more specifically tied to certain proficiency levels (see Figure 58). In general, the results clearly show that more than one network growth dynamic is involved in phonological growth of the learner networks.

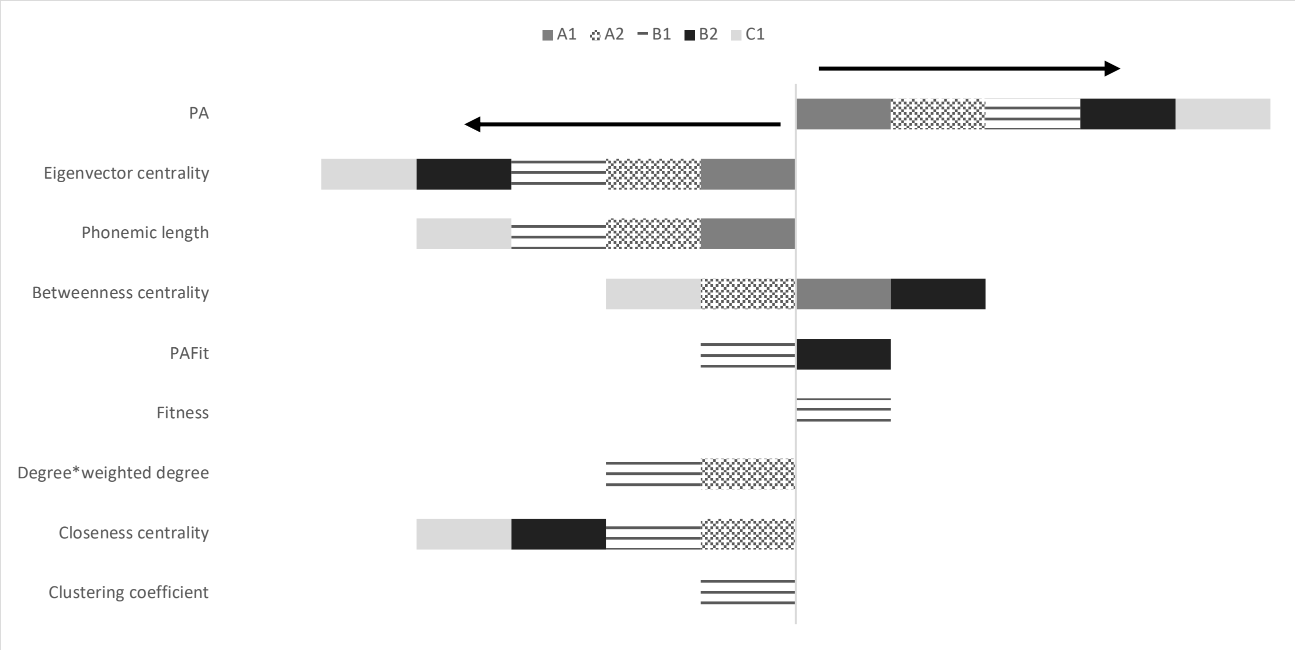


Figure 58: Significant network growth variables in the learner networks and their direction of influence on growth.

For each significant network growth variable, the direction of its influence on growth rate is shown in Figure 58. High preferential attachment scores predicted high growth in all proficiency levels, displaying the trend of rich words getting richer. Eigenvector centrality had the inverse effect and low values predicted growth in all networks. Closeness centrality displayed a similar trend and low values were predictive for growth. The implication here is that network growth occurs primarily by adding new words to those with sparser extended neighborhoods (i.e., low eigenvector and closeness centralities) across all proficiency networks. One-degree neighborhoods tend to grow denser through learning over time, while extended neighborhoods (a few hops removed) preferentially remain sparse. This growth dynamic (schematized in Figure 59) leads to a network structure where large hubs are surrounded by sparse neighborhoods (compare to “hub-and-spoke” typology of networks, see section 4.1.1.1.).

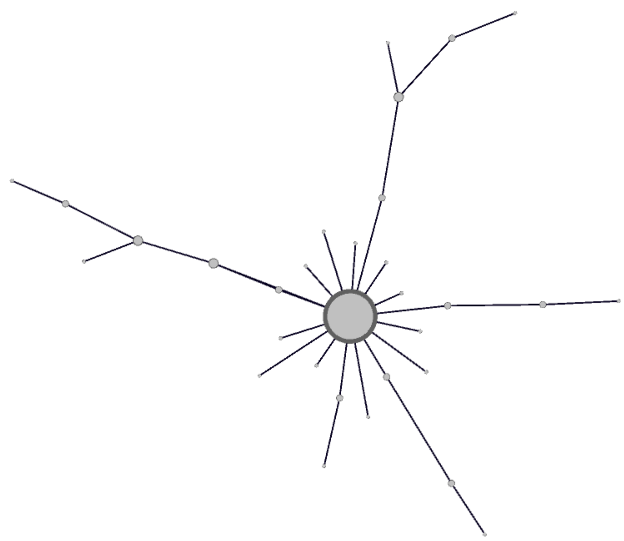


Figure 59: Network structure favored by growth according to preferential attachment in combination with low closeness and eigenvector centralities: dense phonological neighborhoods (i.e., larger hubs) surrounded by a sparser extended neighborhood.

When betweenness centrality was significant (A1, A2, B2, C1), the direction of influence varied depending on the proficiency level: high betweenness centrality furthered growth at the A1 level but impeded growth at the C1 level. Nodes of high betweenness centrality in the beginning A1 level had less opportunity to benefit from a lot of co-activation as compared to such nodes in the C1 network, as activation is assumed to accrue over time (Chan & Vitevitch, 2009; Goldstein & Vitevitch, 2014). In initial build-ups of networks, there may be a tendency to avoid words of high betweenness centrality, as they represent a particular vulnerability of networks due to their central positions, and to strengthen such frail linking points in a network by growth. In larger networks, more numerous links exist between all nodes in the network, relieving some of the burden placed on nodes with high betweenness centralities. In such a case, growth at key betweenness centrality positions may become less prevalent.

The fact that node fitness and PAFit only played a minor role in growth rules out the possibility that phonological networks follow fitness models of network growth. While these models offer intriguing hypotheses for lexical learning by focussing on word characteristics, such as lexical frequency and phonotactic probability, the results of the present study clearly demonstrate the superiority of preferential attachment models for phonological network growth.

5.3. Developmental growth trends

5.3.1. Growth fluctuations: Aging effects

There are nodes which grow only once during their lifetime in a lexicon and those which show more consistent growth patterns. Of the latter group, a question concerns the developmental pattern of growth. Specifically, at which point of their tenure in a lexical system do words show their largest growth potential? In addition, at what age in a lexical system do words attract the most neighbors? To address these issues, a sub-analysis was performed to test the growth development of the continuously growing nodes, which keep adding neighbors from their introduction to the system to the final ESL proficiency level C2. Figure 60 displays a declining trend in growth rate in relation to word age. Continuously growing words go through their largest growth spurt when they initially appear in a lexicon, at which point their growth always starts to decline. A word's growth rate is thus highly predictable by their point of entry into the lexical system. Initial growth lags cannot be overcome by growth spurts at a later time.

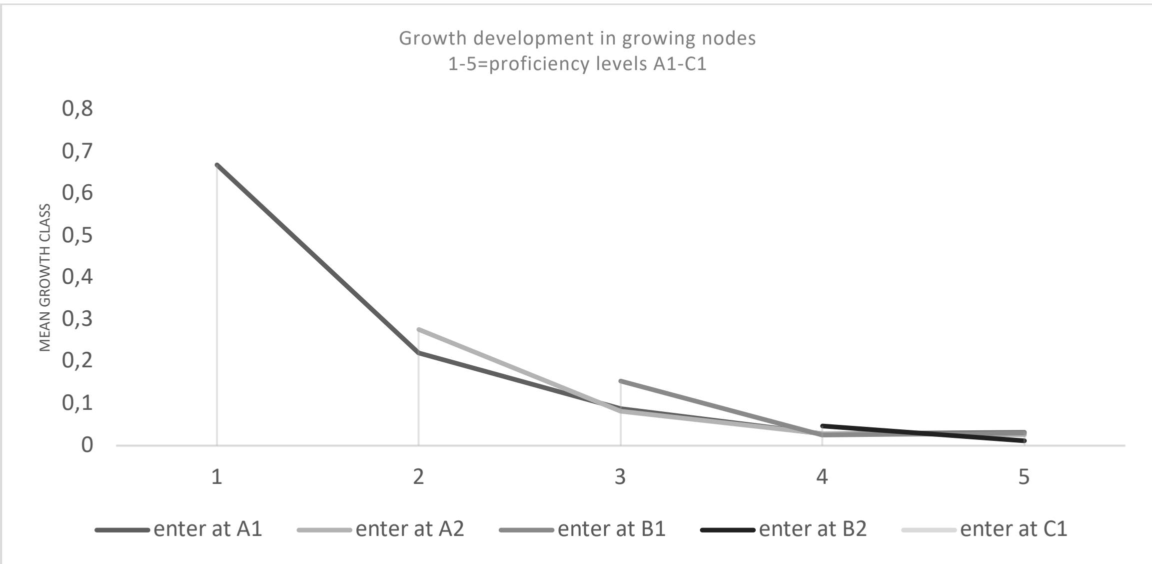


Figure 60: Growth patterns of consistently growing nodes from their entry points in the lexicon until C1.

This developmental growth dynamic indicates a preference for new words to attach to dense neighborhoods with many neighbors initially (a known tendency in word learning), thereby gaining the maximum number of neighbors upon entry in the system. This largest initial growth potential declines with higher proficiency level, and after some time in the lexicon, words decrease their growth potential, indicating an aging effect in the phonological networks of ESL learners. This could reflect a tendency to learn phonologically similar words (=phonological neighbors) at the same time.

5.3.2. Initial attractiveness

The likelihood of a word to acquire at least one neighbor in the learning process from A1 to C1 is generally not high. The majority of singleton words never grow any neighbors and remain singletons throughout all proficiency levels. Of 161 singleton nodes that first appear in the A1 network, only 24 will eventually grow a neighbor by the C1 level; 85% will remain unconnected. Of the 532 singletons that appear in the A2 network, 390 (=74%) will remain singletons; and of the 1017 singletons in the B1 network, 854 (=84%) will never grow a neighbor. The B2 network houses 1529 singletons of which 1357 (=89%) will remain singletons in the C1 network. There are a number of lexical characteristics which raise the initial attractiveness of a word, thus making it more likely to gain a neighbor. Phonemically shorter words tend to have more opportunities to grow at least one neighbor (see Figure 61).

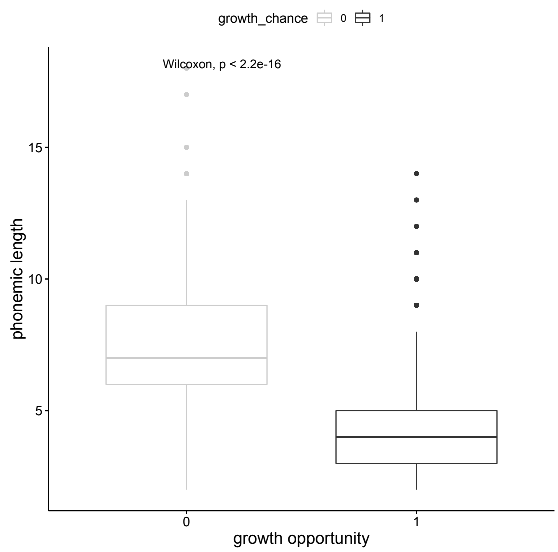


Figure 61: Short words are more likely to eventually grow a neighbor.

Phonotactic probability is another good indicator of growth opportunities, and words containing lower phonotactic probability biphones are more likely to eventually acquire at least one neighbor (see Figure 62).

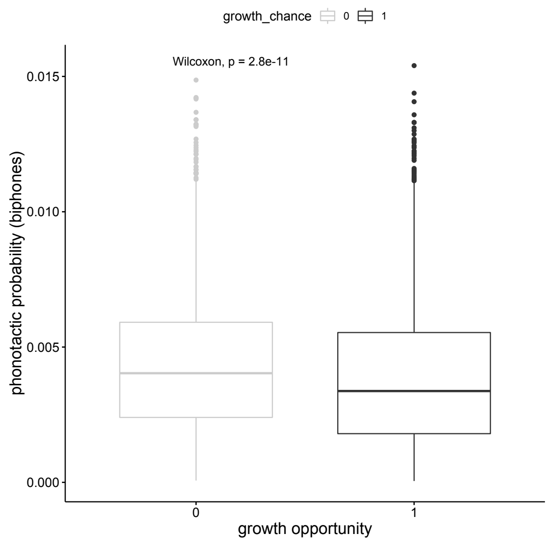


Figure 62: Low phonotactic probability raises the likelihood of a word growing at least one neighbor.

In addition, high lexical frequency raises the probability of acquiring at least one neighbor throughout the learning stages (see Figure 63).

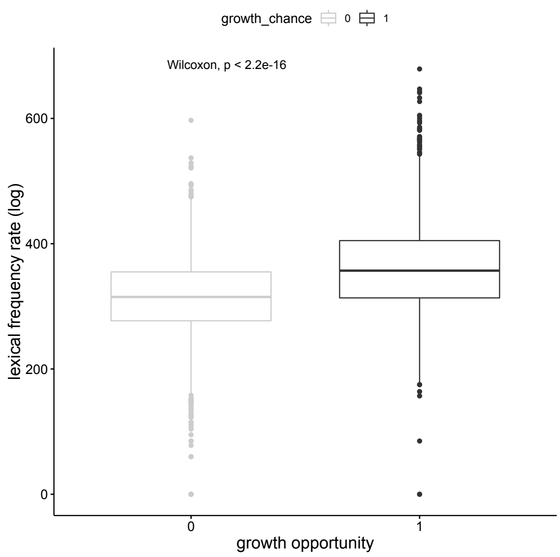


Figure 63: High lexical frequency rate raises the likelihood of a word growing at least one neighbor.

A generalized linear regression model (R function "glm") with the dependent variable "growth opportunity" (0=never grow a neighbor, 1=grow at least one neighbor at some proficiency level) was calculated to determine the influence of phonotactic probability, phonemic length, and lexical frequency rates, using the pseudo model code

Growth opportunity~phonotactic probability+phonemic length+lexical frequency(log), family=binomial

Results show that all three factors play a role in a node's ability to acquire at least one neighbor (see Table 31 for the results).

Table 31: Initial attractiveness is influenced by phonotactic probability, phonemic length, and lexical frequency rate.

	Estimate	Std. Error	z value	p
(Intercept)	0.7	0.23	3.0	
Phonotactic probability	56.25	13.3	4.24	<0.001***
Phonemic length	-0.68	0.02	-31.3	<0.001***
Lexical frequency	0.007	0.0005	12.67	<0.001***

Initial attractiveness can be calculated with a modified preferential attachment equation, following Dorogovtsev, Mendes, and Samukhin (2000b), where the probability of the acquisition of at least one neighboring node is defined as $k+A$, with A being the 'initial attractiveness constant' inferred from the preferential attachment rule incorporating initial attractiveness

$$p(k_i) = (A+k_i)/\sum_j (A+k_j)$$

Thus, it can be calculated that p(0)~A (Barabási, 2016). In the learner networks, A decreases from the beginning to more advanced stages (see Table 32), mirroring the findings displayed in Figure 60.

Table 32: Initial attractiveness constants for each learner network.

	A1	A2	B1	B2	C1
A	0.39	0.21	0.16	0.06	0.05

The A values from Table 32 indicate that initial attractiveness is much higher in early acquired words of the A1 level but decreases substantially over the course of learning history, in accordance with findings from growth development. At the B2 and C1 levels, it becomes highly unlikely for a singleton node to grow its initial and exclusive neighbor. According to Barabási (2016), the presence of initial attractiveness in a network has two consequences: it makes the network more homogenous by decreasing hub sizes, and it influences the degree distribution and deviates it from a power law. Node growth not only takes place in highly connected nodes but also may spread to some non-connected nodes, thereby adding a random component to the attachment probability.

5.3.3. First-mover-advantage

Words that are acquired early during the vocabulary build-up may have an advantage in terms of overall neighbor acquisition, a concept called "first-mover-advantage" (Barabási, 2016). Preferential attachment incapsulates the first-mover-advantage, as only the well-connected nodes gain links from the beginning of the growth process. This leads to a large link advantage of nodes that are initially present in the network, with later arriving nodes lagging behind in obtaining new links. In some networks a certain chronological preference for growth exists where, for instance, new words tend to attach to young nodes, which have only recently been added to the network (also see section 4.3.1.). Alternatively, a tendency to attach to old nodes in a network could indicate a stronger foothold of those older nodes in the network, where new nodes preferentially attach to them. A recent study on phonological neighborhood density (but not from the network viewpoint) showed that in child first language acquisition, early learned words play a more important role for lexical competition than late-acquired words (Karimi & Diaz, 2020). Specifically, it was shown that the age of acquisition affected how strongly or weakly a word was activated in its phonological neighborhood. It is generally assumed that the earlier the acquisition age of a word, the stronger its lexical activation (also see Alario et al., 2004; Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Gilhooly & Watson, 1981; Navarrete, Pastore, Valentini, & Preressotti, 2015, for similar findings). This has implications for word age and its role in lexical networks. The Barabási-Albert network model predicts that older nodes accumulate more growth over time, expressed as $k(t)=t^{1/2}$, with t representing a specific time step (Barabási, 2016).

In order to address the question of the first-mover-advantage in the learner networks, the number of overall neighbors of early-acquired words (A1, A2 levels) was compared to those of late-acquired words (B2, C1). Given the fact that preferential attachment has such a large influence on network growth (see section 5.2.), early acquisition is expected to lead to a larger accumulation of neighbors across the learning time span (up to C1). An overview of node growth classes (see section 5.1.1.) and age of acquisition of words demonstrates a neighbor-growth advantage for early-learned words, while late-learned words grow only slowly. In fact, early-learned words by far exceed all other proficiency levels in terms of neighbor accumulation (mean growth class grouping per proficiency level: A1=2.6, A2=1.3, B1=0.6, B2=0.3, C1=0.1). This indicates a tendency for older nodes that were acquired early in language learning to gain more new neighbors, underscoring the importance of word age for lexical processing (also see Karimi & Diaz, 2020).

5.4. Growth algorithms in communities

Given that communities of different sizes are not homogenous in terms of lexical characteristics (see section 3.4.1. and Siew, 2013), community growth in the learner networks was examined more closely in relation to network growth algorithms. Growth algorithms were calculated per proficiency level in small, medium-sized, and large communities (see Table 33, see also Figure 46 for community growth rates) and results showed a tendency for preferential attachment probability to be higher in mid-sized and larger communities. Node fitness and PAFit were also generally higher in the medium and larger communities, indicating a growth advantage for larger communities. Small communities seem to be awarded fewer opportunities to grow neighbors via the investigated growth algorithms. Here, rich (=large) communities become richer through growth.

Table 33: Growth algorithms in network communities of different sizes.

Level	*Community size*	*Growth rate*		*PA (normalized)*		*Fitness*		*PAFit (normalized)*	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
A1	Small ($<25^{th}$ percentile)	1.6	1.4	0.5	0.2	1.6	1.1	-0.04	0.05
	Medium ($<25^{th}>75^{th}$ percentile)	3.4	2.5	0.6	0.3	1.5	0.9	-0.07	0.09
	Large ($>75^{th}$ percentile)	1.9	1.8	0.6	0.4	1.4	1.3	-0.1	0.2
A2	Small ($<25^{th}$ percentile)	0.3	0.6	0.004	0.009	-0.56	0.5	-0.04	0.03
	Medium ($<25^{th}>75^{th}$ percentile)	2.3	1.9	0.1	0.07	1.4	1.4	-0.22	0.3
	Large ($>75^{th}$ percentile)	2	1.8	0.1	0.06	1.22	1.3	-0.18	0.3
B1	Small ($<25^{th}$ percentile)	0.18	0.4	0.001	0.002	-0.7	0.5	0.006	0.004
	Medium ($<25^{th}>75^{th}$ percentile)	1.15	1.5	0.024	0.02	1.1	1.4	0.028	0.04
	Large ($>75^{th}$ percentile)	1.4	1.4	0.03	0.02	1.3	1.4	0.04	0.06
B2	Small ($<25^{th}$ percentile)	0.07	0.27	0.0006	0.001	-0.72	0.5	1.07e-06	8.29e-06
	Medium ($<25^{th}>75^{th}$ percentile)	0.33	0.67	0.007	0.005	0.75	1.3	-7.65e-05	1.32e-04
	Large ($>75^{th}$ percentile)	0.36	0.6	0.013	0.01	1.29	1.4	-2.06e-04	2.99e-04

C1	Small (<25th percentile)	0.05	0.23	0.0004	0.0007	-0.78	0.5	1.84e-07	1.06e-06
	Medium (<25th>75th percentile)	0.36	0.6	0.007	0.007	1.1	1.6	-2.21e-05	4.1e-05
	Large (>75th percentile)	0.63	0.9	0.009	0.007	1.12	1.4	-2.34e-05	3.6e-05

As can be seen in Figure 64a-c, preferential attachment, fitness, and PAFit differ between differently-sized communities across all learner networks.

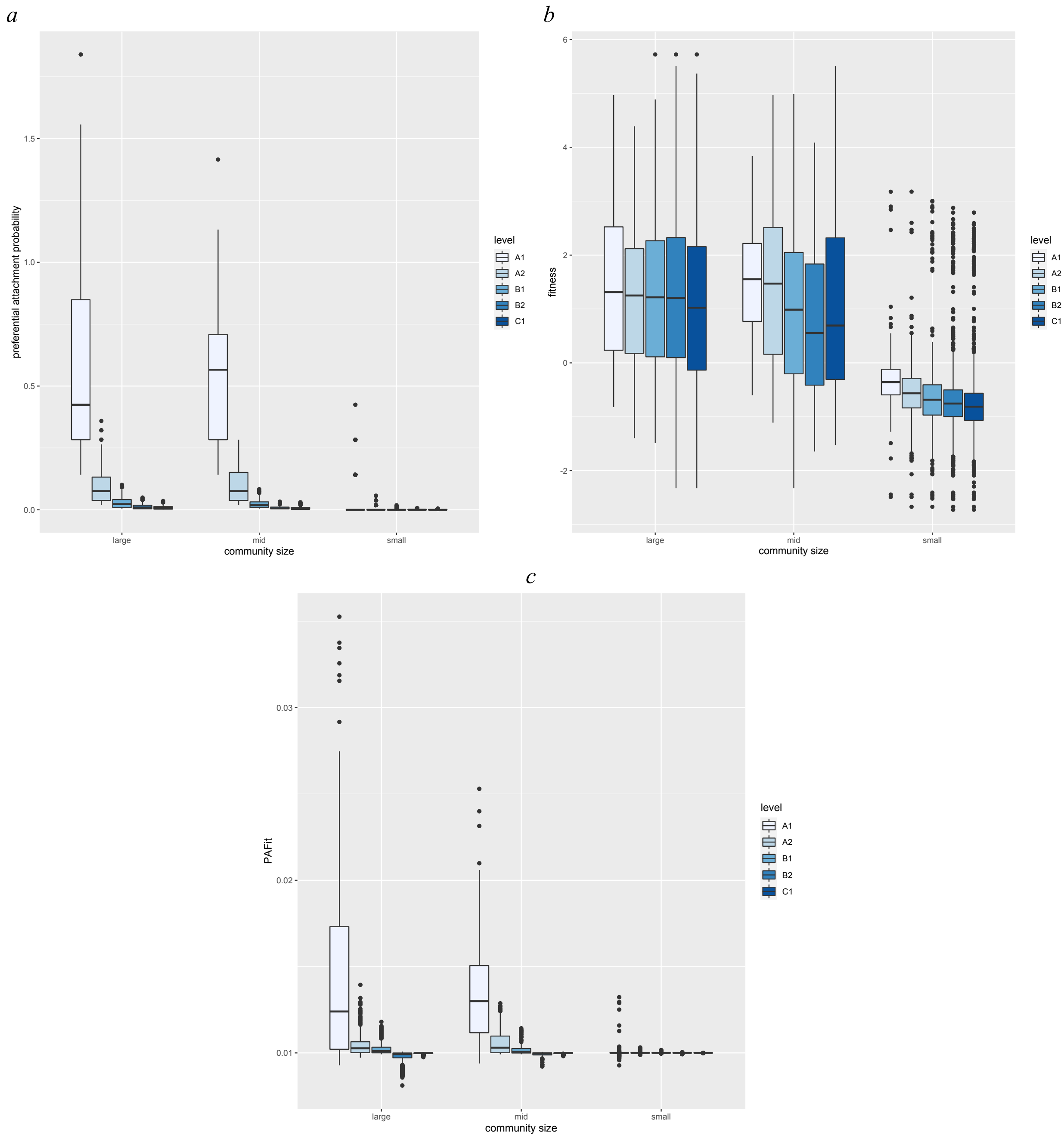


Figure 64a-c: Preferential attachment probability, node fitness, and PAFit differ between proficiency levels and community sizes.

The fact that preferential attachment and PAFit values are the highest in mid-sized and large communities of the A1 network leads to the prediction that these factors can have a more substantial influence on growth in these particular network communities. Node fitness was relatively low in small communities, which shows that any advantage gained from high fitness is primarily bestowed upon words in mid-sized and larger communities.

The results in Figure 64a-c show one clear trend: the patterns of actual community growth rates (see Figure 47) are not identical to the patterns suggested by preferential attachment (64a), fitness (64b), and PAFit (64c). While preferential attachment seems the best match in patterns as judged by visual inspection, it is clear that community growth must be additionally governed by other growth principles. Communities of medium and larger size record the highest growth rates, supporting findings by Siew (2013) who suggested that large communities are built up first in child language acquisition to form the basis for rapid and efficient lexical processing. This trend is particularly relevant for earlier language proficiency stages (A1, A2) where growth rates are the highest of all networks, and here preferential attachment and PAFit might be the primary mechanisms to support the development of robust and larger communities.

5.5. Conclusions

In this chapter, four main strands of growth perspectives were discussed. First, growth rates of individual nodes and their placement in the larger structure of the proficiency network were analyzed to gain insights into which network-mathematical and psycholinguistic factors predispose a word to higher growth. Results demonstrated differences between proficiency levels, lexical characteristics, as well as network parts. Second, the influence of growth algorithms and network centrality measures on growth were tested, with findings indicating that phonological ESL networks can be described as Barabási-Albert growth models, where preferential attachment is the dominant mode. Eigenvector and closeness centralities also affected growth, and the network structure supported by the observed growth patterns of the phonological ESL networks indicate a preference for hub formation, with dense near neighborhoods (one phonological segment distance) but sparser distant neighborhoods (more than one phonological segment distance). Third, extensions of the Barabási-Albert model that include initial attractiveness and aging mechanisms were also demonstrated to influence growth in the ESL networks. Their contributing effect weakens preferential attachment and allows for random attachment, making their presence crucial to explain growth in singleton nodes. Fourth, community growth patterns show unequal growth opportunities for differently-sized communities, with medium-sized and larger communities displaying clear advantages over small communities in terms of growth probabilities resulting from the growth algorithms.

The picture that emerges from the growth analyses of the ESL phonological networks presented here is one of unequal growth distributions: the rich get richer on all levels of the network. On the micro level, dense phonological neighborhoods tend to become denser through growth; on the meso level, mid-sized and larger communities receive the majority of growth opportunities; and on the macro level, giant components tend to grow at the expense of smaller, less connected network parts. As such extreme imbalances are commonly observed in social, technological, and biological networks, it comes as no surprise that "rich words" in the network-theoretical and psycholinguistic sense would gain growth advantages. The findings presented here have direct implications for spoken word recognition and production at different levels of language proficiency. Future research may focus on developmental trajectories concerning the ease and/or difficulty of lexical access of specific words during various stages of vocabulary build-up.

6. Growth and the learning environment

6.1. The ‘yet-to-be-learned’ words

A so-far neglected aspect of lexical networks and word learning is the consideration of the general learning environment in the sense of the ‘yet-to-be-learned’ words at each proficiency level, which formed the basis of previous research on phonological network growth in first-language acquisition. Siew and Vitevitch (2020a) investigated network growth algorithms in their study of phonological networks of L1-learning children and analyzed them from the viewpoint of the not-yet-known words at each acquisition stage. They compiled vocabulary lists of adult L1 English users, subtracted the words known by children at each analyzed acquisition stage, and computed the growth algorithms per unknown word in relation to the known words of the learners. Essentially, network growth algorithms were calculated for each yet-to-be-learned word, and it was expected that high growth algorithm scores would predict the likelihood that a word would be learned at a given acquisition stage. This methodology is fundamentally different from the current approach to network growth in the ESL learner networks and does not enable direct comparison of the results. Therefore, the data of the present study was re-calculated to replicate the Siew and Vitevitch study for the ESL learner networks to gain insights into possible parallels of second-language and first-language vocabulary growth from the network perspective.

Consistent with Siew and Vitevitch (2020a), vocabulary lists of yet-to-be-learned words at each proficiency level were compiled, which only included the new words that appear at a specific proficiency level but are not present in the previous level (see Table 34). The ESL learning environment was defined as the words known at the C2 proficiency level: 6714 words (see section 3.1.1.).[4]

Table 34: Newly learned words that appear in the lexicon of learners at each proficiency level. References to proficiency levels indicate the end point of growth, meaning the higher proficiency level achieved through growth.

A2	B1	B2	C1	C2
877	1426	1727	1018	1084

Learners at the A1 proficiency level have 6132 English words that they have yet to learn (A2 to C2 growth), while at the C1 level learners have 1084 new words to learn to reach the C2 level. Based on these word growth lists, growth algorithms were calculated in accordance with Siew and Vitevitch, who adapted theirs from the semantic network study by Hills and colleagues (Hills et al., 2009b). For each newly learned word (per proficiency level) the following three growth variables were calculated: preferential acquisition growth value (PACQ), preferential attachment growth value (PATT), and the "lure of the associates" growth value (LA). These values were not stable for a given word but changed with each proficiency level, as the learners' vocabulary continued to grow. An explanation of each algorithm with examples is presented below.

6.1.1. Preferential acquisition (PACQ)

PACQ takes into account the set of words that will eventually be known by a learner if they achieve L1 (native) proficiency. In the context of the present study this is first-language English, and the adult vocabulary of L1 English users thus serves as a reference point for vocabulary learning. Preferential acquisition assumes that new words that are phonological

[4] This definition excludes an analysis of the C2 levels, as there is no defined learning environment for that level. It was opted against using L1 English as a measure of the learning environment for C2 learners, since L1 and L2 word learning are conceptually different processes from the network perspective (see section 3.1.) and a conflation of the two would have rendered the C2 analysis out of line with the other proficiency level analyses.

neighbors of a high number of words in L1 English are learned earlier than those that have few neighbors in L1 English. Thus, the PACQ value of a yet-to-be-learned word is the number of phonological neighbors of that word in the adult English L1 lexicon (see Figure 65 for a schematization). Neighborhood densities of L1 English words were calculated with the Clearpond Database for English (Marian, Bartolotti, Chabal, & Shook, 2012), and all neighbors of each of the yet-to-be-learned words were determined. For instance, at the B2 level, the yet-to-be-learned words included *aboard*, which has ten phonological neighbors in the L1 English lexicon (*afford, award, abort, abide, adored, abroad, accord, abode, board, bored*), with the PACQ value equaling "10" in this case.

6.1.2. Preferential attachment (PATT)

This preferential attachment/PATT growth value is not computed following the Albert-Barabási equation (see Barabási, 2016, and chapter 3 of this manuscript) but the phonological neighborhood density of anchor words in the known lexicon, i.e., node degree, was used to represent PATT[5]. For each word in the yet-to-be-learned vocabulary list (per proficiency level), the phonological neighbors of those words to which a yet-to-be-learned word can potentially become a phonological neighbor were determined. For instance, the yet-to-be-learned word *aboard* at the B2 level has five phonological neighbors in the C1 lexicon (*board, bored, abroad, afford, award*). The mean number of phonological neighbors of those five words – and thus the PATT growth value for "aboard" at B2 – is 7.4. PATT growth values for a given word are different at each proficiency level, and *aboard* has a growth value of 6 and 5 at the B1 and A2 levels. As the vocabulary of a learner grows, PATT values change. Figure 65 graphically shows the PATT calculations.

The inverse variant of preferential attachment (iPATT) stipulates that nodes with few links (=low-degree nodes) receive the largest share of new nodes in more developed phonological networks (Siew & Vitevitch, 2020a, 2020b). In order to determine the influence of iPATT on the ESL learner networks, the measurements obtained for PATT (see above) were co-opted for the reverse hypothesis that low PATT values would be better predictors of neighbor growth.

[5] Values calculated with the original preferential attachment equation and those calculated with the proxy variable PATT as used by Siew and Vitevitch were correlated at r=0.65 (Pearson's r).

6.1.3. Lure-of-the-associates (LA)

LA growth values represent the number of phonological neighbors in a learner's lexicon to which the yet-to-be-learned words can connect to. The word *aboard* has 5 phonological neighbors at the B2 level (*board, bored, abroad, afford, award*) and thus its LA growth value equals 5. The difference to PATT is that only the phonological neighbors are counted, whereas PATT counts the phonological neighbors of the phonological neighbors (=degree of a node) at a given proficiency level. Words with high LA growth values are expected to be learned preferentially as opposed to words with few phonological neighbors (=low LA growth values) in a learner's existing vocabulary. See Figure 65 for a schematization of the different growth algorithms.

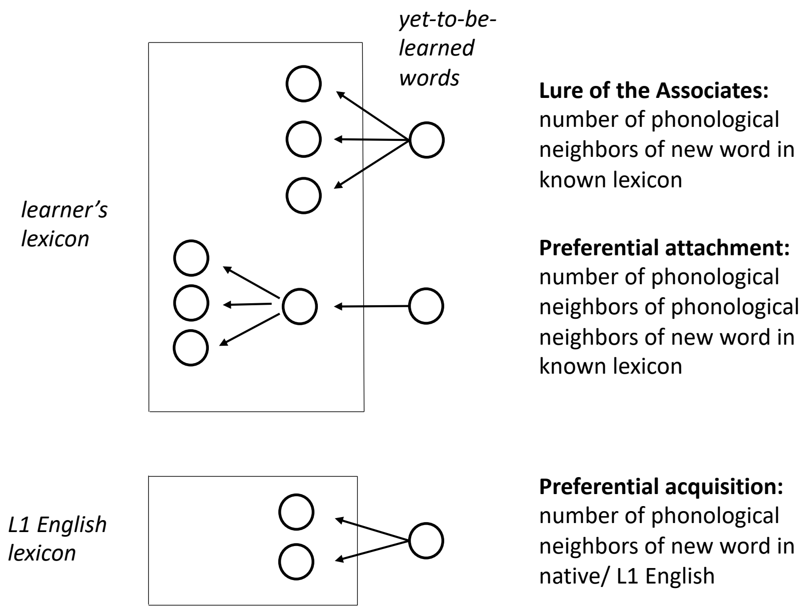


Figure 65: Schematization of three network growth algorithms calculated by Hills et al. (2009) and Siew and Vitevitch (2020b) as adapted for the present study.

The distributions of the growth algorithms LA, PACQ, and PATT are displayed in Figures 66, 67, and 68.

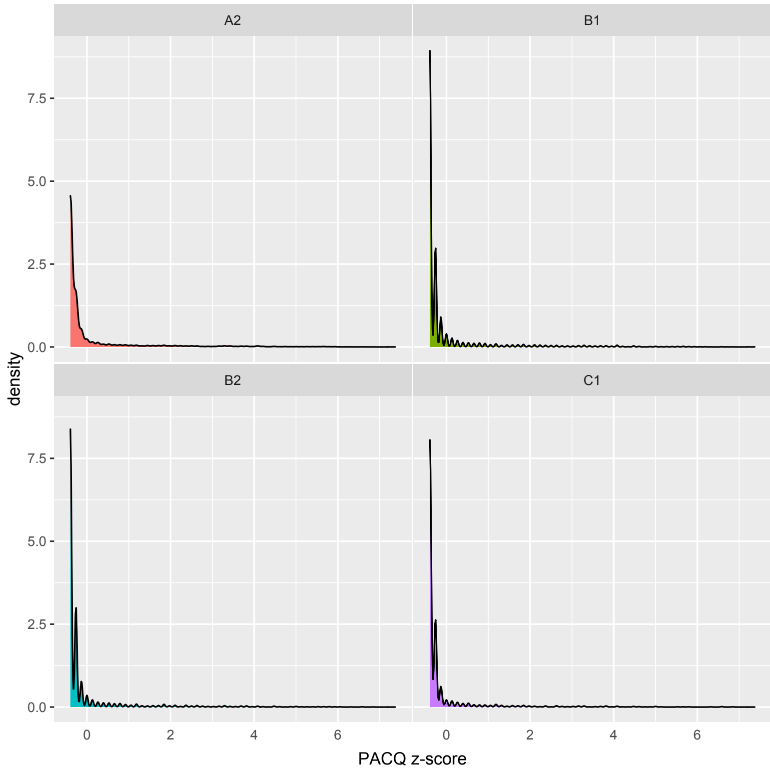


Figure 66: Distributions of PACQ growth algorithm across the growth levels.

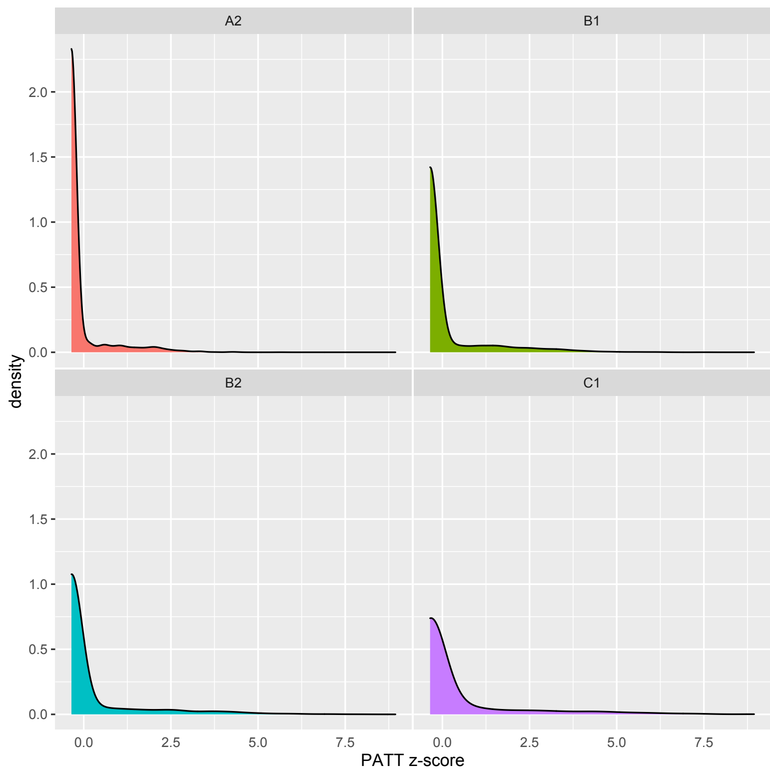


Figure 67: Distributions of PATT growth algorithm across the growth levels.

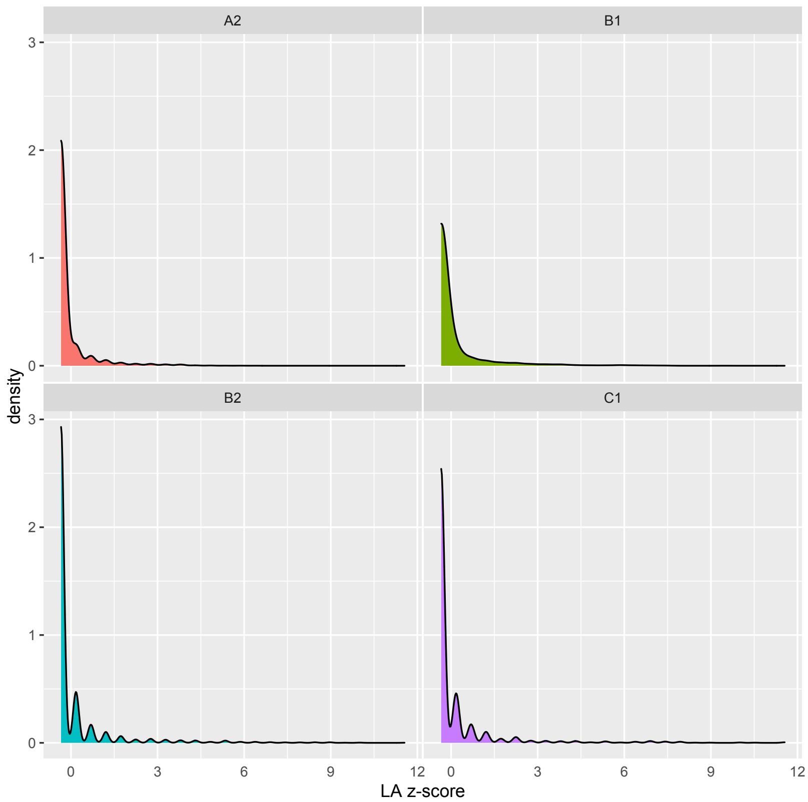


Figure 68: Distributions of LA growth algorithm across the growth levels.

The distributions of the three algorithms of the yet-to-be-learned words in the learner networks are presented in Table 35.

Table 35: Distribution types of network growth in yet-to-be-learned words at each proficiency level.

Best-fitting distribution	*A2*	*B1*	*B2*	*C1*
PACQ	Log-normal	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)
PATT	Power law	Power law	Power law	Power law
LA	Log-normal	Log-logistic	Multimodal (dip test p: <0.001)	Multimodal (dip test p: <0.001)

The PACQ score is related to the phonological neighbors of yet-unknown words in the English L1 lexicon, which is much larger than the learner lexica, thus leading to the PACQ score to be similar across most ESL proficiency levels. The fact that PACQ is evidently smaller in the A2 network indicates that the large number of yet-to-be-learned words of ESL beginners find fewer phonological neighbors (or anchor words) in the small A2 vocabulary. As lexical knowledge and networks grow, the list of yet-to-be-learned words becomes shorter but the anchor points in the larger lexica become more numerous. The PATT score develops from a significant skew to a much less-skewed distribution as proficiency increases but overall follows a power law distribution (see Table 35). Only a small sub-group of the yet-to-be-learned words of A2 learners show high PATT values, with the majority displaying low values. As the learner networks grow, the set of unknown words at the advanced stages (C1) become characterized by less skew, which could be a reflection of the fact that the yet-to-be-learned words are relatively few at this point. The LA score is generally distributed rather unevenly, with a small number of nodes showing a high score and a large majority of nodes showing a small score across all proficiency levels. What these distribution-related findings clearly demonstrate is that characteristics of the yet-unknown vocabulary of learners change as learning progresses. The shrinking learning potentials leave the yet-to-be-learned vocabulary exposed to different dynamics regarding their links to L1 English words (PACQ) and the constantly expanding vocabulary of a learner (PATT, LA).

6.2. Assessing the statistical effects in the learner networks

In line with Siew and Vitevitch (2020a), the network growth algorithms were entered into a spreadsheet, together with all yet-to-be-learned words and level-of-acquisition information for each. One column indicated whether a word was learned at a specific proficiency level (yes/no). The growth levels were defined as ranging from proficiency level A2 to C1 (e.g., "A2" represents the growth from A1 to A2). Regression modelling was used to test the effects of the algorithms on network growth.

First, the contribution of each growth algorithm to the overall probability of word learning was assessed in order to ensure they were actual predictors of learning probability (also see Siew & Vitevitch, 2020a). Identical to Siew and Vitevitch, the variables phonemic length and lexical frequency rate were included in the models as they have a well-known effect on word learning probability. Five generalized linear mixed-effects models were constructed and a series of likelihood ratio tests were calculated to compare the model fits to the data. Model 1 included the following interactions (indicated by *) and their main effects: phonemic length *proficiency level, and lexical frequency rate* proficiency level. Model 2 included all predictors from model 1 and the PACQ*proficiency level interaction; model 3 included the predictors from model 1 and the LA*proficiency level interaction; model 4 included all predictors from model 1 plus the interaction PATT*proficiency level. Model 5 represented the full model, including interaction terms between proficiency level and all predictors (phonemic length, lexical frequency rate, PACQ, LA, PATT growth values). The function "lrtest" of the R package "lmtest" was used for likelihood ratio tests for nested models. Results of the tests are summarized in Table 36 below. All three network growth algorithms are significant predictors of word learning beyond the influence of the variables phonemic length and lexical frequency rate.

Table 36: Results of likelihood ratio tests for the models.

	Model 2: PACQ	Model 3: LA	Model 4: PATT
Model 1: Baseline	$\chi2=6.59$ df=2 p=0.037	$\chi2=35.38$ df=2 p<0.001	$\chi2=22.23$ df=2 p<0.001
Model 5: Full model	$\chi2=32.96$ df=4 p<0.001	$\chi2=4.17$ df=4 p=0.38	$\chi2=17.32$ df=4 p=0.002

The full model was then submitted to a generalized linear mixed-effects model analysis with binomial error structure, using the function "glmer" of the R package "lme4" (Bates et al., 2014), in order to determine which of the three growth algorithms can best account for word learning. The dependent variable was binary for "learned" (=1) or "not learned" (=0) for each word, describing whether or not they appeared in the vocabulary of the next higher proficiency level. As fixed effects, the three growth mechanisms were added, in addition to the control variables phonemic length and lexical frequency rate (calculated with the BNC, see section 3.1.1.). All variables were first z-scored before being entered into the model. In the full model, interaction terms were constructed between proficiency level and all other predictors. Inclusion of the interaction terms allows for the accounting of the magnitude of the effect of the growth algorithms varying over time/proficiency. As a random effect, "word" was specified in order to account for possible individual variation associated with particular words. Mixed effects analysis was performed to model the relationship of the fixed effects and the probability of word learning (yes/no). P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question (Dobson, 2002; Forstmeier & Schielzeth, 2011). Residuals were checked to ensure homoscedasticity and normality of the distribution. Cases of multicollinearity were searched by computing the variance inflation factors of the model (R package "car", function "vif"), which were all below 3.5. Even though the LA and PATT scores are correlated at r=0.71 (Pearson's r), their variance inflation factors were below 3.5 in the model. The correlation of >7 can hardly be ignored, however, even if variance inflation factors were not affected in a way that made the model untenable.[6] Therefore, two models were computed: (1) model 1: an identical model to Siew and Vitevitch (2020b), with each of the three growth algorithms, and (2) model 2: an alternative model accounting for the correlation between LA and PATT by creating an interaction term for the two. The pseudo-code for the full model 1 was

```
learned~LA*level+PACQ*level+PATT*level+phonemic length*level+lexical frequency*level+(1|word),
family=binomial
```

Table 37 summarizes the descriptive statistics for lexical characteristics and network growth values for yet-to-be-learned words at a given proficiency level.

Table 37: Descriptive statistics of predictors for unknown words at the different proficiency levels.

		A1	A2	B1	B2
	N (yet-to-be-learned)	*6132*	*5246*	*3821*	*2097*
Phonemic length	M	6.42	6.64	6.93	7.2
	SD	2.44	2.44	2.47	2.49
	Min	1	1	1	1
	Max	18	18	18	18
Lexical frequency (log)	M	3.13	3.06	2.97	2.8
	SD	0.63	0.59	0.54	0.49
	Min	0	0	0	0.48
	Max	5.71	5.5	5.32	4.95
PACQ	M	3.55	3.05	2.53	2.15
	SD	8.32	7.65	6.86	6.28
	Min	0	0	0	0
	Max	59	59	55	55
LA	M	0.43	0.68	0.85	0.93
	SD	1.29	1.9	2.33	2.59
	Min	0	0	0	0
	Max	13	19	23	23
PATT	M	0.45	0.77	0.97	1.3
	SD	1.36	2.1	2.55	3.16
	Min	0	0	0	0
	Max	13	17	18	20.1

Results of the mixed-effects logistic regression show that numerous main effects and interactions terms were significant (see Table 38). As expected, short phonemic length and high lexical frequency rates (and their interaction terms with proficiency level) were significant predictors of higher likelihood of word learning, providing further support for a large body of research on the influence of the two variables on word learning. The only network growth algorithm that had a measurable effect on probability of word learning was the lure of the

[6] Siew and Vitevitch (2020a) do not provide information on correlations between any of the growth algorithms in their study.

associates (LA) score (in isolation and in interaction with proficiency level). Here, a higher LA score raised probability of learning in the yet-to-be-learned words (see Figure 69).

Table 38: Final mixed-effects logistic regression with lexical and network growth values to predict learning (model 1).

Predictors	Estimate	SE	*z*	*p*
(Intercept)	-4.055	0.588	-6.89	< 0.001 ***
Phonemic length*proficiency	0.082	0.009	8.53	< 0.001 ***
Lexical frequency* proficiency	0.079	0.034	2.29	0.022 *
PACQ* proficiency	0.014	0.023	0.59	0.48
LA* proficiency	-0.096	0.031	-3.09	0.008 **
PATT* proficiency	0.016	0.029	0.53	0.42
Phonemic length	-0.48	0.044	-10.74	< 0.001 ***
Lexical frequency	0.81	0.15	5.38	< 0.001 ***
PACQ	-0.078	0.09	-0.82	0.41
LA	0.45	0.15	3.05	0.0023 **
PATT	-0.039	0.14	-0.27	0.79

Figure 69 shows the general trend of the lower proficiency levels (A2, B1, B2) to have higher values of LA, PATT, and PACQ in the learned words, while this trend was reversed in the higher proficiency level (C1) where the three growth algorithms showed higher values in non-learned words. This clearly represents a similar developmental dynamic as outlined by Siew and Vitevitch of the more advanced lexicon relying on lower rather than higher PATT values to raise probability of word learning. The regression analysis yielded only the LA score as significant, so PACQ and PATT scores in Figure 69 should be interpreted with caution.

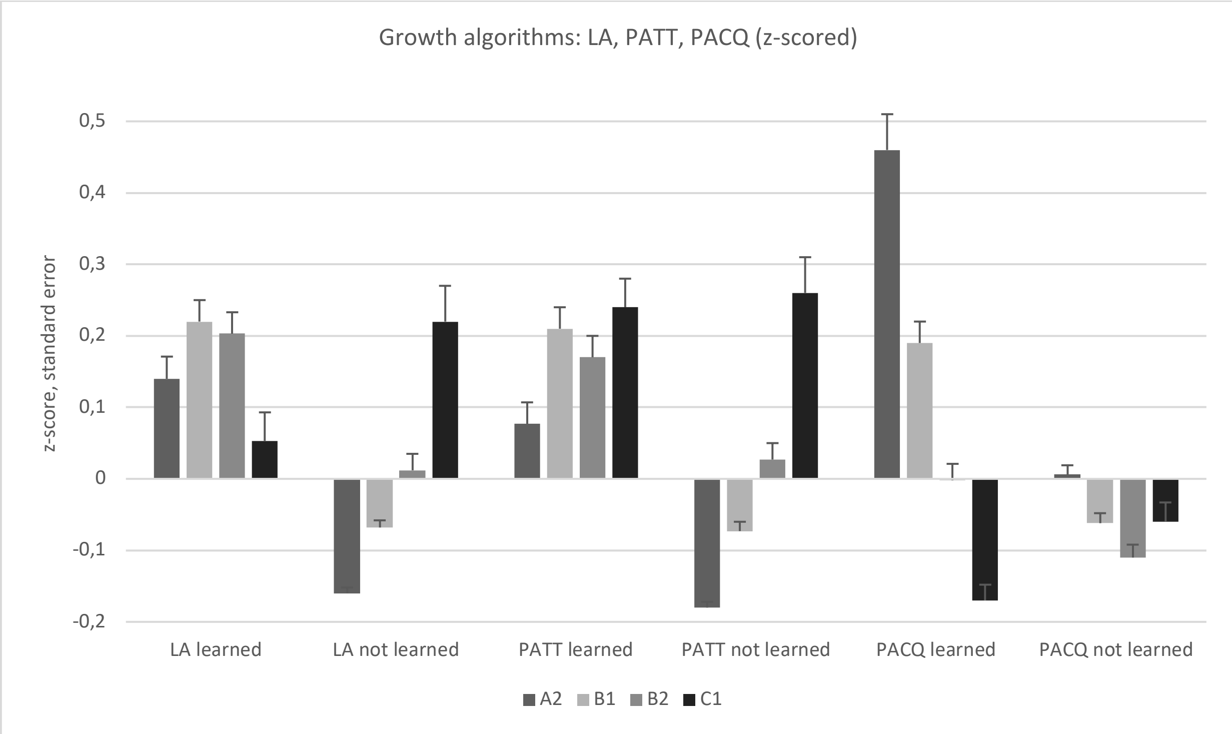


Figure 69: Overview of network growth algorithms and their contributions to word learning probability. Note the opposite trend in learned/non-learned words at the C1 level.

Due to the correlation between LA and PATT (r=0.71), an alternative model was computed where LA and PATT were centered and combined as an interaction variable. The pseudo-code of the full model 2 was

```
learned~PACQ*level+LA*PATT*level+phonemic length*level+lexical frequency*level+(1|word),
family=binomial
```

Here, the LA*PATT interaction variable was shown to have the same effect on learning probability that was described by Siew and Vitevitch (2020b) for PATT, with the smaller vocabularies at the beginning stages of learning showing higher scores in learned words (see Figure 70), which becomes reversed at the C1 level, reflecting what Siew and Vitevitch termed the iPATT effect (Siew & Vitevitch, 2020a, 2020b). PACQ did not have an effect on learning probability in this new model (see Table 39 for results).

Table 39: Final mixed-effects logistic regression with lexical and network growth values to predict learning, including the LA*PATT interaction term (model 2).

Predictors	Estimate	SE	*z*	*p*
(Intercept)	-2.04	0.04	-56.19	< 0.001
Phonemic length*proficiency	0.49	0.06	0.28	< 0.001

Lexical frequency*proficiency	0.79	0.13	8.43	<0.001
PACQ*proficiency	-0.016	0.06	-0.13	0.47
LA*PATT*proficiency	0.14	0.036	3.86	< 0.001***
Phonemic length	-0.64	0.04	-16.1	< 0.001***
Lexical frequency	0.61	0.04	15.27	< 0.001***
PACQ	-0.04	0.03	-1.23	0.22
LA*PATT	-0.14	0.033	-4.34	<0.001***

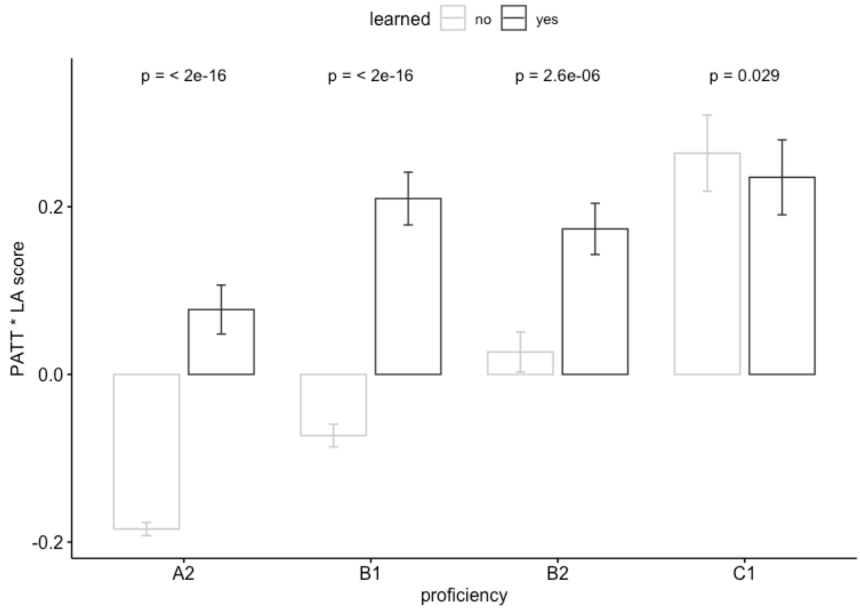


*Figure 70: LA*PATT scores are higher in learned words in the lower proficiency levels; in the highest proficiency level, the dynamic is inverse and LA*PATT scores are lower in learned words.*

The iPATT mechanism has been suggested to avoid clustering of too many similar word forms in the advanced lexicon (Siew & Vitevitch, 2020b). This has benefits for lexical processing by minimizing competitor words and enhances network robustness (see Zhao & Xu, 2009). Human linguistic cognition must strike a balance between the efficiency of similarity-based learning on the one hand, and maximal discriminability on the other hand. According to the iPATT hypothesis, preferential attachment functions as the initial mechanism by which new words are added to the lexicon. At this stage, the lexicon is small and phonological similarity clusters provide learning advantages, without impeding lexical processing. Once the lexicon has grown, the PATT mechanism is reversed to the iPATT mechanism, and new neighbors are added to low-degree network words. Through this process, unequal distribution of growth is restricted and growth can supplement low-degree and singleton nodes to build a more robust network. It remains to be investigated when and how the switch between the two mechanisms PATT and iPATT takes place and what factors contribute to its cause in first and second languages.

The present study suggests that phonological similarity-guided learning is most pronounced in beginners, when there are actually only few words that new neighbors can attach to (see Figure 71). In larger lexica, new words can find a multitude of cognitive anchors to attach to phonologically. At this stage, network growth becomes more diversified through iPATT. The learning environment of yet-to-be-learned words, on the other hand, shrinks as proficiency grows. At the A2 stage, more than 6000 words are "available" for learning but they only find a small number of possible anchor words in a learner's dictionary. At the C1 level, word learning potential has shrunk to approximately 1000 words, but anchor points in the lexicon have grown by a multitude.

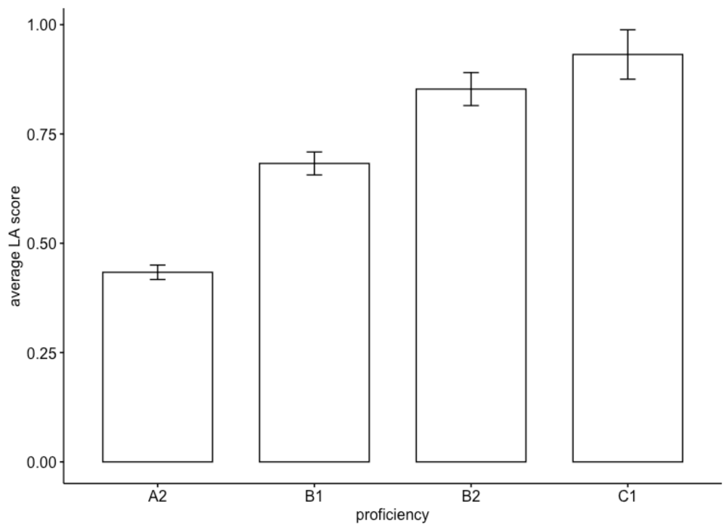


Figure 71: Potential anchor words become more frequent as the lexicon grows.

The findings of the present study align with the hypothesis advanced by Siew and Vitevitch (2020a, 2020b) that the current vocabulary retained by a learner plays a central role in growth processes (as opposed to what was found by Fourtassi et al., 2020). Both LA and PATT were shown to be influential for word learning, with PACQ not having an effect on the results. The interaction between LA and PATT is initially higher in learned words but this dynamic is turned on its head in the advanced proficiency level C1, where LA*PATT is lower in learned words. This parallel pattern of growth dynamic documented for first and second language learners of

English is certainly noteworthy, as it may represent a basic and common principle of language learning. Language proficiency changes many of the parameters of the learning process, and there is an undeniable role of the existing knowledge for the acquisition of new knowledge. What the results presented in this chapter suggest is that the lexical knowledge that a learner possesses at a given time is indicative of what they will learn in the future. The adult L1 lexicon of a particular language is the target for all learners, be it children or second language learning adults, however, lexical acquisition – like all learning – is a process where existing knowledge has to be interwoven with new knowledge in a systematic way in order to create a harmonious web of knowledge in the learners' minds. Anchoring new words to known words can represent an efficient method for maximizing similarity-guided learning benefits in language learners' mental lexica, all the while reducing cognitive effort committed to linguistic processing and storage.

7. Conclusion

7.1. Utility of network science for lexical learning

The current book has highlighted the advantages of network science for the study of the mental lexicon of (second) language learners, and it is easy to see what network-theoretical concepts can contribute to our understanding of word learning and lexical access. Underlying the methodology of network theory are the assumptions that a system is governed by local as well as global rules, which work in tandem to influence the behaviors of the individual constituents. By contrast, a reductionist view utilizes only a fraction of the available information within a (lexical) system and tends to underestimate its complexity by failing to account for global properties that may impose unexpected constraints. Until recently, the mental lexicon was exclusively studied from the “bottom-up” perspective of local rules governing small groups of interrelated words, and these findings greatly furthered our understanding of the dynamical underpinnings of lexical learning and processing. The introduction of network science has now offered a tool to take a “top-down” approach to study the mental lexicon, following the rationale of complex systems. Rather than conceiving of the mental lexicon as an accumulation of independent, co-existing local word neighborhoods, lexical networks posit a complex lexical storage system characterized by far-reaching interconnections that span a large number of words. In this view, lexical learning is seen as a process of network growth, emphasizing the fact that network changes are dependent on network structure, encapsulating Steven Strogatz’ rule that “structure always affects function”. This interplay between the grander layout of the lexical network and the possibilities it offers for investigations of word learning are significant to acknowledge, especially in light of long-known evidence that not all words are equipotential in their learning probability by (second) language learners. Network science now opens the view on how individual words can embed themselves in the pre-existing lexical knowledge of learners and can, thus, derive new predictions about learning probabilities.

The findings presented in this book can contribute to theories on second language word memory and learning. The structural organization of word knowledge at different proficiency levels can help inform predictions on activation spreading by making visible and analyzable the larger neighborhoods in which words are embedded and in which activation can spread, following the activation diffusion hypothesis put forth by Michael Vitevitch and colleagues for first language processing. The lexical dynamic of co-activation leaking out of the immediate neighborhood, essentially thinning out activation and thereby delaying lexical access of a target word when there is a denser larger neighborhood, may well be identical in second language learning, and the present findings can form the methodological scaffold upon which future studies of L2 activation spreading can build.

The results of this book can also make contributions to scale-free network growth theories by detailing growth algorithms in evolving phonological networks. Preferential attachment was shown to have an overwhelming influence on growth at all stages of network development, supporting a dominant theory of evolving networks. The increasingly popular fitness growth models, on the other hand, are less applicable to phonological networks, even though they take into account lexical characteristics indicative of better word learning and, thus, would seem a better theoretical fit. Seemingly, cumulative advantages in phonological neighbors excel over quality of those neighbors, with ensuring implications for the growing architecture of the investigated network over time. Principles of rich-get-richer, initial attractiveness, aging effects, as well as the first-mover-advantage in phonological networks, place them well within the mathematical norms of scale-free networks, further underscoring the utility of network-mathematical growth algorithms for phono-lexical data.

Lastly, the data presented in this book can help answer questions as to what role the external and internal language environments play in word learning. Growth of the ESL learner networks suggests that the current lexical knowledge of learners at a given time is more influential, as opposed to the L1 environment of the target language. Because literature on this issue is scarce and conflicting, with only two network-theoretical studies with contrary findings being available, any new study contributing evidence to one or the other hypothesis can add valuable information to this ongoing debate in second language learning and psycholinguistics.

An online appendix to this book makes the neighborhood data as calculated from the EVP website available. All one-phoneme neighbors (Levenshtein distance) of all words in the learner networks are presented per proficiency level and in a format that can easily be navigated by interested researchers or used to replicate the present study. Furthermore, this resource can be used for neighborhood density estimations of ESL/ L2 English. As opposed to relying on neighborhood density data from L1 English, future studies involving English-as-a-second-language learning are given the option to calculate more L2-specific and potentially more applicable neighborhood density data that can better account for learners' actual lexical knowledge at a given proficiency level.

7.1.1. Future directions

A number of relevant issues have arisen from the findings presented here, which were beyond the scope of this book but may be explored by future studies incorporating the network-theoretical approach to vocabulary storage and learning in second languages. A few will be mentioned here.

First, the developmental trend of hub formation during network growth carries potential consequences for the learning of phonological detail. At initial learning stages, not all phonological contrasts are successfully implemented in a learner's language, as for instance, dental fricatives are difficult to acquire for many learners whose first language does not incorporate such consonants, and they may be substituted by sibilants or labio-dental fricatives, resulting in words such as "sink" instead of "think" in a German learner's ESL vocabulary. As learning progresses, phonological (and phonetic) learning may improve, leading to the acquisition of the correct dental fricative [θ]. An interesting question concerns the role of hubs in this phonological learning/phonological changes in the network: do hubs suppress the introduction of a new phonological feature, due to the many members exerting pressure to maintain a conservative phonological standard in the neighborhood? Or are hubs the main facilitators of phonological changes, disseminating them most efficiently in the network? A similar question could be posed in relation to sound changes in first languages in general. The role of hubs in phonological or phonetic changes with potentially far-reaching consequences for the whole network remains to be described.

Another intriguing research question relates to language loss or network decline. Analyzing which network parts start to disintegrate first and which parts are able to maintain their structure the longest, whether build-up and decline are correlated, and what constitutes the most robust core of a network can yield valuable information for strategies on how to fight language loss, be it temporary or pathological. Language educators as well as clinical linguistics studying language impairments stand to benefit from such an analysis.

The network-theoretical approach to the mental lexicon has yet to be broadly evaluated by psycholinguistics and tested under different experimental conditions, but there can be no doubt that it has potential to complement current theoretical frameworks of word memory and learning. The most promising possibility is that network science will be best understood as a natural extension of the neighborhood analysis methodology to study the inner workings and developmental functions of the mental lexicon.

8. References

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. London: John Wiley & Sons.

Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, and Computers, 36*(1), 140-155.

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

Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature, 406*, 378-382.

Alexy, O., George, G., & Salter, A. J. (2013). Cui bono? The selective revealing of knowledge and its implications for innovative activity. *Academy of Management Review, 38*(2). doi:https://doi.org/10.5465/amr.2011.0193

Alvarez-Ponce, D., Feyertag, F., & Chakraborty, S. (2017). Position matters: Network centrality considerably impacts rates of protein evolution in the human-protein interaction networks. *Genome Biology and Evolution, 9*, 1742-1756.

Amaral, L. A. N., Scala, A., Barthélémy, 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.

Anwar, R., Yousuf, M. I., & Abid, M. (2021). Uniform preferential selection model for generating scale-free networks. *MEthodology and Computing in Applied Probability*. doi:https://doi.org/10.1007/s11009-021-09854-w

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos, 20*(3), 679-685.

Arutiunian, V., & Lopukhina, A. (2020). The effects of phonological neighborhood density in childhood word production and recognition in Russian are opposite to English. *Journal of Child Language, 47*(6), 1244-1262.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics of lexical neighborhoods? *Journal of Memory & Language, 44*, 568-591.

Banerjee, S., & Bhamidi, S. (2021). Persistence of hubs in growing random networks. *Probability Theory and Related Fields*. doi:https://doi.org/10.1007/s00440-021-01066-0

Barabási, A. L. (2009). Scale-free networks: A decade and beyond. *Science, 325*, 412-413.

Barabási, A. L. (2016). *Network science*. Cambridge, UK: Cambridge University Press.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512.
Bard, E. G., & Shillcock, R. C. (1993). Competitor effects during lexical access: Chasing Zipf's tail. In G. E. Altmann & R. C. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 235-275). Hillsdale, NJ: Lawrence Erlbaum Associates.
Bard, E. G., Shillcock, R. C., & Altmann, G. E. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Evidence of subsequent context. *Perception and Psychophysics, 44*, 395-408.
Baronchelli, R., iCancho, R. F., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Science, 17*, 348-360.
Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks.* Paper presented at the International AAAI Conference on Weblogs and Social Media.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). {lme4}: Linear mixed-effects models using Eigen and S4. *R Package version 1.1-7*.
Bauer, R., & Kaiser, M. (2017). Nonlinear growth: An origin of hub organization in complex networks. *Royal Society Open Science, 4*(3). doi:https://doi.org/10.1098/rsos.160691
Beckage, N. M., Aguilar, A., & Colunga, E. (2015). *Modeling lexical acquisition through networks*. Paper presented at the Proceedings of the 37th Annual Conference of the Cognitive Science Society, Austin, TX.
Beckage, N. M., & Colunga, E. (2016). Language networks as models of cognition: Understanding cognition through language. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks: Understanding complex systems* (pp. 2-20). Berlin, Heidelberg: Springer.
Beckage, N. M., & Colunga, E. (2019). Network growth modeling to capture individual lexical learning. *Complexity, 2019*, 7690869.
Beckage, N. M., Smith, L., & Hills, T. T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE, 6*(5), e19348.
Bedogne', C., & Rodgers, G. (2006). Complex growing networks with intrinsic vertex fitness. *Physical Review E, 74*(4), 046115.
Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition, 96*(2), B45-54.
Bell, M., Perera, S., Piraveenan, M., Bliemer, M., Latty, T., & Reid, C. (2017). Network growth models: A behavioural basis for attachment proportional to fitness. *Scientific Reports, 7*, 42431.
Benham, S., Goffman, L., & Schweickert, R. (2018). An application of network science to phonological sequence learning in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 61*(9), 2275-2291.
Berg, T., & Schade, U. (1992). The role of inhibition in a spreading-activation model of language production. I. The psycholinguistic perspective. *Journal of Psycholinguistic Research, 21*, 405-434.
Bialystok, E. (2010). Bilingualism. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 559-572.
Bianconi, G., & Barabási, A. L. (2001). Bose-Einstein condensation in complex networks. *Physical Review Letters, 86*(24), 5632-5635.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 10008*(6). doi:https://doi.org/10.1088/1742-5468/2008/10/P10008
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physical Reports, 424*, 175-308.
Boccara, N. (2010). *Modeling complex systems*. Berlin et al. : Springer.
Bock, K., & Levelt, W. J. M. (1994). Languge production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). Sand Diego, CA: Academic Press.
Bonacich, P. (1972). Technique for analyzing overlapping memberships. *Sociological Methodology, 4*, 176-185.
Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks, 29*(4), 555-564.
Bonacich, P., & Lloyd, P. (2015). Eigenvector centrality and structural zeroes and ones: When is a neighbor not a neighbor? *Social Networks, 43*, 86-90.
Bose, S. (1924). Planck's law and the light quantum hypothesis. *Zeitschrift für Physik, 26*, 178.
Broersma, M., & Cutler, A. (2008). Phanton word activation in L2. *System: An International Journal of Educational Technology and Applied Linguistics, 36*, 22-34. .
Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications, 10*, 1017.
Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychologial Review, 114*, 539-576.
Bryla, B. (2015). *Oracle Database 12c handbook: Manage a scalable, secure Oracle enterprise database environment*. New York et al.: McGraw Hill.
Brysbaert, M., Mandera, P., & Keuleers, E. (2017). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45-50.
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.
Buchwald, A. (2011). Neighborhood effects. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology: Volume IV Phonological Interfaces* (pp. 2070-2088). Chichester, UK: Wiley-Blackwell.
Caldarelli, A., Capocci, A., De Los Rios, P., & Munoz, M. A. (2002). Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters, 89*, 258702.
Callaway, D. S., Hopcroft, J. E., Kleinberg, J. M., Newman, M. E. J., & Strogatz, S. H. (2001). Are randomly grown graphs really random? *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 64*(4), 041902.
Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implication for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1430-1450.
Carlson, M. T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory & Language, 75*, 159-180.
Carrasco-Ortiz, H., Midgley, K. J., & Frenck-Mestre, C. (2012). Are phonological representations in bilinguals language specific? An ERP study on interlingual homophones. *Psychophysiology, 49*(4), 531-543.

Carreras, I., Miorandi, D., Canright, G. S., & Engo-Monsen, K. (2007). Eigenvector centrality in highly partitioned mobile networks: Principles and applications. *Studies in Computational Intelligence, 69*, 123-145.
Castro, N., & Siew, C. S. Q. (2020). Contributions of modern network science to the cognitive sciences: Revisiting research spirals of representation and process. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 476*, 20190825.
Castro, N., & Stella, M. (2018). The multiplex interplay between phonological and semantic networks impacts word production across different types of aphasia. *Frontiers in Human Neuroscience, 12*. doi:10.3389/conf.fnhum.2018.228.00035
Castro, N., Stella, M., & Siew, C. S. Q. (2019). Quantifying the interplay of semantics and phonology during failures of word retrieval by people with aphasia using a multiplex lexical network. *Cognitive Science, 44*(9), e12881.
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*(6), 1934-1949.
Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science, 34*, 685-697.
Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicon. *Journal of Child Language, 17*(1), 205-215.
Chen, B.-S., & Lin, Y.-P. (2011). On the interplay between evolvability and network robustness in an evolutionary biological network: A systems biology approach. *Evolutionary Bioinformatics*. doi:10.4137/EBO.S8123
Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693-713.
Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417-430.
Chen, Q., & Mirman, D. (2014). Interaction between phonological and semantic representations: Time matters. *Cognitive Science, 39*(3), 538-558. doi: https://doi.org/10.1111/cogs.12156
Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51*(4), 661-703.
Cohen, R., & Havlin, S. (2010). *Complex networks: Structure, robustness and function*. New York: Cambridge University Press.
Collins, A., & Loftus, E. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.
Cong, J., & Haitao, L. (2014). Approaching human language with complex networks. *Physics of Life Reviews, 11*(4), 598-618.
Cooper, C., & Frieze, A. (2003). A general model of web graphs. *Random Structures & Algorithms, 22*(3), 311-335.
Crossley, S. A., Skalicky, S., Kyle, K., & Monteiro, K. (2019). Absolute frequency effects in second language lexical acquisition. *Studies in Second Language Acquisition, 41*, 721-744.
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal: Complex Systems, 1695*.
D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit techniques*. New York, Basel: Marcel Dekker, Inc. .
Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, 28*, 125-127.

De Ambroggio, U., Polito, F., & Sacerdote, L. (2020). On dynamic random graphs with degree homogenization via anti-preferential attachment probabilities. *Physica D: Nonlinear Phenomena, 414*, 132689.
de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language, 67*, 297-319.
De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. (2017). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Frontiers of cognitive psychology: Big data in cognitive science* (pp. 174-202). New York: Routledge/ Taylor & Francis Group.
De Deyne, S., Navarro, D., Perfors, A., Brysbaert, M., & Storms, G. (2019). "The small world of words" English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*, 987-1006.
Dearing, J. W. (2008). Evolution of diffusion and dissemination theory. *Journal of Public Health Management & Practice, 14*(2), 99-108.
Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software, 64*(4), 1-34.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychologial Review, 93*(3), 283-321.
Dell, G. S., & Gordon, J. (2003). Neighbors in the lexicon: Friends or foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9-38). London: De Gruyter Mouton.
Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory & Language, 56*, 490-520.
Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. G., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychologial Review, 104*(4), 801-838.
Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience, 21*, 1893-1906.
Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, R. H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language, 62*, 284-301.
Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*(4), 445-464.
Dobson, A. J. (2002). *An introduction to generalized linear models*. Boca Raton: Chapman & Hall/ CRC.
Dóczi, B. (2019). An overview of conceptual models and theories of lexical representation in the mental lexicon. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 46-65). New York: Routledge.
Dóczi, B. (2020). An overview of conceptual models and theories of lexical representation in the mental lexicon. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 46-65). London, New York: Routledge.
Donnelly, S., & Kidd, E. (2020). Individual differences in lexical processing efficiency and vocabulary in toddlers: A longitudinal investigation. *Journal of Experimental Child Psychology, 192*, 104781.
Dorogovtsev, S. N., & Mendes, J. F. F. (2003). Accelerated growth of networks. In S. Bornholdt & H. G. Schuster (Eds.), *Handbook of graphs and networks* (pp. 318-341). Berlin: Wiley-VCH.

Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (2000a). Structure of growing networks with preferential linking. *Physical Review Letters, 85*, 4633.
Dorogovtsev, S. N., Mendes, J. F. F., & Samukhin, A. N. (2000b). Structure of growing networks with preferential linking. *Physical Review Letters, 85*(21), 4633-4636.
Downey, B., Hallmark, B., Cox, M. P., Norquest, P., & Lansing, J. S. (2008). Computational feature-sensitive reconstruction of language relationships: Developing the ALINE distance for comparative historical linguistic reconstruction. *Journal of Quantitative Linguistics, 15*(4), 340-369.
Downey, S. S., Sun, G., & Norquest, P. (2017). alineR: An R package for optimizing feature-weighted alignments and linguistic distances. *The R Journal, 9*(1), 138-152.
Einstein, A. (1924). Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, 261-267.
Elgort, I., Beliaeva, N., & Boers, F. (2020). Contextual word learning in the first and second language. *Studies in Second Language Acquisition, 42*, 7-32.
Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.
Epskamp, S., Cramer, A. O. J., Waldorp, L., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1-18.
Erdős, P., & Rényi, A. (1959). On random graphs I. *Publ. Math. Debrecen, 6*, 290-297.
Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology, 12*, 263-282.
Estrada, E. (2006). Network robustness to targeted attacks: The interplay of expansibility and degree distribution. *The European Physical Journal B - Condensed Matter and Complex Systems, 52*, 563-574.
Estrada, E. (2016). When local and global clustering of networks diverge. *Linear Algebra and its Applications, 488*, 249-263.
Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry, 8*(3), 505-520.
Feather, N. (1971). Organization and discrepancy in cognitive structures. *Psychological Review, 78*, 355-379.
Ferretti, L., Cortelezzi, M., Yang, B., Marmorini, G., & Bianconi, G. (2012). Features and heterogeneities in growing network models. *Physical Review E, 85*, 066110.
Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.
Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology, 65*, 47-55.
Fotouhi, B., & Rabbat, M. G. (2013). Network growth with arbitrary initial conditions: Degree dynamics for uniform and preferential attachment. *Physical Review E, 88*(6), 062801.
Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children’s semantic and phonological networks: Insights from 10 languages. *Cognitive Science, 44*(7), e12847.
Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory & Language, 43*, 182-216.
Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215-239.
Fricke, M., Baese-Berk, M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience, 31*(5), 639-645.

Friedrich, C. K., Felder, V., Lahiri, A., & Eulitz, C. (2013). Activation of words with phonological overlap. *Frontiers in Psychology*. doi:https://doi.org/10.3389/fpsyg.2013.00556
Frisch, S. A. (2011). Frequency effects. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology: Phonological interfaces* (pp. 2137-2163). Chichester, UK: Blackwell Publishing Ltd.
Gagen, M. J., & Mattick, J. S. (2005). Accelerating, hyperaccelerating, and decelerating networks. *Physical Review Letters, 72*, 016123.
Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*(4), 789-806.
Garlaschelli, D., & Loffredo, M. (2004). Fitness-dependent topological properties of the word trade web. *Physical Review Letters, 93*, 188701.
Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105-132.
Gass, S., & Selinker, L. (2008). *Second language acquisition: An introductory course*. New York: Routledge.
Gerometta, J. A. (2015). *Network modeling of the developing mental lexicon: Phonological links within and between lexical communities.* (PhD). The City University of New York,
Ghoshal, G., Chi, L., & Barabási, A. L. (2013). Uncovering the role of elementary processes in network evolution. *Scientific Reports, 3*, 2920.
Gilhooly, K. J., & Watson, F. L. (1981). Word age-of-acquisition effects: A review. *Current Psychological Reviews, 1*(3), 269-286.
Gillespie, C. S. (2014). Fitting heavy tailed distributions: The poweRlaw package. *arXiv [Preprints].*
Glazer, A. M., & Wark, J. S. (2001). *Statistical mechanics*. Oxford, UK: Oxford University Press.
Goh, W. D., Suarez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word frequency effects. *Psychonomic Bulletin & Review, 16*, 882-887.
Goldbeck, J. (2015). *Introduction to social media investigation*. Amsterdam: Elsevier.
Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language, 28*(5), 501-518.
Goldrick, M., Folk, J. R., & Rapp, B. (2010). Mrs. Malaprop's neighborhood: Using word errors to reveal neighborhood structure. *Journal of Memory and Language, 62*(2), 113-134.
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.
Goldstein, R., & Vitevitch, M. S. (2017). The influence of closeness centrality on lexical processing. *Frontiers in Psychology, 8*, 1683.
Gomez-Rodriguez, M., Leskovec, J., & Krause, A. (2012). Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data, 5*(4), 1-37.
Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515-531.
Gordon, J. K. (2002). Phonological neighborhood effects in aphasic speech errors: Spontaneous and structured contexts. *Brain and Language, 82*(2), 113-145.

Gráf, T. (2017). The story of the learner corpus LINDSEI_CZ. *Studies in Applied Linguistics, 8*(2), 22-35.
Gravino, P., Servedio, V. D. P., Barrat, A., & Loreto, V. (2012). Complex structures and semantics in free word association. *Advances in Complex Systems, 15*(3-4), 1250054-1250075.
Greenberg, J. H. (1966). *Language universals: With special reference to feature hierarchies*. Berlin: Mouton de Gruyter.
Griffiths, T., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science, 18*, 1069-1076.
Gruenenfelder, T. M., & Pisoni, D. B. (2009). The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons. *Journal of Speech, Language, and Hearing Research, 52*(3), 596-609.
Haggett, P., & Chorley, R. J. (1972). *Network analysis in geography*. London: Edward Arnold.
Haigh, C. A., & Jared, D. (2007). The activation of phonological representations by bilinguals while reading silently: Evidence from interlingual homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 623-644.
Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology, 4*, 563.
Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89*, 151-174.
Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *Annals of Statistics, 13*(1), 70-84.
Havas, B., Taylor, J. S. H., Vaquero, L., de Diego-Balaguer, R., Rodriguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *Quarterly Journal of Experimental Psychology, 71*(6), 1469-1481.
He, H., & Deng, Y. (2015). The mental lexicon and English vocabulary teaching. *English Language Teaching, 8*(7), 40-45.
Hébert-Dufresne, L., Allard, A., Marceau, V., Noël, P.-A., & Dubé, L. J. (2011). Structural preferential attachment: Network organization beyond the link. *Physical Review Letters, 107*(15), 158702.
Helsen, J., Frickel, J., Jelier, R., & Verstrepen, K. J. (2019). Network hubs affect evolvability. *PloS Biology, 17*(1), e3000111.
Hills, T. T., Maouene, J., Riordan, B., & Smith, L. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory & Language, 63*(3), 259-273.
Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Categorical structure among shared features in networks of early-learned nouns. *Cognition, 112*(3), 381-396.
Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Longitudinal analysis of early semantic networks. *Psychological Science, 20*(6), 729-739.
Hoffmann, T., Peel, L., Lambiotte, R., & Jones, N. S. (2020). Community detection in networks without observing edges. *Science Advances, 6*(4). doi:10.1126/sciadv.aav1478
Hohenberger, A., & Peltzer-Karpf, A. (2009). Language learning from the perspective of nonlinear dynamic systems. *Linguistics, 47*(2), 481-511.

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probabillity and neighborhood density on word learning by preschool children. *Journal of Memory & Language, 63*(1), 100-116.
Hu, P., & Lee, L. (2020). Community-based link-addition strategies for mitigating casading failures in modern power systems. *Processes, 8*(2). doi:https://doi.org/10.3390/pr8020126
Huang, S., Bian, X., Wu, G., & McLemore, C. (1997). CALLHOME Mandarine Chinese Lexicon. *Linguistic Data Consortium, University of Pennsylvania*.
Humphries, M. D., & Gurney, K. (2008). Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS ONE, 3*(4), e0002051.
Humphries, M. D., Gurney, K., & Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not scale-free, network. *Proceedings of the Royal Society B: Biological Sciences, 273*(1585), 503-511.
iCancho, R. F., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B: Biological Sciences, 268*, 2261-2265.
Iyengar, S. R., Veni Madhavan, C. E., Zweig, K. A., & Natarajan, A. (2012). Understanding human navigation using network analysis. *Topics in Cognitive Science, 4*, 121-134.
Jakobson, R. (1941/68). *Child language, aphasia and phonological universals*. The Hague: Mouton.
James, L. E., & Burke, D. M. (2000). Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1378-1391.
Janson, S., Luczak, T., & Rucinski, A. (2000). *Random graphs*. New York: Wiley.
Jescheniak, J. D., & Schriefers, H. (1998). Serial discrete versus cascaded processing in lexical access in speech production: Further evidence from the co-activation of near-synonyms. *Journal of Experimental Psychology: Language, Memory, and Cognition, 24*, 1256-1274.
Jiang, J., Yu, W., & Liu, H. (2019). Does scale-free syntactic network emerge in second language learning? *Frontiers in Psychology, 10*, 925.
Joanisse, M. F., & McClelland, J. L. (2015). Connectionist perspectives on language learning, representation and processing. *WIREs Cognitive Science, 6*(3), 235-247. doi: doi: 10.1002/wcs.1340
Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26*, 240-346.
Kane, E. A., & Higham, T. E. (2015). Complex systems are more than the sum of their parts: Using integration to understand performance, biomechanics, and diversity. *Integrative and Comparative Biology, 55*(1), 146-165.
Kapatsinski, V. (2006). Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech Research Lab Progress Report, 27*, 133-152.
Karimi, H., & Diaz, M. (2020). When phonological neighborhood density both facilitates and impedes: Age of acquisition and name agreement interact with phonological neighborhood during word production. *Memory & Cognition, 48*, 1061-1072.
Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika, 18*, 39-43.
Kauffman, S. A. (1993). *The origin of order: Self-organization and selection in evolution*. London: Oxford University Press.
Kaushanskaya, M., Yoo, J., & Van Hecke, S. (2013). Word learning in adults with second-language experience: Effects of phonological and referent familiarity. *Journal of Speech, Language, and Hearing Research, 56*(2), 667-678.

Kaushanskaya, M., Yoo, J., & Van Hecke, S. (2014). Word learning in adults with second language experience: Effects of phonological and referent familiarity. *Journal of Speech and Hearing Research, 56*(2), 667-678.
Keeling, M. J., & Eames, K. T. D. (2005). Networks and epidemic models. *Journal of the Royal Society: Interfact, 22*(4), 295-307.
Kephart, J. O., & White, S. R. (1991). *Directed-graph epidemiological models of computer viruses*. Paper presented at the Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA.
Kim, P. M., Korbel, J. O., & Gerstein, M. B. (2007). Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 20274-20279.
Kjaergaard, M., Brander, S., & Poulsen, F. M. (2010). Small but slow world: How network typology and burstiness slow down spreading. *Physical Review E, 83*(83), 602-608.
Kleinberg, J. M. (2000). The small-world phenomonon: An algorithmic perspective. *Proceedings of the 32nd Annual ACN Symposium on Theory of Computing*, 163-170.
Ko, K., Lee, K. J., & Park, C. (2008). Rethinking preferential attachment scheme: Degree centrality versus closeness centrality. *Connections, 28*(1), 4-15.
Kondrak, G. (2000). A new algorithm for the alignment of phonetic sequence. *Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference*, 288-295.
Kong, J. S., & Roychowdhury, V. P. (2008). Preferential survival in models of complex ad hoc networks. *Physica A: Statistical Mechanics and its Applications, 387*(13), 3335-3347.
Koubkova-Yu, T. C.-T., Chao, J.-C., & Leu, J.-Y. (2018). Heterologous Hsp90 promotes phenotypic diversity through network evolution. *PloS Biology, 16*(11), e2006450.
Krapivsky, P. L., & Redner, S. (2001). Organization of growing random networks. *Physical Review E, 6*(2), 066123.
Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990.
Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.
Lara-Martinez, P., Quintana-Obregon, B., Reyes-Manzano, C. F., Lopez-Rodriguez, I., & Guzman-Vargas, L. (2021). Comparing phonological and orthographic networks: A multiplex analysis. *PLoS ONE*. doi:https://doi.org/10.1371/journal.pone.0245263
Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters, 87*, 198701.
Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*, 306-353.
Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Science, 3*(6), 223-232.
Levelt, W. J. M. (2001). Relations between speech production and speech perception: Some behavioral and neurological observations. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 241-256). Cambridge, MA: MIT Press.
Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-38.
Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*, 707-710.

Levy, O., Kenett, Y. N., Oxenberg, O., Castro, N., De Deyne, S., Vitevitch, M. S., & Havlin, S. (2021). Unveiling the nature of interaction between semantics and phonology in lexical access based on multilayer networks. *Scientific Reports, 11*, 14479. doi:https://doi.org/10.1038/s41598-021-93925-y
Li, C., Wang, M., & Davis, J. A. (2017). The phonological preparation unit in spoken word production in a second language. *Bilingualism: Language and Cognition, 20*(2), 351-366.
Li, D., Zhang, Q., Zio, E., Havlin, S., & Kang, R. (2015). Network reliability analysis based on percolation theory. *Reliability Engineering & System Safety, 142*, 556-562.
Li, L., Alderson, D., Doyle, J. C., & Willinger, W. (2005). Towards a theory of scale-free graphs: Definitions, properties, and implications. *Internet Mathematics, 2*(3), 431-523.
Li, X., & Chen, G. (2003). A local-world evolving network model. *Physica A: Statistical Mechanics and its Applications, 328*(102), 274-286.
Li, X., Jin, Y. Y., & Chen, G. (2003). Complexity and synchronization of the world trade web. *Physica A: Statistical Mechanics and its Applications, 328*(1-2), 287-296.
Linehan, J., Gross, M., & Finn, J. (1995). Greenway planning: developing a landscape ecological network approach. *Landscape and Urban Planning, 33*, 179-193.
Liu, J., Li, J., Chen, Y., Chen, X., Zhou, Z., Yang, Z., & Zhang, C.-J. (2019). Modeling complex networks with accelerating growth and aging effect. *Physics Letters A, 383*(13), 1396-1400.
Liu, J. S., & Kuan, C.-H. (2016). A new approach for main path analysis: Decay in knowledge diffusion. *Journal of the Association for Information Science and Technology, 67*(2), 465-476.
Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics, 39*, 155-158.
Luce, P. A., Goldinger, S., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics, 62*(3), 615-625.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1-36.
Luce, R. D., & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika, 14*, 95-116.
Luef, E. M. (submitted). Phonological neighborhood complexity and sound change.
Luef, E. M., Ghebru, B., & Ilon, L. S. (2018). Apps for language learning: their use across different languages in a Korean context. *Interactive Learning Environments*. doi:https://doi.org/10.1080/10494820.2018.1558255
Lugosi, G., & Pereira, A. S. (2019). Finding the seed of uniform attachment trees. *Electronic Journal of Probability, 24*(18), 1-15.
Macklin-Cordes, J. L., & Round, E. R. (2020). Re-evaluating phoneme frequencies. *Frontiers in Psychology, 11*, 3181. doi:10.3389/fpsyg.2020.570895
Maddieson, I. (2009). Calculating phonological complexity. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupe (Eds.), *Approaches to phonological complexity* (pp. 83-110). Berlin, New York: de Gruyter Mouton.
Maechler, M. (2021). Package 'diptest'. *R package*. doi:https://cran.r-project.org/web/packages/diptest/diptest.pdf
Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133-156.

Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired-associate learning. *Psychonomic Bulletin & Review, 27*, 1059-1069.
Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Foryś-Nogala, M., Kutyłowska, K., & Wodniecka, Z. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition, 206*, 104477.
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.
Marian, V., & Blumenfeld, H. K. (2006). Phonological neighborhood density guides: Lexical access in native and non-native language production. *Journal of Social and Ecological Boundaries, 2*(1), 3-35.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.
Marslen-Wilson, W. D. (1990). Activation, competition and frequency in lexical access. In G. T. M. Altman (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148-172). Cambridge, M. A.: MIT Press.
Marslen-Wilson, W. D., Brown, C. M., & Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes, 3*, 1-16.
Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access. *Psychological Review, 101*, 653-675.
Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.
Martin, C., & Niemeyer, P. (2020). On the impact of network size and average degree on the robustness of centrality measures. *Network Science*, 1-22.
Martindale, C., Gusein-Zade, S. M., McKenzie, D., & Borodovsky, M. Y. (1996). Comparison of equations describing the ranked frequency distributions of graphemes and phonemes. *Journal of Quantitative Linguistics, 3*, 106-112.
Mayor, J., & Plunkett, K. (2009). Using TRACE to model infant sensitivity to vowel and consonant mispronunciations. *Proceedings of the Annual Meeting of the Cognitive Science Society, 31*(31), 1816-1821.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.
McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4*(4), 310-322.
McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 3-44). Cambridge, MA: MIT Press.
McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics, 39*, 74-92.
Mendes, G., & Da Silva, L. (2009). Generating more realistic complex networks from power-law distribution of fitness. *Brazilian Journal of Physics, 39*(2A), 423-427.
Mendes, J. F. F. (2003). Effect of accelerated growth on network dynamics. In R. Pastor-Satorras, M. Rubi, & A. Diaz-Guilera (Eds.), *Statistical mechanics of complex networks*. Berlin, Heidelberg: Springer.
Metcalf, L., & Casey, W. (2016). *Cybersecurity and applied mathematics*. Amsterdam: Elsevier.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*(1), 69-89.
Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience, 23*(7), 1634-1647.
Milli, L., Rossetti, G., Pedreschi, D., & Giannotti, F. (2018). Active and passive diffusion processes in complex networks. *Applied Network Science, 3*, 42.
Milojević, S. (2010). Power law distributions in information science: Making the case for logarithmic binning. *Journal of the American Society for Information Science and Technology, 61*(12), 2417-2425.
Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 65-79.
Moreira, S., & Hamilton, M. (2010). Goats don't wear coats: An examination of semantic interference in rhyming assessments of reading readiness for English language learners. *Bilingual Research Journal, 30*(2), 547-557.
Mulík, S., Carrasco-Ortiz, H., & Amengual, M. (2018). Phonological activation of first language (Spanish) and second language (English) when learning third language (Slovak) novel words. *International Journal of Bilingualism, 23*(5), 1024-1040.
Naug, D. (2008). Structure of the social network and its influence on transmission dynamics in a honeybee colony. *Behavioral Ecology and Sociobiology, 62*, 1719-1725.
Navarrete, E., Pastore, M., Valentini, R., & Peressotti, F. (2015). First learned words are not forgotten: Age-of-acquisition effects in the tip-of-the-tonngue experience. *Memory & Cognition, 43*(7), 1085-1103.
Neergaard, K. D., Britton, J., & Huang, C.-R. (2019). *Neighborhood in decay: Working memory modulates effect of phonological similarity on lexical access.* Paper presented at the 41st Annual Conference of the Cognitive Science Society.
Neergaard, K. D., Luo, J., & Huang, C.-R. (2019). Phonological network fluency identifies phonological restructuring through mental search. *Scientific Reports, 9*, 15984.
Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 404-409.
Newman, M. E. J. (2002a). Assortative mixing in networks. *Physical Review Letters, 89*(20), 208701.
Newman, M. E. J. (2002b). Assortative mixing in networks. *Physical Review Letters, 89*, 20889701.
Newman, M. E. J. (2003a). Properties of highly clustered networks. *Physical Review E, 68*, 026121.
Newman, M. E. J. (2003b). The structure and function of complex networks. *SIAM Review, 45*(3), 167-256.
Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 5200-5205.
Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*(5), 323-351.
Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 8577-8582.
Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, UK: Oxford University Press.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 026113.
Newman, M. E. J., Strogatz, S. H., & Watts, D. J. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 2566-2572.
Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech communication, 47*, 43-58.
Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring familiarity of 20,000 words. *Research on Speech Perception Progress Report, 10*, 357-376.
O'Séaghdha, P. G., & Frazer, A. K. (2014). The exception does not rule: Attention constraints form preparation in word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 797-810.
O'Séaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282-302.
Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18*, 340-353.
Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2010). Phonological advance planning in sentence production. *Journal of Memory and Language, 63*(4), 526-540.
Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks, 32*, 245-251.
Pachon, A., Sacerdote, L., & Yang, S. (2018). Scale-free behavior of networks with the copresence of preferential and uniform attachment rules. *Physica D: Nonlinear Phenomena, 371*, 1-12.
Pachon, L., Sacerdote, L., & Yang, S. (2018). Scale-free behavior of networks with the copresence of preferential and uniform attachment rules. *Physica D: Nonlinear Phenomena, 371*, 1-12.
Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank citation ranking: Bringing order to the web (technical report)*. Retrieved from Stanford, CA:
Papadopoulos, F., Kitsak, M., Serrano, A., Boguna, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature, 489*, 537-540.
Payne, J. L., & Wagner, A. (2014). The robustness and evolvability of transcription factor binding sites. *Science, 343*, 875-877.
Peel, L., Delvenne, J.-C., & Lambiotte, R. (2018). Multiscale mixing patterns in networks. *Proceedings of the National Academy of Sciences of the United States of America, 115*(16), 4057-4062.
Perline, R. (2005). Strong, weak and false invese power law. *Statistical Sciences, 20*(12), 68-88.
Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Language, Memory, and Cognition, 24*, 539-557.
Pham, T., Sheridan, P., & Shimodaira, H. (2015). PAFit: A statistical method for measuring preferential attachment in temporal complex networks. *PLoS ONE, 10*(9), e0137796.
Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*, 1112-1130.
Piantadosi, S. T., Tily, H. J., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280-291.
Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics, 9*, 75-82.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*(1), 75-95.

Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult learning of novel words in a non-native languagew: Consonants, vowels, and tones. *Frontiers in Psychology, 9*, 1211. doi:10.3389/fpsyg.2018.01211

Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communities in networks. *Notices of the American Mathematical Society, 56*, 1082-1097.

Quillian, R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science, 12*, 410-430.

Quinn, G. P., & Keough, M. J. (2002). *Experimental designs and data analysis for biologists*. Cambridge, UK: Cambridge University Press.

Ravasz, E., & Barabási, A. (2003). Hierarchical organization in complex networks. *Physical Review E, 67*, 026112.

Redner, S. (2005). Citation statistics from 110 years of Physical Review. *Physics Today, 58*, 49-54.

Ren, H.-P., Song, J. Y., Yang, R., Baptista, M. S., & Grebogi, C. (2016). Cascade failure analysis of power grid using new load distribution law and node removal rule. *Physica A: Statistical Mechanics and its Applications, 442*, 239-251.

Ren, J., Yang, X., Yang, L.-X., Xu, Y., & Yang, F. (2012). A delayed computer virus propagation model and its dynamics. *Chaos, Solitons & Fractals, 45*(1), 74-79.

Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health, 61*(1), 933-937.

Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage*. New York: Columbia University Press.

Roach, P. (2004). British English: Received Pronunciation. *Journal of the International Phonetic Association, 34*(2), 239-245.

Roberts, F. S. (1976). *Discrete mathematical models*. New Jersey: Prentice-Hall.

Rodriguez, F. A. (2019). Network centraliry: An introduction. In E. Macau (Ed.), *A mathematical modeling approach from nonlinear dynamics to complex systems.* (Vol. Nonlinear Systems and Complexity 22, pp. 177-196). Cham: Springer.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review, 12*, 33-37.

Rogers, E. M. (2003). *Diffusion of innovations (5th ed.)*. New York: Free Press.

Romani, C., Galuzzi, C., Guariglia, C., & Goslin, J. (2017). Comparing phoneme frequency, age of acquisition, and loss in aphasia: Implications for phonological universals. *Cognitive Neuropsychology, 34*(7-8), 449-471.

Rosvall, M., Delvenne, J.-C., Schaub, M. T., & Lambiotte, R. (2017). The many facets of community detection in complex networks. *Applied Network Science, 1*.

Saavedra, S., Reed-Tsochas, F., & Uzzi, B. (2008). Asymmetric disassembly and robustness in declining networks. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 16466-16471.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology, 68*, 33-58.

Salavati, C., Abdollahpouri, A., & Manbari, Z. (2019). Ranking nodes in complex networks based on local structure and improving closeness centrality. *Neurocomputing, 336*, 36-45.

Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science, 163*, 292-299.

Sanders, R. (1987). The Pareto principle: Its use and abuse. *Journal of Services Marketing, 1*(2), 37-40.
Scarborough, R. (2013). Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *Journal of Phonetics, 41*(6), 491-508.
Schank, T., & Wagner, D. (2005). Approximating clustering coefficient and transitivity. *Journal of Graph Algorithms and Applications, 9*(2), 265-275.
Schriefers, H., & Vigliocco, G. (2015). Psychology of speech production. In J. D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 225-258). Amsterdam et al.: Elsevier.
Schulpen, B., Dijkstra, T., Schriefers, H., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human perception and performance, 29*(6), 1155-1178.
Schur, E. (2007). Insights into the structure of L1 and L2 vocabulary networks: Intimations of small worlds. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 182-204). Cambridge: Cambridge University Press.
Scott, J. (1991). *Social network analysis: A handbook*. London: Sage.
Sebastián Gallés, N., Cuetos Vega, F., Carreiras Valiña, M. F., & Martí Antonin, M. A. (2000). *Lexesp: Léxico informatizado del español*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.
Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*, 91-127.
Shai, S., Stanley, N., Granell, C., Taylor, D., & Mucha, P. J. (2017). Case studies in network community detection. In R. Light & J. Moody (Eds.), *The Oxford handbook of social networks*. Oxford, UK: Oxford University Press.
Shao, Z.-G., Zou, X.-W., Tan, Z.-J., & Jin, Z.-Z. (2006). Growing networks with mixed attachment mechanisms. *Journal of Physics A: Mathematical and General, 390*(9), 2035.
Sheridan, P., & Onodera, T. (2018). A preferential attachment paradox: How preferential attachment combines with growth to produce networks with log-normal in-degree distributions. *Scientific Reports, 8*, 2811.
Shoemark, P., Goldwater, S., Kirby, J., & Sarkar, R. (2016). Toward robust cross-linguistic comparisons of phonological networks. *Proceedings of the 14th ACL SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 110-120.
Siegenfeld, A. F., & Bar-Yam, Y. (2020). An introduction to complex systems science and its applications. *Complexity, 2020*, 6105872. doi:https://doi.org/10.1155/2020/6105872
Siew, C. S. Q. (2013). Community structure in the phonological network. *Frontiers in Psychology, 4*, 553.
Siew, C. S. Q. (2019). An R package to simulate spreading activation in a network. *Behavioral Research, 51*, 910-929. doi:https://doi.org/10.3758/s13428-018-1186-5
Siew, C. S. Q., & Vitevitch, M. S. (2016). Spoken word recognition and serial recall of words from components in the phonological network. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 394-410.
Siew, C. S. Q., & Vitevitch, M. S. (2020a). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General, 149*(12), 2376-2394.
Siew, C. S. Q., & Vitevitch, M. S. (2020b). Investigating the influence of inverse preferential attachment on network development. *Entropy, 22*, 1029.

Siew, C. S. Q., 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.

Simmons, E. S., & Magnuson, J. S. (2018). Word length, proportion of overlap, and phonological competition in spoken word recognition. In C. Kalish, M. Rau, J. Zhu, & T. Rogers (Eds.), *Proceedings of the Cognitive Science Society* (pp. 1062-1067).

Smits, E., Sandra, D., Martensen, H., & Dijkstra, T. (2009). Phonological inconsistency in word naming: Determinants of the interference effect between languages. *Bilingualism: Language and Cognition, 12*, 23-39.

Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism: Language and Cognition, 15*(3), 490-502.

Stella, M. (2020). Multiplex networks quantify robustness of the mental lexicon to catastrophic concept failures, aphasic degradation and ageing. *Physica A: Statistical Mechanics and its Applications, 554*, 124382.

Stella, M., Beckage, N. M., Brede, M., & de Domenico, M. (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific Reports, 8*, 2259.

Stella, M., & Brede, M. (2015). Patterns in the English language: Phonological networks, percolation and assembly models. *Journal of Statistical Mechanics: Theory and Experiment, 2015*(5), P05006.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*(1), 41-78.

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

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

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

Storkel, H. L., & Lee, S.-Y. (2011). The independent effect of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*(2), 191-211.

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE: A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39*, 19-30.

Stumpf, M. P. H., & Porter, M. A. (2012). Mathematics: Critical truths about power laws. *Science, 335*(6069), 665-666.

Suarez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighborhood effects without neighbors. *Psychonomic Bulletin & Review, 18*, 605-611.

Sun, E. D., Michaels, T. C. T., & Mahadevan, L. (2020). Optimal control of aging in complex networks. *Proceedings of the National Academy of Sciences of the United States of America, 117*(34), 20404-20410.

Tambovtsev, Y., & Martindale, C. (2007). Phoneme frequencies follow a Yule distribution. *SKASE Journal of Theoretical Linguistics, 4*, 1-11.

Tanaka, M., & Takahashi, R. (2019). The dynamic characteristics in the L2 mental lexicon (technical report). *IEICE, 119*, T2019-2023.

Tang, J., & Liu, P. (2014). Synchronization in a novel local-world dynamical network model. *Mathematical Problems in Engineering, 2014*, 851403.

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The ubiquity of small-world networks. *Brain Connectivity, 1*(5), 367-375.

The British National Corpus, v. B. X. E. (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. doi:URL: http://www.natcorp.ox.ac.uk/

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 145-152.

Trubetzkoy, N. (1939). *Grundzüge der Phonologie*. Göttingen: van der Hoeck & Ruprecht.

Turnbull, L., Huett, M.-T., Ioannides, A. A., Kininmonth, S., Poeppl, R., Tockner, K., . . . Parsons, A. J. (2018). Connectivity and complex systems: Learning from a multi-disciplinary perspective. *Applied Network Science, 3*, 11. doi:https://doi.org/10.1007/s41109-018-0067-2

Turnbull, R. (2021). Graph-theoretic properties of the class of phonological neighbourhood networks. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 233-240.

Turnbull, R., & Peperkamp, S. (2017). What governs a language's lexicon? Determining the organizing principles of phonological neighbourhood networks. In H. Cherifi, S. Gaito, W. Quattrociocchi, & A. Sala (Eds.), *Complex networks and their applications V* (pp. 83-94). Cham, Switzerland: Springer.

Ullman, M. T. (2007). The biocognition of the mental lexicon. In M. G. Gaskell, G. T. M. Altmann, P. Bloom, A. Caramazza, & P. Levelt (Eds.), *The Oxford handbook of psycholinguistics* (pp. 267-288). Oxford, UK: Oxford University Press.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*(4), 780-789.

van Hemmen, J. L., & Schulten, K. (1995). *Models of neural networks I*. New York: Springer.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*, 458-483.

Van Rensbergen, B., Storms, G., & De Deyne, S. (2015). Examining assortativity in the mental lexicon: Evidence from word associations. *Psychonomic Bulletin & Review, 22*, 1717-1724.

Velleman, S. L., & Vihman, M. M. (2007). Phonology development in infancy and early childhood: Implications for theories of language learning. In M. C. Pennington (Ed.), *Phonology in context* (pp. 25-50). London: Palgrave Macmillan.

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

Vitevitch, M. S. (2002a). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human perception and performance, 28*(2), 270-278.

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

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

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

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.

Vitevitch, M. S. (2021). What can network science tell us about phonology and language processing. *Topics in Cognitive Science*. doi:https://doi.org/10.1111/tops.12532

Vitevitch, M. S., Armbruster, J., & Chu, S. (2004). Sub-lexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 514-529.

Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology, 17*, 13-25.

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

Vitevitch, M. S., Chan, K. Y., & Roodenrys, S. (2012). Complex network structure influences processing in long-terem and short-term memory. *Journal of Memory & Language, 67*(1), 30-44.

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

Vitevitch, M. S., & Goldstein, R. (2014). Keywords in the mental lexicon. *Journal of Memory & Language, 73*, 131-147.

Vitevitch, M. S., Goldstein, R., & Johnson, E. (2016). Path-length and the misperception of speech: Insights from network science and psycholinguistics. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Toward a theoretical framework for analyzing complex linguistic networks: Understanding complex systems* (pp. 29-45). Berlin: Springer.

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

Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2*, 75-94.

Vitevitch, M. S., & Rodriguez, E. (2004). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders, 3*(1), 64-73.

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

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

Vlachos, D. S., Parousis-Orthodoxou, K. J., & Simos, T. E. (2008). Winner-takes-all strategies for complex network evolution. *AIP Conference Proceedings, 1048*. doi:https://doi.org/10.1063/1.3040246

Wagner, A. (2007). Robustness and evolvability: A paradox resolved. *Proceedings of the Royal Society B: Biological Sciences, 275*(1630). doi:https://doi.org/10.1098/rspb.2007.1137

Walsh, M., & Gluck, K. A. (2015). Mechanisms for robust cognition. *Cognitive Science, 39*, 1131-1171.

Wang, D., Wen, Z., Tong, H., Lin, C.-Y., Song, C., & Barabási, A. L. (2011). *Information spreading in context*. Paper presented at the International World Wide Web Conference, Hyderabad, India.

Wang, T., & Zhang, P. (2021). Directed hybrid random networks mixing preferential attachment with uniform attachment mechanisms. *arXiv*. doi:https://arxiv.org/abs/2101.04611

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge, UK: Cambridge University Press.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature, 393*, 440.
Weaver, I. S. (2015). Preferential attachment in randomly grown networks. *Physica A: Statistical Mechanics and its Applications, 439*, 85-92.
Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*, 387-401.
Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of standards and resources for spoken language systems*. Berlin, New York: Mouton de Gruyter.
Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research, 18*, 303-324.
Wilks, C., Meara, P., & Wolter, B. (2005). A further note on simulating word association behaviour in an L2. *Second Language Research, 21*(4), 359-372.
Xu, Y., Feng, Z., & Qi, X. (2021). Signless-laplacian eigenvector centrality: A novel vital nodes identification method for complex networks. *Pattern Recognition Letters, 148*, 7-14.
Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports, 6*, 30750.
Yates, M. (2009). Phonological neighbourhood spread facilitates lexical decisions. *62*(1304-1314).
Yates, M. (2013). How the clustering of phonological neighbors affects visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1649-1656.
Yelland, G. W. (1994). Word recognition and lexical access. In A. Bowers (Ed.), *Encyclopedia of language and linguistics* (pp. 1-11). Edinburgh: Pergamon Press.
Ying, D. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Infometrics, 5*(1), 4633-4636.
Zhang, X., & Gan, C. (2018). Global attractivity and optimal dynamic countermeasure of a virus propagation model in complex networks. *Physica A: Statistical Mechanics and its Applications, 490*, 1004-1018.
Zhao, J., & Xu, K. (2009). Enhancing the robustness of scale-free networks. *Journal of Physics: Mathematical and Theoretical, 42*(19), 195003.
Zhou, B., Meng, X., & Stanley, H. E. (2020). Power-law distribution of degree-degree distance: A better representation of the scale-free property of complex networks. *Proceedings of the National Academy of Sciences of the United States of America, 117*(26). doi:https://doi.org/10.1073/pnas.1918901117
Zhu, H., Wang, X., & Zhu, J.-Y. (2003). Effect of aging on network structure. *Physical Review E, 68*(5), 056121.
Zhu, Q., & Cen, C. (2017). A novel computer virus propagation model under security classification. *Discrete Dynamics in Nature and Society, 2017*, 8609082.
Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory & Language, 48*, 779-793.
Zipf, G. K. (1935). *The psycho-biology of language*. Boston, MA: Houghton-Mifflin.

9. Appendix

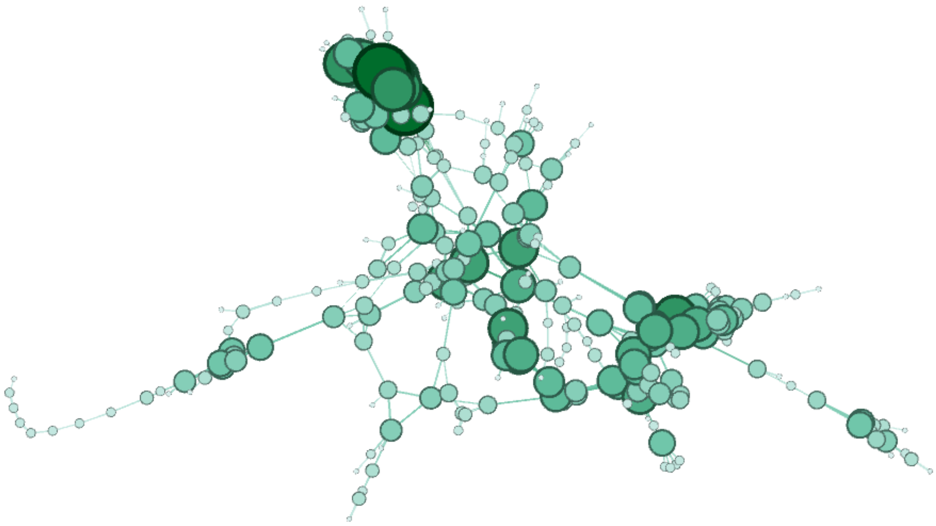
9.1. Appendix A: Network graphs (ESL, BNC)

9.1.1. Micro-level graphs of giant components

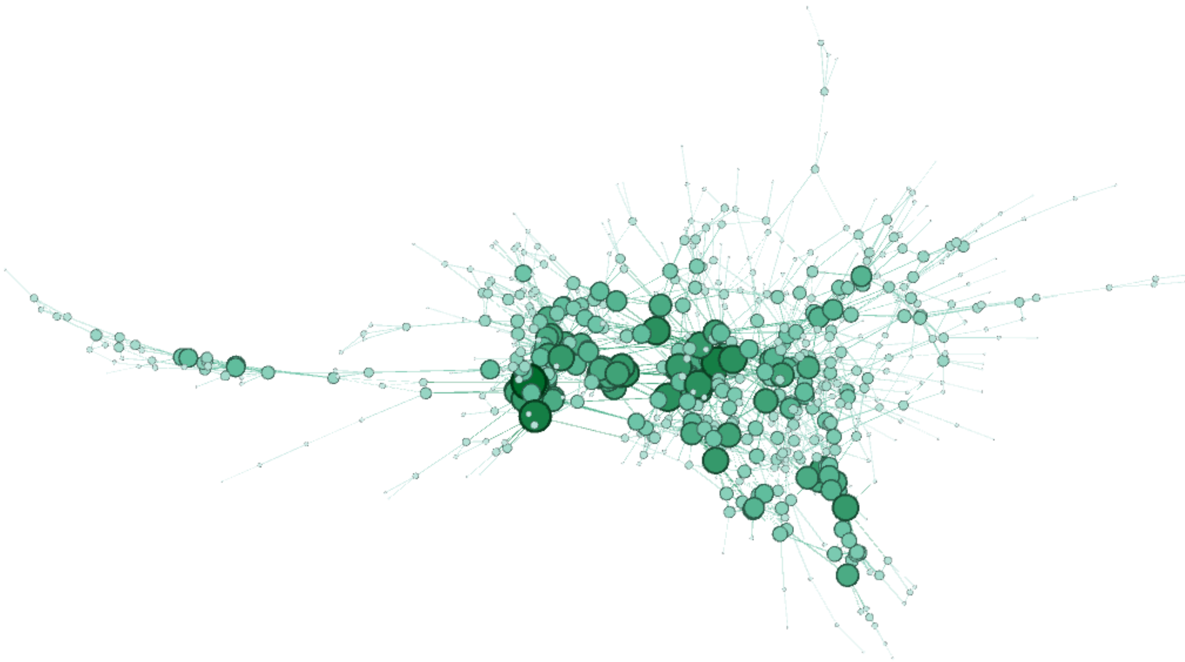
Node sizes and colors according to the respective network centrality measure: larger and darker nodes indicate higher centrality values. Yifan-Hu network layouts (Gephi).

9.1.1.1. Degree centrality

A1

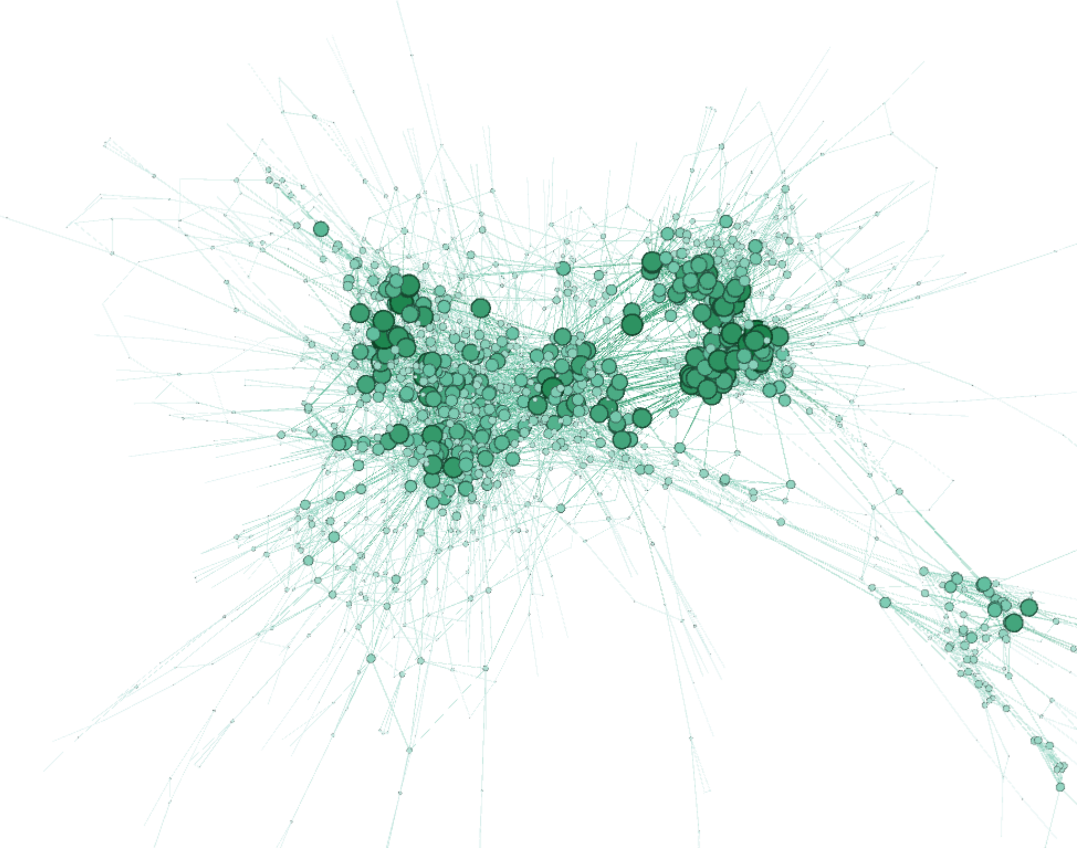


A2



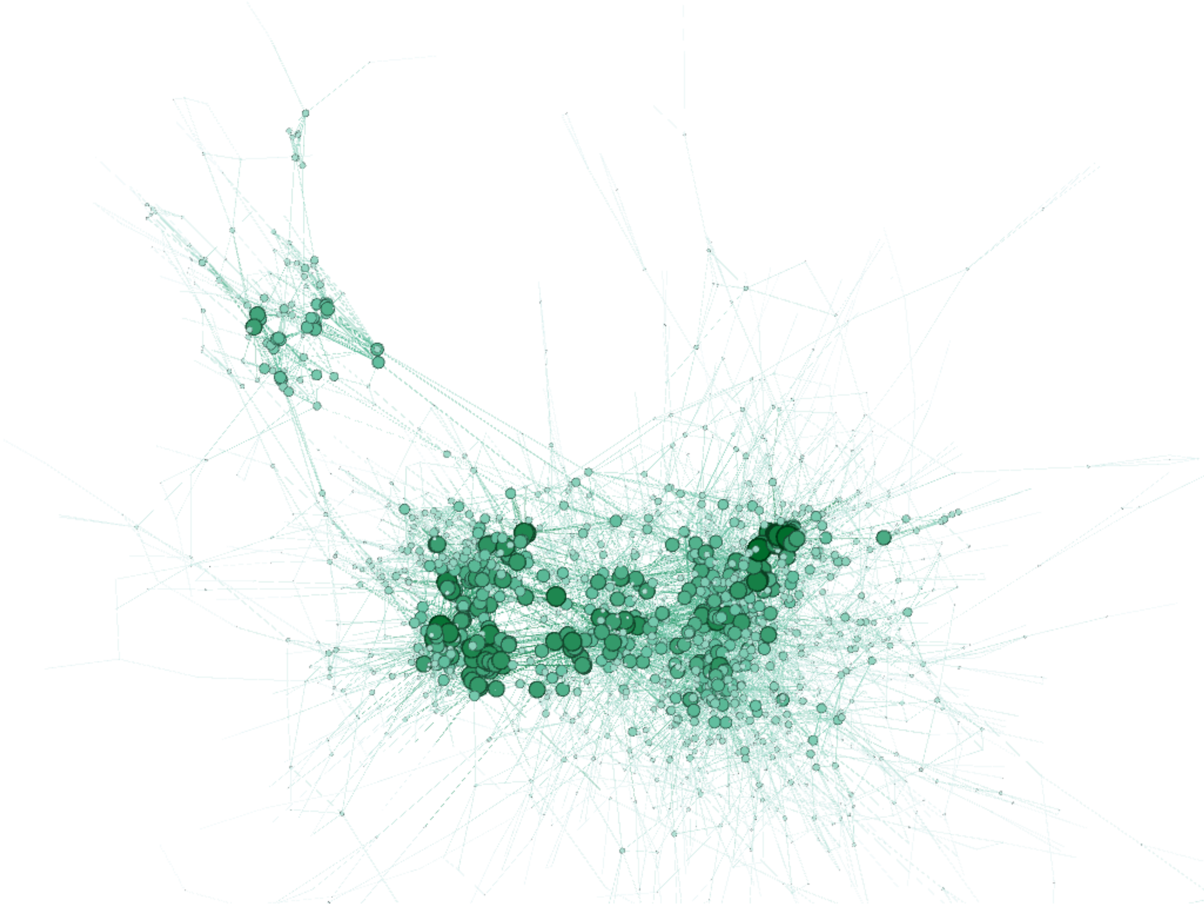
B1

B2



C1

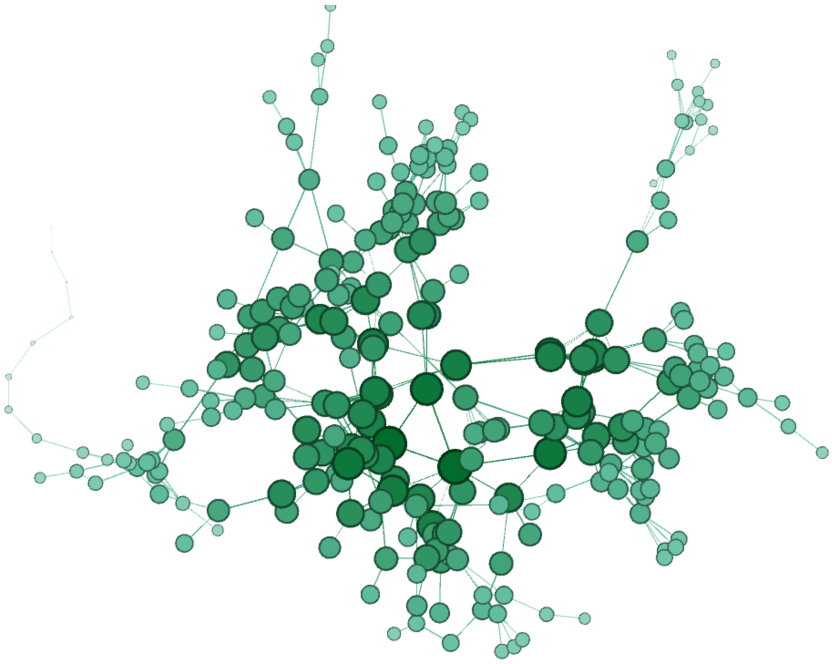
C2



BNC

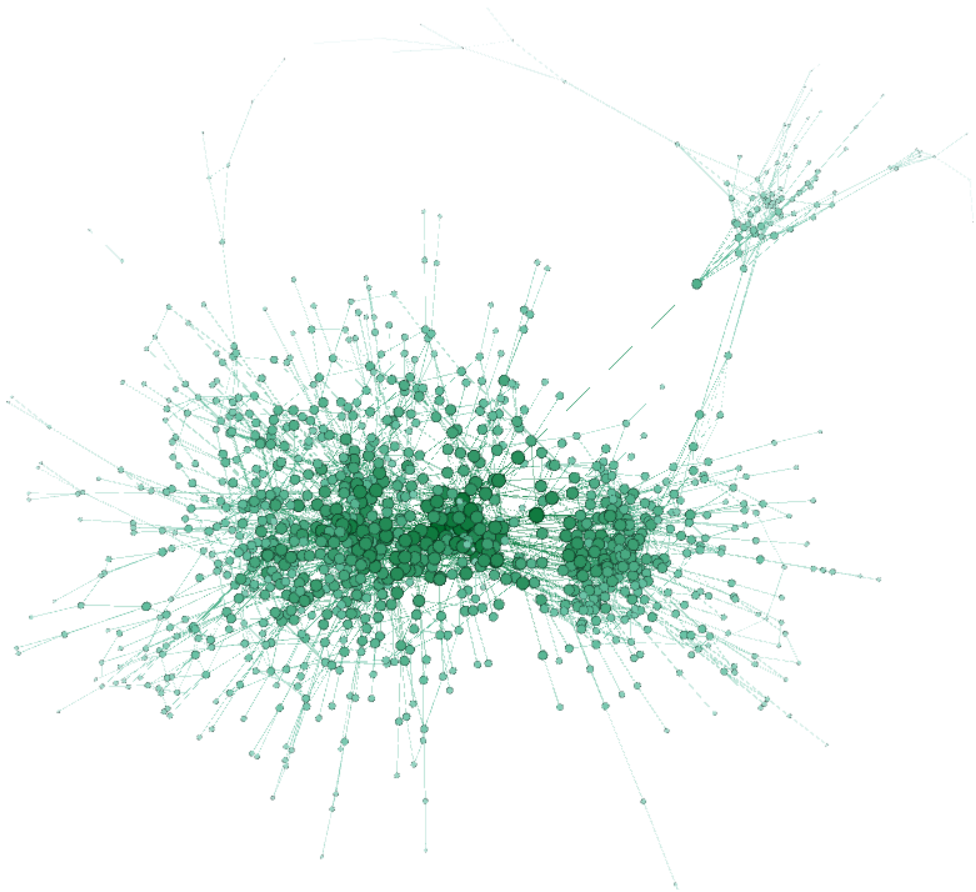
9.1.1.2. Closeness centrality

A1

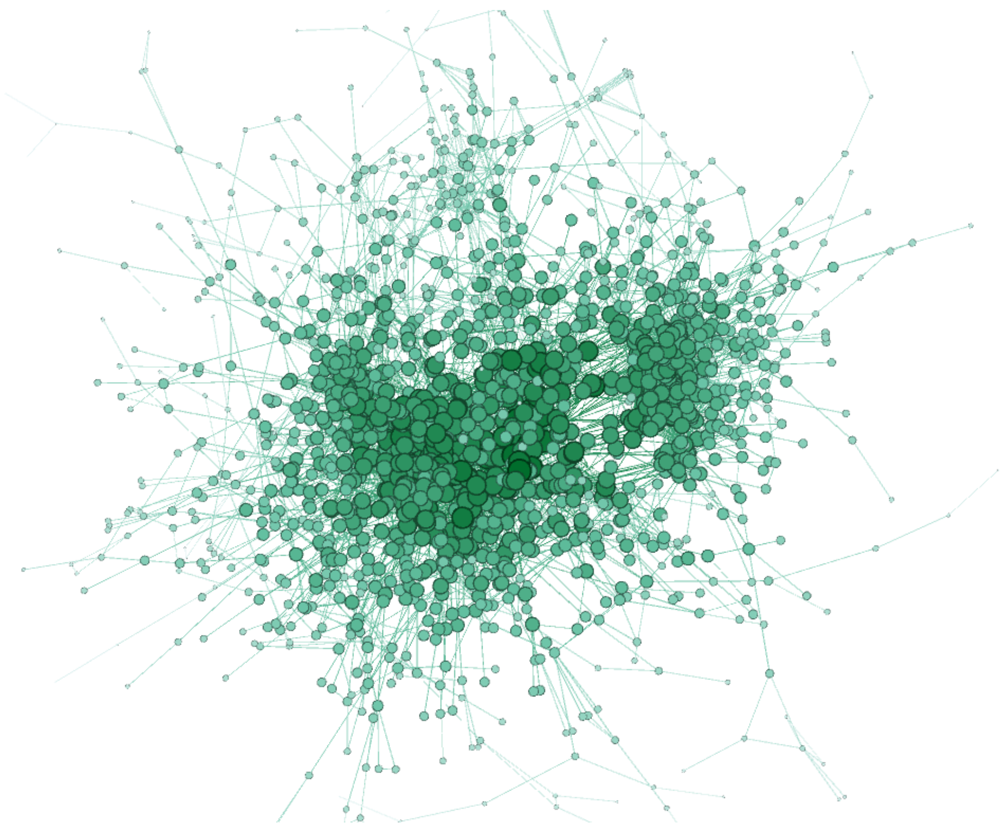


A2

B1



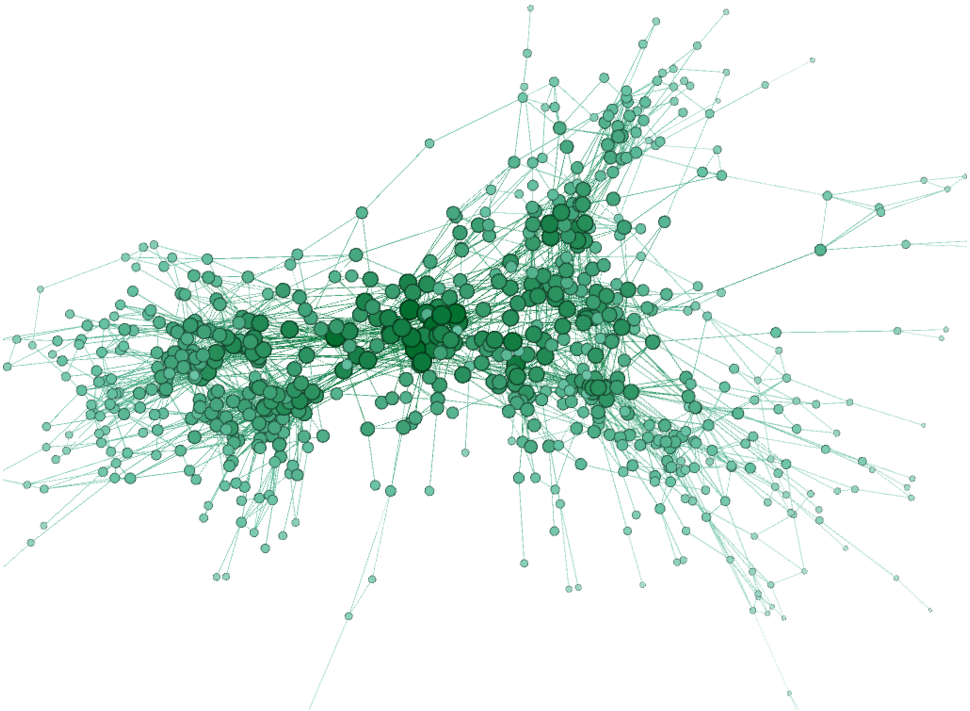
B2



C1

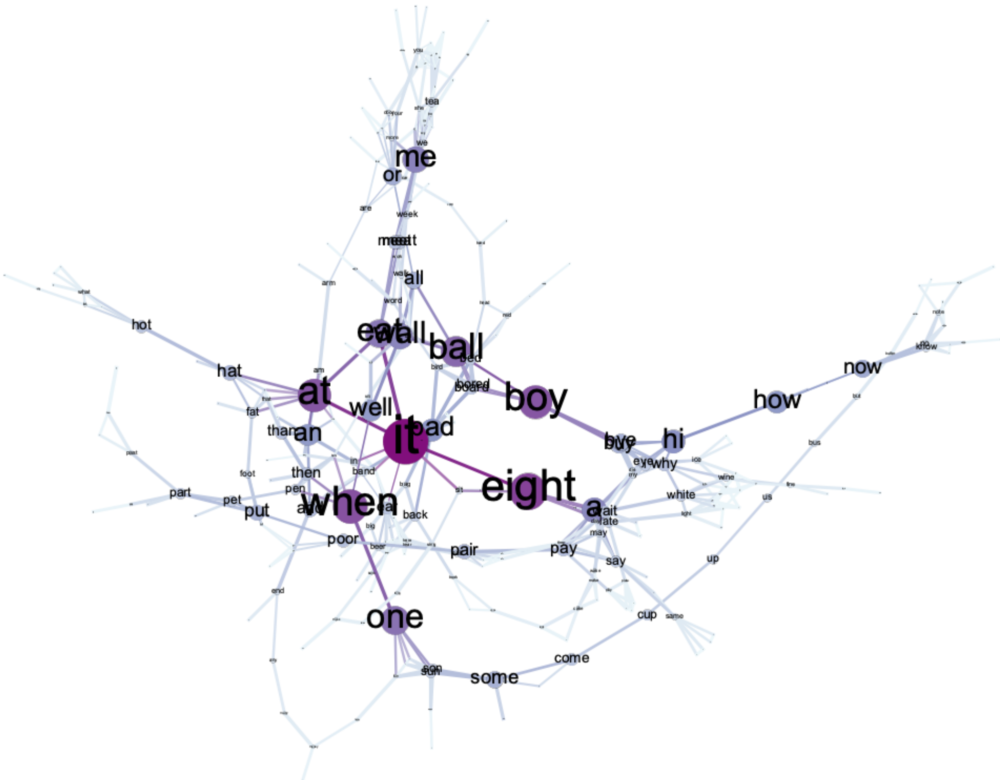
C2

BNC

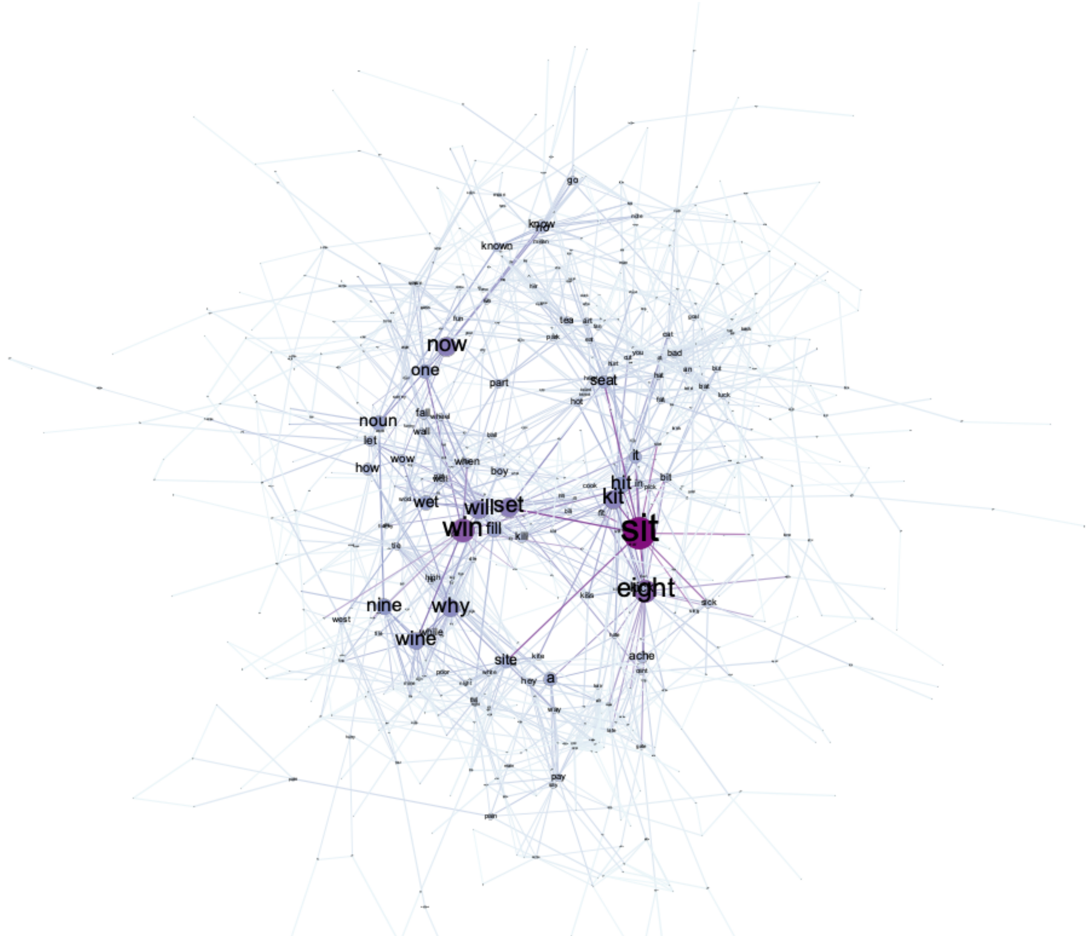


9.1.1.3. Betweenness centrality

A1

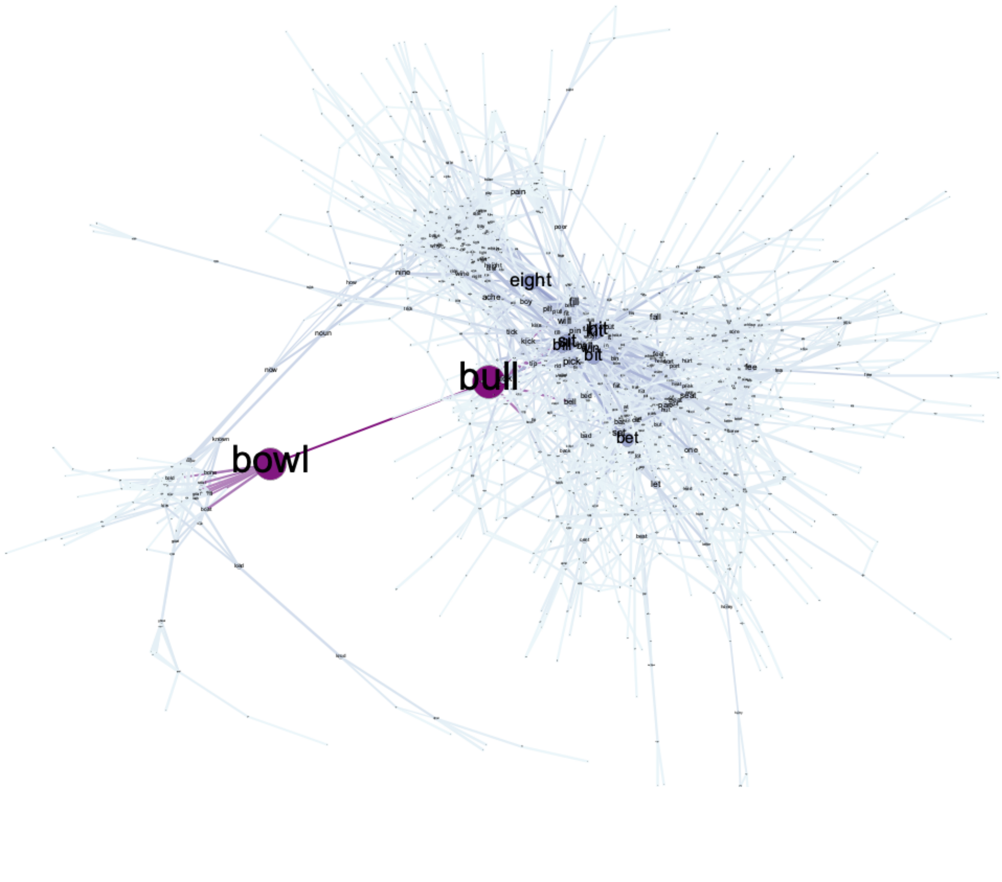


A2



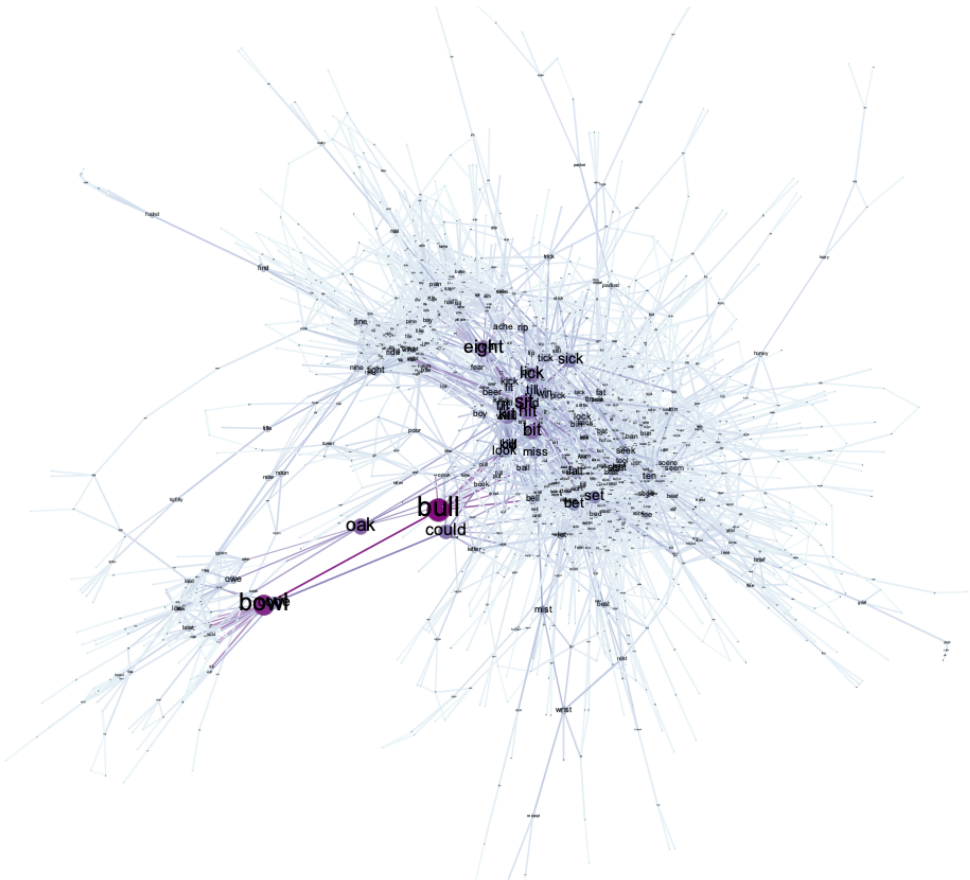
go
know
known
now
one
part
seat
tea
noun
fall
wall
let
how
wow
when
boy
wet
will
set
win
fill
kill
hit
kit
it
sit
eight
nine
why
wine
west
site
a
hey
ache
pay
pain

B1

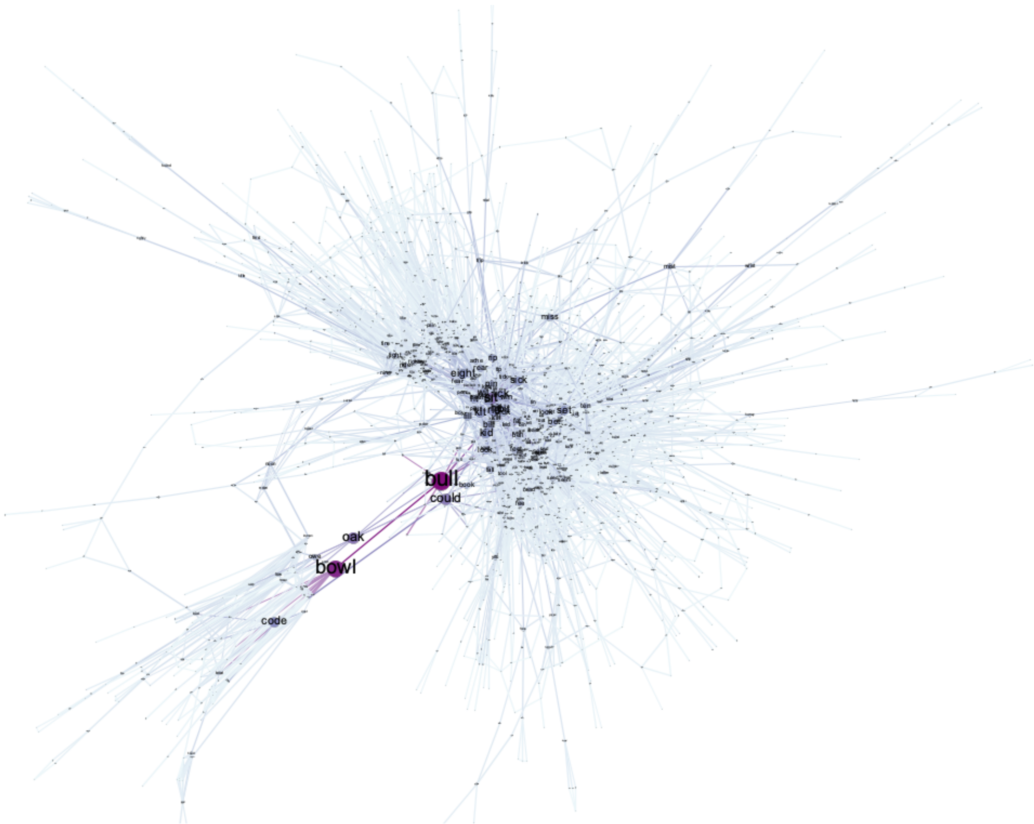


pain
nine
eight
ache
boy
fill
kit
sit
bit
pick
bull
bowl
bet
one
let
noun
now

B2


eight
lick
sick
bit
bull
could
oak
bowl
set
bet
miss
look
mist
wrist

C1


miss
eight
sick
set
bull
could
oak
bowl
code

C2

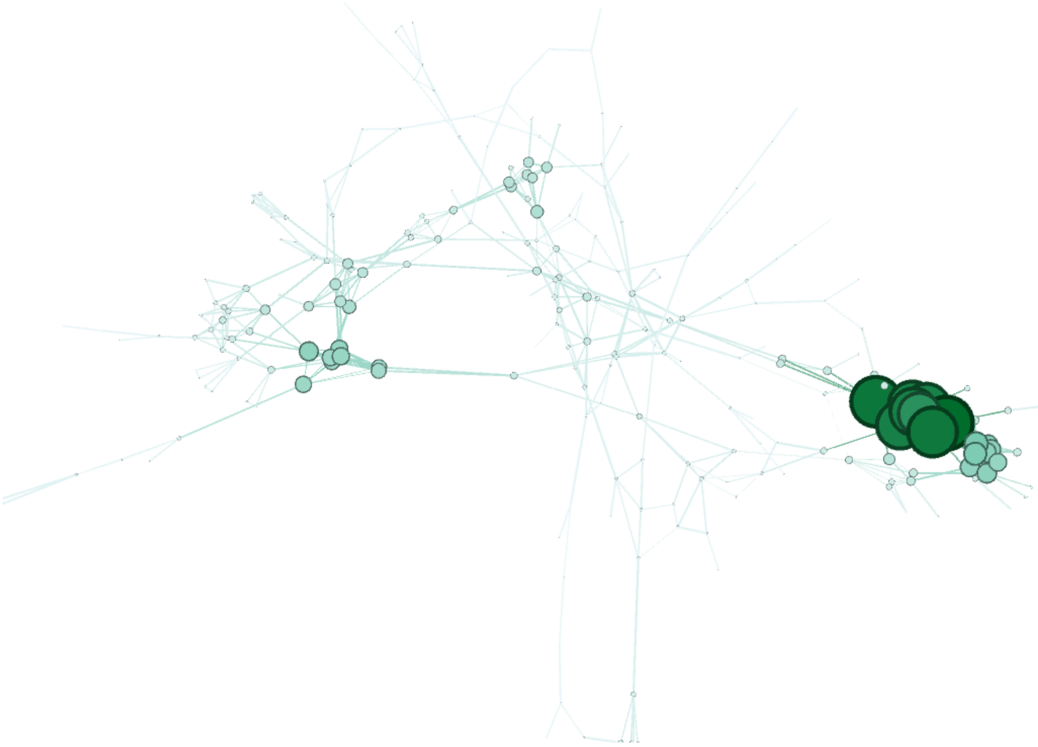

could
oak
bowl
code
pole
fit
rid
eight

BNC


eight
fear
sit
fit
rid
lit
set
bore

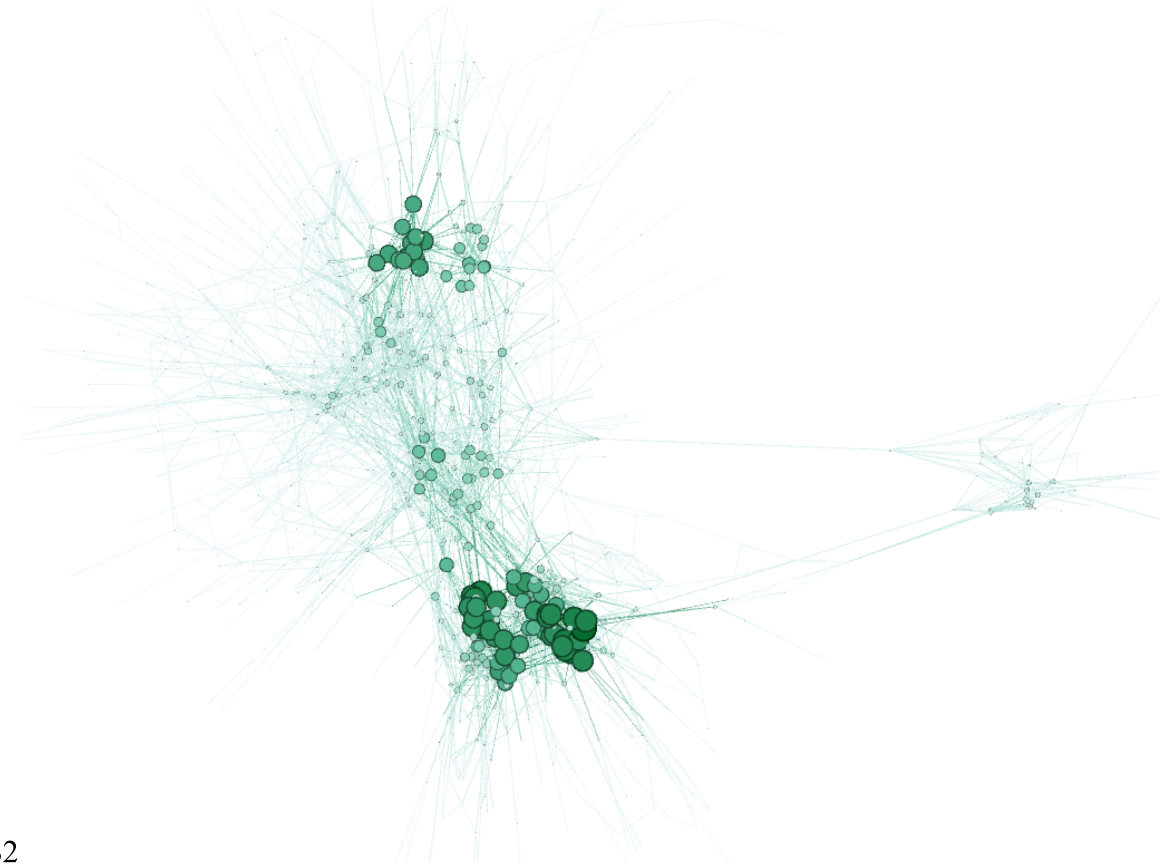
9.1.1.4. Eigenvector centrality

A1

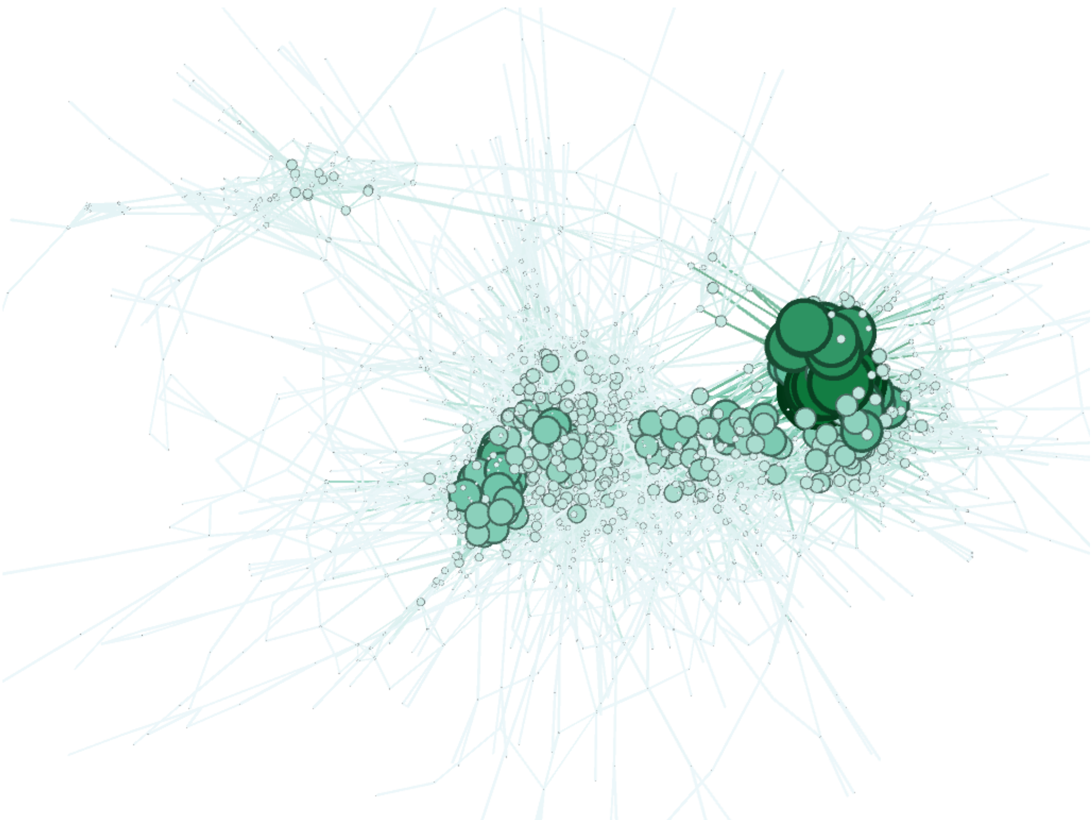


A2

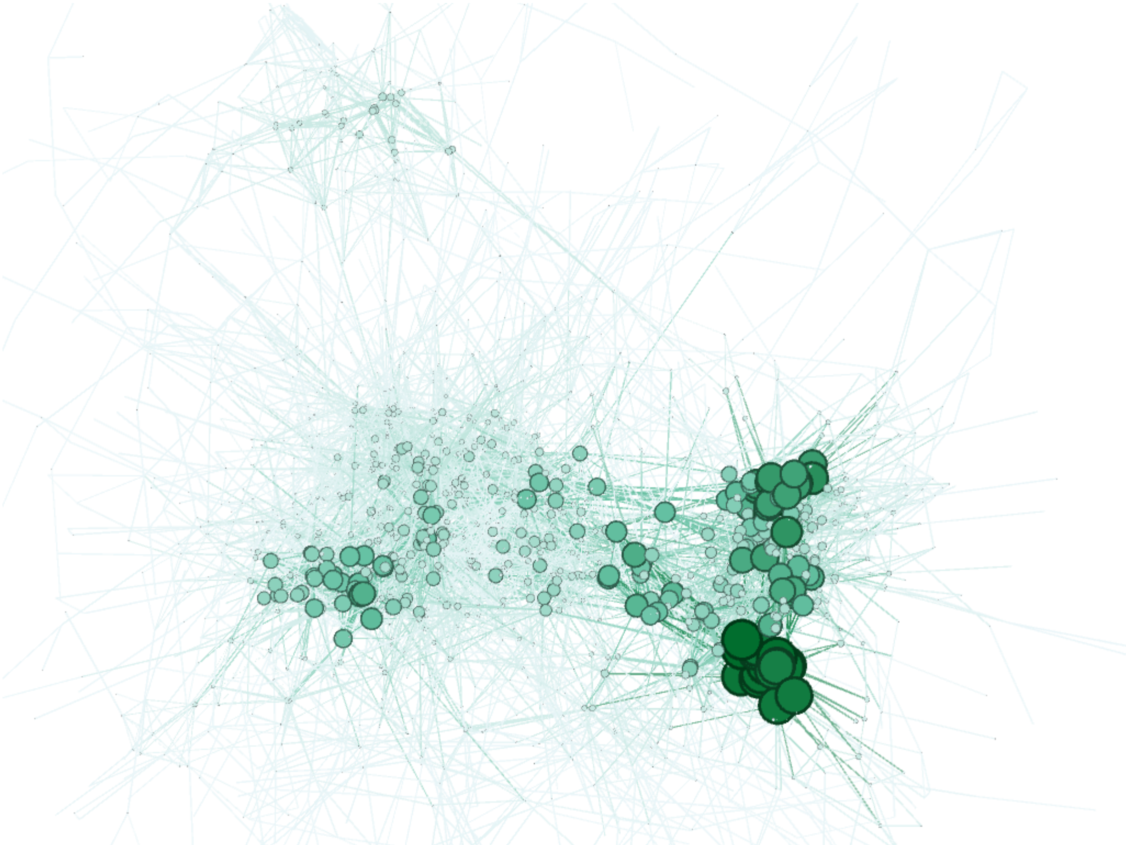
B1



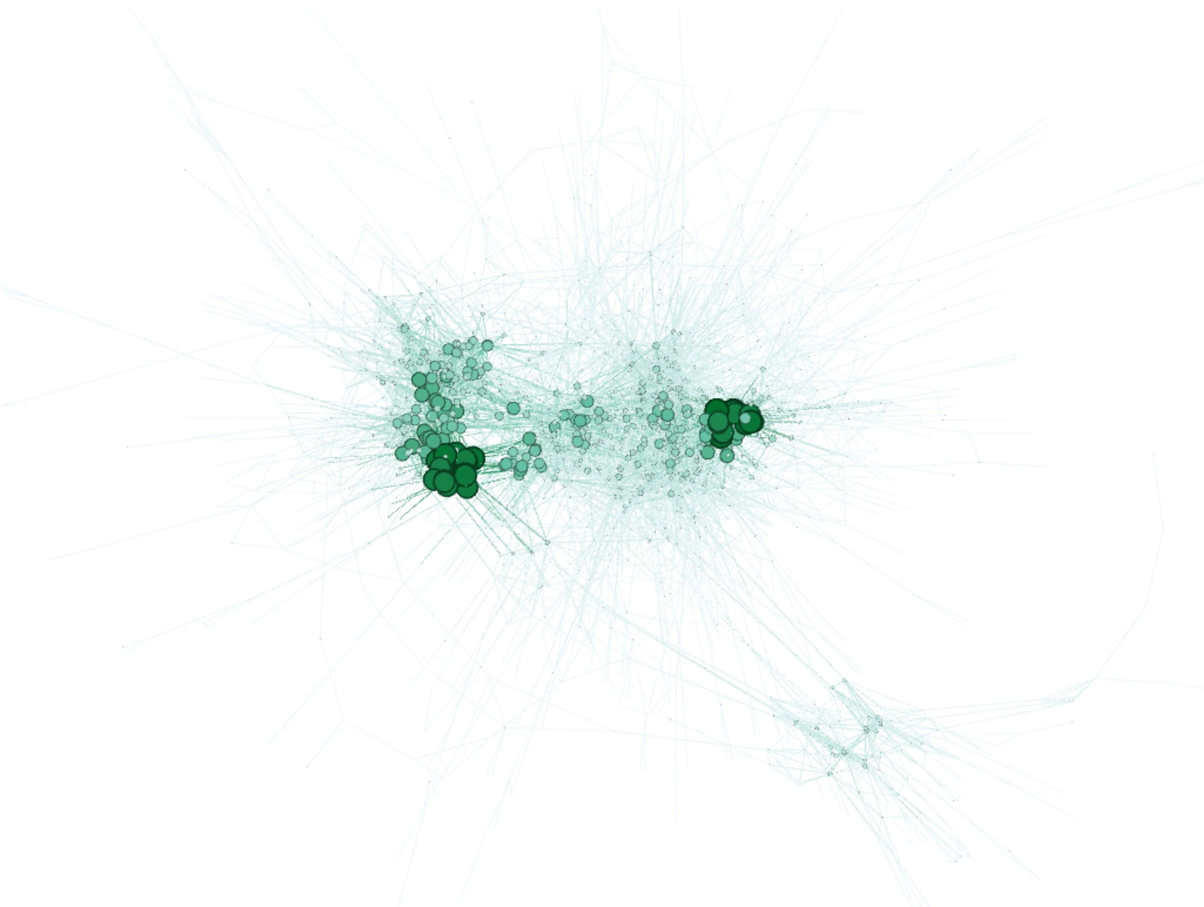
B2



C1



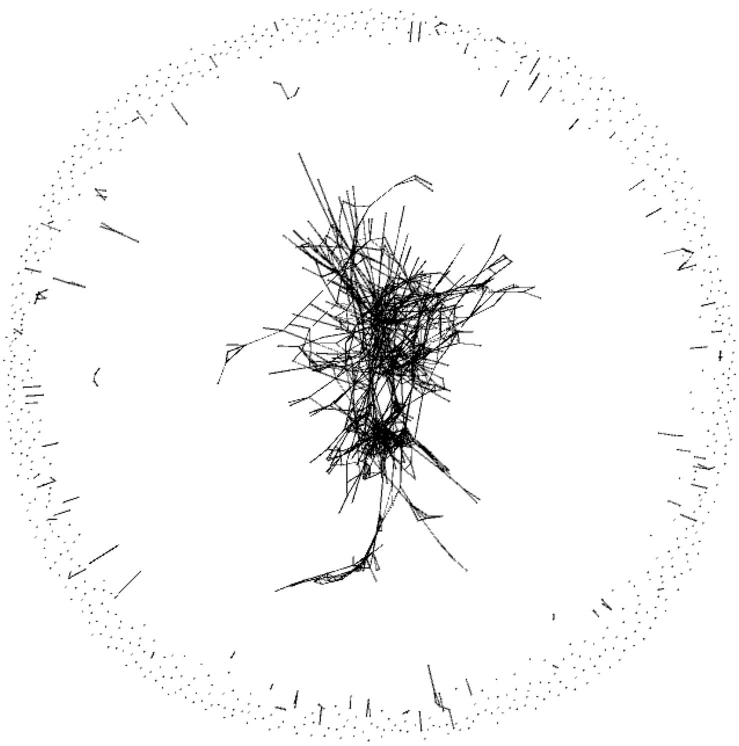
C2



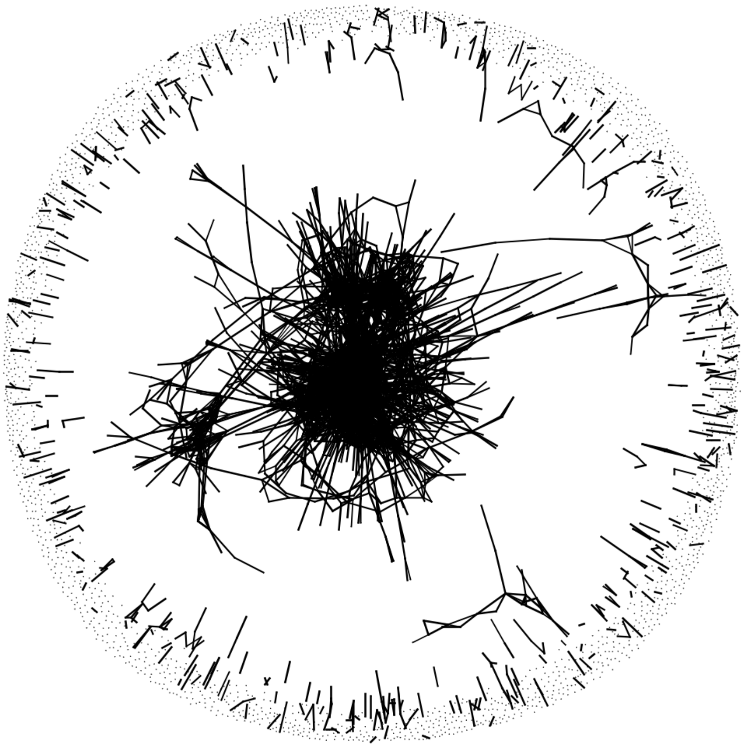
BNC

9.1.2. Macro-level graphs

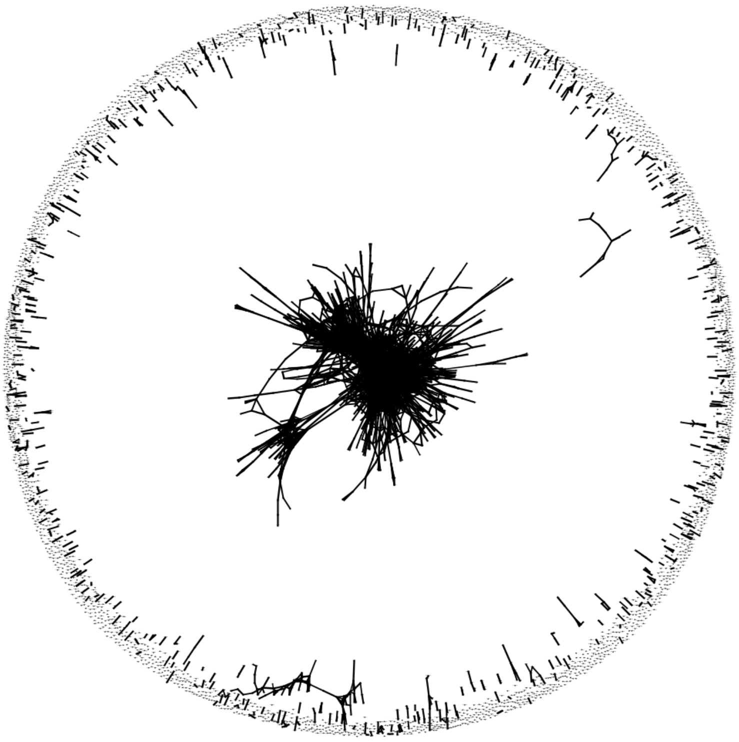
A2



B2

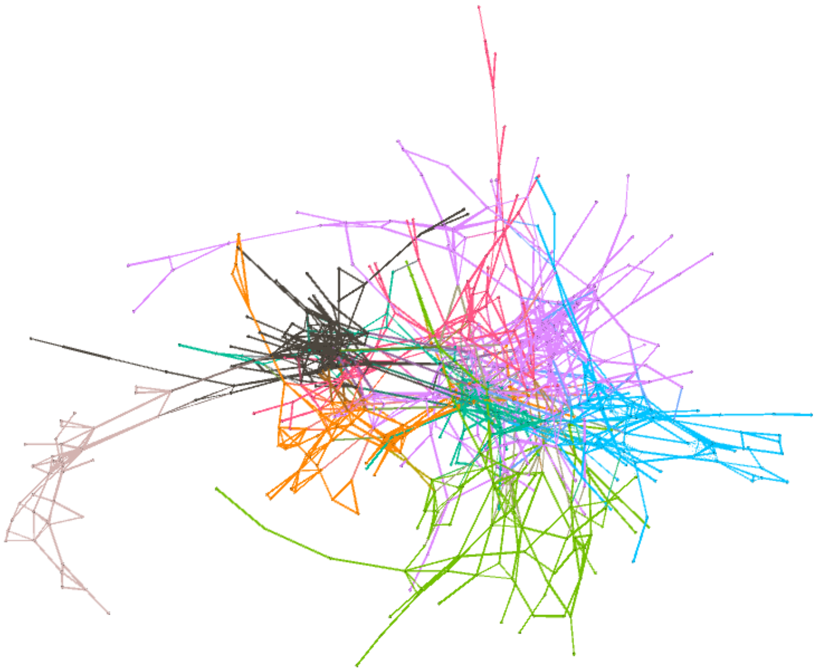


C1

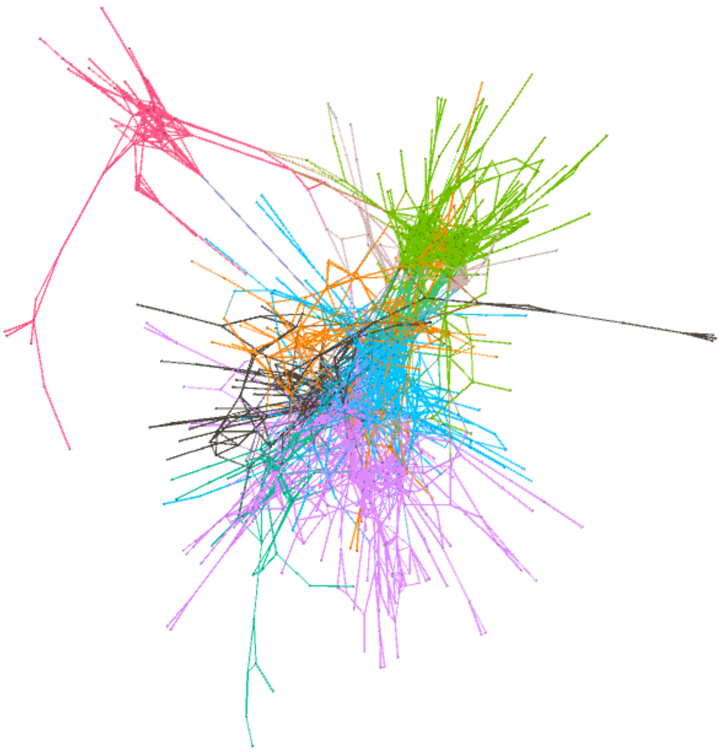


9.1.3. Giant component communities

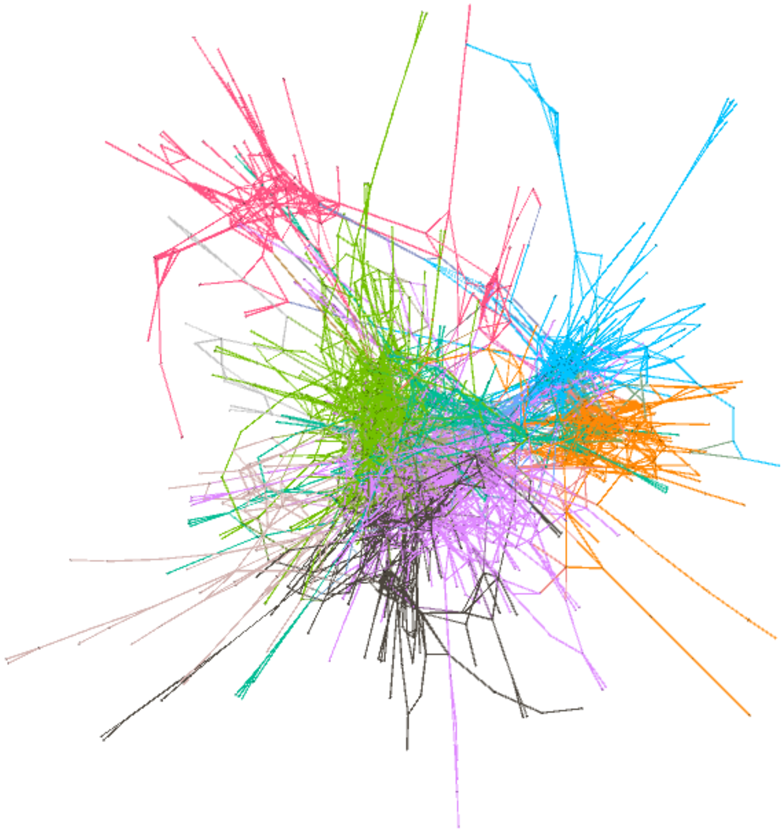
A2



B1



B2



C1