A cognitive network analysis of semantic associates in monolingual English speakers and learners of Kaqchikel

Alysia E. Martinez and Michael S. Vitevitch [ORCID]

Department of Speech Language Hearing: Sciences & Disorders, University of Kansas, Lawrence, KS, USA

Abstract

Network science was used to create and examine semantic networks of cue and response words from a word association task in learners of Kaqchikel (an endangered language indigenous to Guatemala) and monolingual English speakers. English speakers provided a wide range of responses indicative of creative language use, whereas the Kaqchikel learners provided straightforward and utilitarian responses. The observed differences in network structure in the English/ Kaqchikel learners resembled the structural differences previously observed in more/less creative individuals, suggesting an alternative account of the "foreign language effect" based on the size and structure of the semantic network in the native/foreign language.

Highlights

- Native English speakers and learners of Kaqchikel gave semantic associates to words.
- Networks were created of semantically related words for both groups.
- English networks had structures similar to networks of creative individuals.
- Kaqchikel networks had structures similar to networks of less creative individuals.
- Different semantic network structures may account for the foreign language effect.

1. Introduction

Network science is an interdisciplinary field drawing from mathematics, physics, computer science, and other related fields to represent complex systems as networks (Watts & Strogatz, 1998). *Nodes* in the network represent individual entities in the system; *edges* represent a relationship between two nodes. The complex networks used in the present study should not be confused with artificial neural network models (Rogers & McClelland, 2014) where the terms network, nodes, and connections are also used.

Complex networks have been used in various domains, including physics, biology, and the social sciences (Newman, 2008). These computational and mathematical techniques have also been used to study different aspects of cognitive and language processing (Siew et al., 2019). Words in the mental lexicon are represented as nodes, and edges represent phonological (for a review see Vitevitch, 2021), semantic (Gravino et al., 2012), or other types of relationships between words (e.g., Levy et al., 2021). Importantly, the structure of the emerging network influences the efficiency of processes in that system (Kleinberg, 2000), making it useful to analyze the overall structure of various types of cognitive networks (e.g., Marko & Riečanský, 2021).

One common way to examine the mental lexicon is to create a network of semantic associates obtained via a word association task (De Deyne et al., 2018). In a word association task, participants are given a cue word and asked to state the first word(s) that come to mind. Nodes represent the cue and response words, and (directed) edges connect the words given as responses to each cue. In many areas of the social and behavioural sciences, much research has involved participants from *w*estern, *e*ducated, *i*ndustrialized, *r*ich, and *d*emocratic (WEIRD) societies (Henrich et al., 2010). Further, much psycholinguistic research has focused on *w*ritten, *i*nstitutionally-supported, *s*tandardized, and *p*restigious (WISP) languages (Mallinson et al., 2021). To examine a broader range of languages, to assess language learning in a lesser-taught language, and to raise awareness of indigenous languages (note that 2022–2032 was declared the International Decade of Indigenous Languages by the United Nations), we, in the present study, elicited semantic associates to a set of cue words from monolingual English speakers, and from learners of Kaqchikel.

Kaqchikel is a moderately synthetic language from the K'ichean branch of the Mayan language family (Garzon et al., 1998). It is spoken by approximately 500,000 people in the central highlands of Guatemala (in the regions between Guatemala City and Lake Atitlán). The Kaqchikel language has limited syllable structure, uses many fusional affixes, has pro-drop agreement markers that replace pronouns, a highly interconnected morphology and syntax, and

CAMBRIDGE
UNIVERSITY PRESS

has a relatively free word order (Tummons, 2010). The number of Kaqchikel speakers has been declining—making it an endangered language (Instituto Nacional de Estasística, 2018 census)—in part because many families are teaching their children Spanish exclusively due to the perceived social and economic advantages of speaking Spanish in the region. However, many native Kaqchikel speakers, Guatemalan government and education sectors, and other professionals and scholars are engaging in language revitalization efforts in Guatemala (Garzon et al., 1998).

In the present study, we used the techniques of network science to construct and compare the semantic networks of monolingual English speakers and learners of Kaqchikel. In these networks, nodes represented the cue words and the responses, and directed edges connected the cue word to the responses.

2. Methods

2.1. Participants

The study reported here was approved by the Institutional Review Board at the University of Kansas. Participation in this study via Qualtrics was completely voluntary, and participants provided informed consent prior to beginning the word association task. The 90 (self-reported) native English speakers who completed the task were recruited from a General Psychology course at the University of Kansas (KU) and received course credit for their participation. Nine students in Intermediate or Advanced Kaqchikel Maya courses at the University of Kansas completed the task and received extra credit for the class for their participation. The amount of time that each student spent learning Kaqchikel ranged from roughly 5 months (i.e., a little more than a semester of coursework) to 5 years (*n* = 1, via immersive experience). Given that the English speakers and Kaqchikel learners were all students enrolled at KU in the same semester, any social/cultural/and so forth differences typically observed in two independent groups of language users were reduced.

2.2. Procedure

The Kaqchikel learners received 71 Kaqchikel words (from Moore, 2021; see Appendix A) and presented one word at a time on a computer screen as the cue words in a word association task. Participants were asked to respond with the first word that came to mind when they saw the Kaqchikel cue word. Based on vocabulary lists from the Intermediate and Advanced Kaqchikel Maya courses at KU, 57 words should have been known, and 14 words might be novel for students in those courses. Participants were encouraged to respond with Kaqchikel words that they knew and not simply translate the cue words to English. However, some responses were provided in English (indicated with "_E" on the node label) or in Spanish (indicated with "_S" on the node label).

The monolingual English speakers received the same list of 71 Kaqchikel words (translated into English). Participants were asked to respond with the first English word that came to mind when they saw the cue word.

3. Results and discussion

As many have noted (e.g., De Deyne et al., 2018), there is variability across studies using the semantic associate task in the number of responses that participants are asked to provide (e.g., the first, more

than one, as many as possible in a set amount of time, etc.), and in the criteria used to "count" a response (e.g., 2 or more participants provide the response, etc.). Given the limited number of Kaqchikel learners and the small number of cue words (relative to other studies) in the present study, we created our networks (using the Gephi software; Bastian et al., 2009) by having nodes represent the 71 cue words with directed edges connecting to nodes representing all of the (real-word) responses provided by participants (regardless of the number of participants who provided that response). Table 1 lists several commonly used measures for the two resulting networks. See Appendix B for definitions of each measure.

Many of the values reported in Table 1 are "points," not mean values of a distribution. Because there are no degrees of freedom in such measures, conventional statistics cannot be used to assess the statistical significance of observed differences. Previous studies (e.g., Watts & Strogatz, 1998; Vitevitch et al., 2023) used an informal heuristic of a value in one network being "different" if it is more than 1.5 times greater than the same measure from the other network. Using that heuristic, none of the measures differ "significantly," except nodes, edges, graph density and connected components.
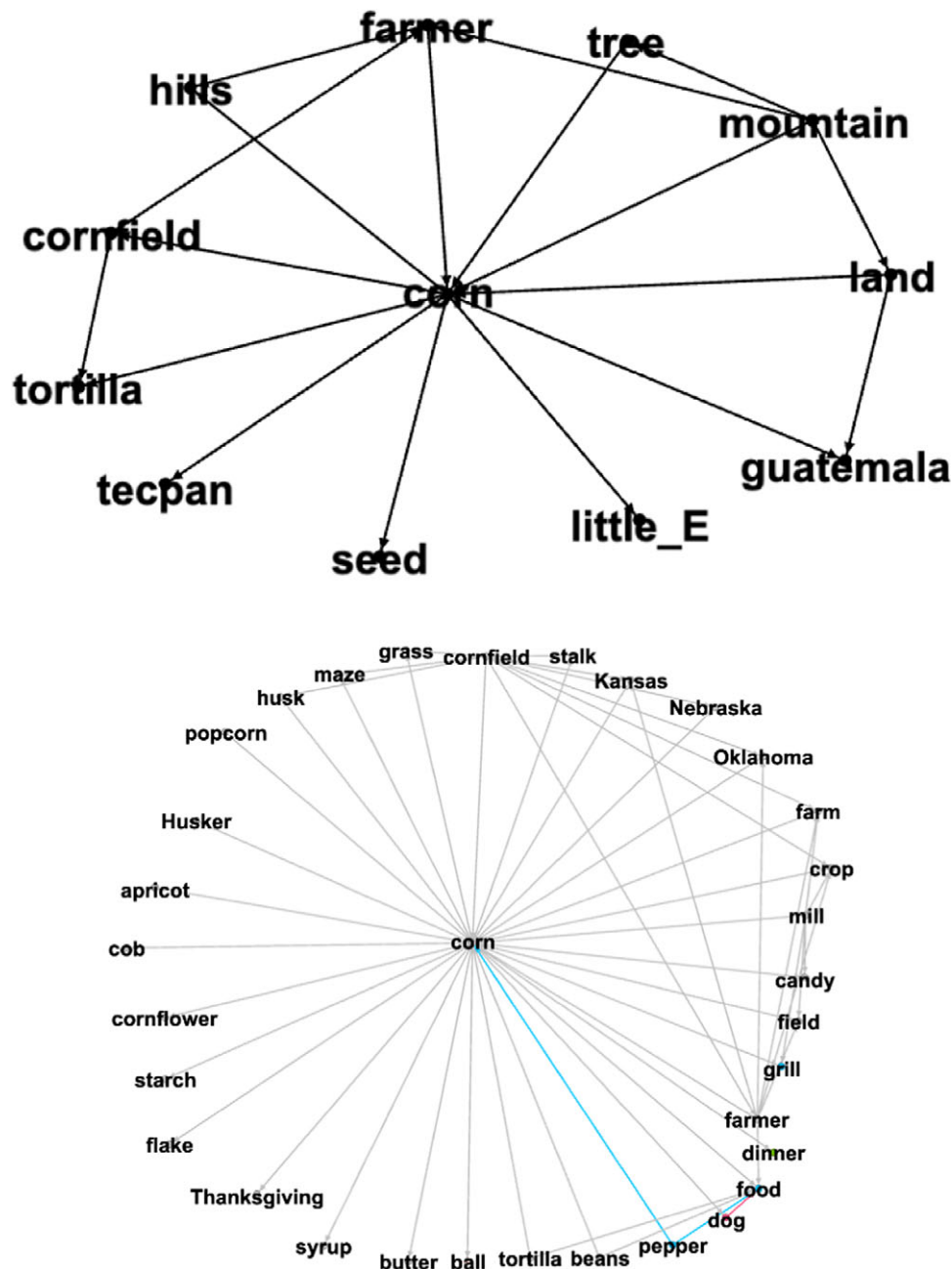
Although the monolingual English speakers and the Kaqchikel learners received the same 71 cue words, the results in Table 1 show that the network created from the responses from the English speakers had approximately 3 times as many nodes (i.e., words) and edges than the network created from the responses from the Kaqchikel learners. Similarly, when we consider the out-degree of just the 71 cue words we again see a statistically significant difference (using an independent samples *t*-test, $t(140) = 24.68, p < .001$), with the English speakers producing more responses to each cue word (*mean* = 26.69 words, *SD* = 6.20) than the Kaqchikel learners (*mean* = 7.27 words, *SD* = 2.36).

The larger number of responses from the monolingual English speakers (*n* = 90) compared to the Kaqchikel learners (*n* = 9) is of course due in part to the difference in sample size for each group. However, the difference in the number of responses from the two groups is also indicative of a difference in the *variety* of the responses provided by each group. Consider for example the responses to the cue word Ixïm (corn) for the Kaqchikel learners and the monolingual English speakers (Figure 1).

Notice (in the examples in parentheses) that both groups produced responses related to locations where corn is grown (English:

Table 1. Network measures of the Kaqchikel learners and monolingual English speakers

Network measure	Kaqchikel	English
Nodes	403	1,176
Edges	516	1,892
Average degree	1.28	1.61
Network diameter	10	11
Graph density	0.003	0.001
Connected components	3	1
Number of communities	17	18
Modularity (Q)	0.763	0.635
Avg. clustering coefficient	0.02	0.029
Avg. path length	3.477	4.63

Figure 1. The top panel shows the ego network for Ixïm (corn) from Kaqchikel learners (with a density of 0.144). The bottom panel shows the ego network for corn from monolingual English speakers (with a density of 0.054).

farm, crop, Kansas, Nebraska, field, cornfield; <u>Kaqchikel</u>: Guatemala, tecpán (Tecpán = a town in Guatemala), land, field, hills, mountain), food items related in some way to corn (<u>English</u>: tortilla, beans, syrup; <u>Kaqchikel</u>: tortilla), components of corn (<u>English</u>: cob, husk, stalk; <u>Kaqchikel</u>: seed), and occupations related to corn (<u>English</u>: Husker, farmer; <u>Kaqchikel</u>: farmer). Thus, both groups tended to focus on similar characteristics or features of the cue words.

However, the responses provided by the Kaqchikel learners appear to be more direct or utilitarian, whereas the responses provided by the English speakers appear to be more original or "creative" in nature (e.g., (corn)dog, (corn)flake, candy(corn), Thanksgiving, maze, (corn)starch, popcorn, etc.). Previous studies suggest that speakers of a foreign language make decisions about moral dilemmas that are more utilitarian in nature than native

speakers of a language (Costa et al., 2014; Planchuelo et al., 2024; *cf.*, Rühle & Lev-Ari, 2024). This so-called "foreign language effect" may have contributed to the difference we observed in how utilitarian (or creative) the responses were from Kaqchikel learners (or English speakers).

The observation that both groups focused on similar characteristics or features of the cue words casts some doubt on an explanation based completely on the foreign language effect. Further doubt is cast on the foreign language effect by findings of a diminished foreign language effect in bilinguals who are proficient in both languages (Degner et al., 2011; Dylman & Champoux-Larsson, 2020), suggesting that proficiency in L2 may influence the extent to which a response is utilitarian/creative in the present study. Recall that the participants in the present study were (self-reported)

native English speakers with 18+ years of exposure to the language and learners of Kaqchikel enrolled at the same Midwestern university with approximately 5 months of exposure to the language. The more proficient English speakers had a richer, more nuanced semantic knowledge to draw upon for each cue word than the Kaqchikel learners who knew a limited number of words in that language. This enabled the English speakers to produce unique and creative responses, whereas the limited vocabulary of the Kaqchikel learners restricted them to common, rigid, utilitarian responses.

The hypothesis that the size or structure of the semantic network in novice/proficient language users might influence the utilitarian/creative nature of their responses is consistent with other work that examined the semantic networks of people who were high or low in creativity (Kenett et al., 2014). Kenett et al. found that less creative individuals had semantic networks that were rigid, and had few connections between nodes. The semantic networks of more creative individuals had broader connections (connecting seemingly distant concepts) and were more spread out. Consistent with the idea that novice language users might produce less creative/utilitarian responses and proficient language users might produce more creative responses is the density of the Kaqchikel and English networks. As shown in Table 1 the Kaqchikel network (density = 0.003) was more rigidly connected than the English network (density = 0.001) where seeming disparate concepts might be connected. A difference in network density can also be seen in the ego-networks of the cue word Ixïm/corn shown in Figure 1 (with the density values reported in the caption).

Additional support for the hypothesis that the structure of the semantic network of novice/proficient language users might influence the utilitarian/creative nature of their responses can be seen in the number of components in each network. The English network had only 1 component, meaning that every node was connected in some way to every other node in the network. The implication of such a structure is that even disparate concepts can be reached and might appear to be related in some way.

In contrast, the network of the Kaqchikel learners had 3 components: a giant component (containing 409 of the 421 nodes that were connected to each other), and 2 smaller components that were not connected to the other components. (One of the small components contained the cue word taluwäch/television and 6 responses, and the other small component contained the cue word ti'ij /meat and 5 responses.) The implication in this case is that there is a literal "gap" in semantic knowledge (i.e., 3 components that are disconnected from each other), which makes it impossible to get from some concepts to others, and makes it increasingly difficult to use words in a flexible, novel, or original way to produce a "creative" response. The observation that learners of a language might have a semantic network with several disconnected components also suggests that language instructors might target vocabulary words that bridge together the components that are disconnected from each other to aid acquisition by better approximating the semantic structure of proficient/native speakers.

Consider further the responses from the proficient/native English speakers (in the bottom panel of Figure 1). Notice that several of the responses are compound words with complex compositional meanings (e.g., corndog, cornflake, candy corn, cornstarch, popcorn). The Common European Framework of Reference (CEFR) states that having "…a good familiarity with idiomatic expressions and colloquialisms…" is indicative of CEFR-level C2 "proficiency" (from the Spoken Interaction criterion) in a foreign language. Perhaps a related marker of proficiency in a foreign language might relate to the facility that speakers have in not only understanding compound words with complex compositional meanings but in creating novel, humorous concepts, such as *cabbage cage*, by pairing existing concepts (Siew et al., 2022). Further investigation of such responses from the English speakers in the present study may be useful in determining how a mechanism known as "generators" develops in the lexical network of a native speaker or an L2 speaker to model the creation of such novel concepts or relationships between existing concepts (Hills & Kenett, 2022).

More broadly, further investigation of the responses from the English speakers in the present study might contribute to questions about the way that morphology is represented and processed in the mental lexicon (Embick et al., 2022). The investigation of Kaqchikel learners in the present study also highlights the importance of considering less taught/non-WISP languages with morphological systems that differ significantly from the English language when addressing questions about morphological processing, or when attempting to use networks to model the lexicon (Vitevitch et al., 2024).

The present study demonstrates a novel way to use a semantic association task and the mathematical techniques from cognitive network science to study the mental lexicon of native speakers and language learners. For other examples of network science being used to investigate bilingualism see Borodkin et al. (2016)), Tiv et al. (2020)), and Luef (2022).

Perhaps unsurprising, the results of the present analysis show that the semantic network of language learners is smaller than the semantic network of native speakers of a language. Crucially, however, the present analysis suggests that the more "loosely" connected structure of the semantic network of native speakers of a language may facilitate the production of linguistically creative responses (i.e., associating seemingly disparate concepts). In contrast, the rigid and disconnected structure of the semantic network of (novice) language learners may constrain their responses to be more common, straightforward, and utilitarian in nature.

Although the utilitarian decisions made by people in L2 compared to their native language (known as the "foreign language effect") are often attributed to the reduced emotional response elicited by the L2 (Costa et al., 2014), the present observation suggests an alternative interpretation. Namely, the size and structure of one's semantic network may allow an experienced speaker to creatively connect disparate concepts (see Kenett et al., 2014), or constrain a less experienced speaker (with a smaller and disconnected network) to produce responses that are more utilitarian and less creative in nature.

The present analysis also draws attention to less taught/non-WISP languages, and illustrates how studying such languages can lead to insights that might help researchers better understand how all languages (even L1) are learned (e.g., Luef, 2022). For example, language instructors might use the tools of network science to map out the semantic or phonological knowledge of their students and target new vocabulary words that would bridge the "gaps" between disconnected clusters of related words. Finally, the present analysis provides another way in which the tools of network science can be used to understand another aspect of bilingualism (see also Borodkin et al., 2016 and Tiv et al., 2020).

Competing interest. The author(s) declare none.

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*.

Borodkin, K., Kenett, Y. N., Faust, M., & Mashal, N. (2016). When pumpkin is closer to onion than to squash: The structure of the second language lexicon. *Cognition*, **156**, 60–70.

Costa, A., Foucart, A., Hayakawa, S., Aparici, M., Apesteguia, J., Heafner, J., & Keysar, B. (2014). Your morals depend on language. *PLoS One*, **9**(4), e94842. https://doi.org/10.1371/journal.pone.0094842

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1115-7

Degner, J., Doycheva, C., & Wentura, D. (2011). It matters how much you talk: On the automaticity of affective connotations of first and second language words. *Bilingualism: Language and Cognition*, **15**, 181–189.

Dylman, A. S., & Champoux-Larsson, M.-F. (2020). It's (not) all Greek to me: Boundaries of the foreign language effect. *Cognition*, **196**, Article 104148. https://doi.org/10.1016/j.cognition.2019.104148

Embick, D., Creemers, A., & Goodwin Davies, A. J. (2022). Morphology and the mental lexicon: Three questions about decomposition. In A. Papfragou, J. C. Trueswell, & L. R. Gleitman (Eds.), *The Oxford handbook of the mental lexicon* (pp. 77–97). Oxford: Oxford University Press.

Garzon, S. R., Brown, M. K., & Becker Richards, J. (1998). *The life of our language: Kaqchikel Maya maintenance, shift, and revitalization*. University of Texas Press.

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), 1250054-1–1250054-22.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, **33**(2–3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Hills, T. T., & Kenett, Y. N. (2022). Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. *Topics in Cognitive Science*, **14**, 189–208.

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience*, **8**. https://doi.org/10.3389/fnhum.2014.00407

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

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**(1), 1–14.

Luef, E. M. (2022). Growth algorithms in the phonological networks of second language learners: A replication of Siew and Vitevitch (2020a). *Journal of Experimental Psychology: General*, **151**(12), e26–e44. https://doi.org/10.1037/xge0001248

Mallinson, C., Cahill, M., Francis, E. J., Namboodiripad, S., Labrada, J. R., & Bender, E. M. (2021). Teaching ethics across the linguistics curriculum. In *Symposium at the Linguistics Society of America Virtual Annual Meeting (Virtual)*.

Marko, M., & Riečanský, I. (2021). The structure of semantic representation shapes controlled semantic retrieval. *Memory*, **29**(4), 538–546. https://doi.org/10.1080/09658211.2021.1906905

Moore, R. J. (2021). *Survey and sociolinguistic variation analysis of Kaqchikel word Associations*. Tulane University.

Newman, M. (2008). The physics of Networks. *Physics Today*, **61**(11), 33–38. https://doi.org/10.1063/1.3027989

Planchuelo, C., Hinojosa, J. A., & Duñabeitia, J. A. (2024). The nature of lexical associations in a foreign language: valence, arousal and concreteness. *Bilingualism: Language and Cognition*, 1–10. https://doi.org/10.1017/S1366728924000117

Rogers, T. T., & McClelland, J. L. (2014). Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, **38**, 1024–1077.

Rühle, M.-C., & Lev-Ari, S. (2024). Do native and non-native speakers make different judicial decisions? *Bilingualism: Language and Cognition*, 1–8. https://doi.org/10.1017/S136672892400018X

Siew, C. S. Q., Engelthaler, T., & Hills, T. T. (2022). Nymph piss and gravy orgies: Local and global contrast effects in relational humor. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **48**(7), 1047–1063. https://doi.org/10.1037/xlm0001120

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**(1), 2108423.

Tiv, M., Gullifer, J., Feng, R., & Titone, D. (2020). Using network science to map what Montréal bilinguals talk about across languages and communicative contexts. *Journal of Neurolinguistics*, **56**, 100913. https://doi.org/10.1016/j.jneuroling.2020.100913

Tummons, E. J. (2010). *Positional roots in Kaqchikel Maya*. University of Kansas. http://hdl.handle.net/1808/7426

Vitevitch, M. S. (2021). What can network science tell us about phonology and language processing? *Topics in Cognitive Science*, **14**, 127–142. https://doi.org/10.1111/tops.12532

Vitevitch, M. S., Castro, N., Mullin, G. J. D., & Kulphongpatana, Z. (2023). The resilience of the phonological network may have implications for developmental and acquired disorders. *Brain Science*, **13**, 188. https://doi.org/10.3390/brainsci13020188

Vitevitch, M. S., Martinez, A. E., & England, R. (2024). Defining nodes and edges in other languages in cognitive network science—moving beyond single-layer networks. *Information*, **15**(7), 401. https://doi.org/10.3390/info15070401

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

Appendix A

Table A1. The 71 cue words used in the semantic associate task

Kaqchikel	English
räx	green
wuj	book
xajab'	sandal
ruk'u'x tinamït	town center
sib'	smoke
kinäq'	beans
tzuyül	sitting
ulew	land
k'exelom	midwife
jub'ül	smell
wexaj	pants
ko'öl	small
k'üch'	vulture
uq	skirt
nïm	big
awän	cornfield
k'olon	curled up
ajkem	weaver
ajq'ij	spirit guide
ak'wal	child
aq'omab'äl jay	hospital
b'ey	road
b'ojo'y	clay pot
chakäch	basket
che'	tree
chi'aj	mouth
ch'atal	table
ik	pepper
ixim	corn
ixöq	woman
itzel	bad
jeb'äl	beautiful
juxb'äl	marker
juyu'	mountain
kumätz	snake
k'ayb'äl	market
mama'aj	grandfather
muxu'x	belly button
nimajay	living room
oj	avocado
pa'äl	standing

(*Continued*)

Table A1. (*Continued*)

Kaqchikel	English
paläj	face
pamaj	stomach
pawi'aj	hat
po't	blouse
q'aq'	fire
raqän ya'	river
saqmolo'	egg
taluwäch	television
ti'ij	meat
tijob'äl	school
tijonel	teacher
tikonel	farmer
tukr	owl
tz'i'	dog
umül	rabbit
ütz	good
wäy	tortilla
winäq	person
xikinaj	ear
jewël	seated
jut	worm
mesonel	housekeeper
pik'ïl	on tippy toes
q'inob'äl	mill
raq'äl	leaning
setël	circular
tunül	stacked
tzalan	lie on your side
tz'eqël	hanging
xukül	on one's knees

Appendix B

Definitions of common network measures. For formal definitions and information on how to calculate these measures, see Bastian, Heymann, and Jacomy (2009).

Nodes. Nodes represent individual entities in the system being studied, such as people who are friends with each other in a social network. In the present study, nodes represented words (i.e., cue words and responses).

Edges. Edges are connections placed between nodes that are related in some way. In a social network, edges would connect nodes/people who are friends with each other. The absence of an edge between two nodes/people would indicate that those two people do not know each other. Edges can be directed (as in the present study) indicating that the cue word/node elicited various response words. Edges can also be undirected indicating a bidirectional flow of information/goods/knowledge etc. In some networks, edges can be weighted to indicate the strength of the relationship between two nodes. For example, A and B might be friends since elementary school (and have a larger weight on the edge), whereas A and C met only recently (and would have a smaller weight on

the edge). The edges in the present study were unweighted, meaning the strength of the relationships among nodes was assumed to be similar.

Average Degree. The average degree refers to the mean number of edges per node in the whole network.

Network Diameter. The network diameter refers to the shortest number of connections that must be traversed to connect the two nodes that are farthest from each other in the network.

Graph Density. Graph density is the ratio between the actual number of edges and the number of possible edges in the network. Values closer to 0 indicate that there is a small number of connections (compared to the number that is possible); such networks are said to be sparsely connected. Values closer to 1 indicate that there is a large number of connections (compared to the number that is possible); such networks are said to be densely connected. A value of 1 indicates that the network is fully connected (i.e, every node is connected to every other node in the network).

Connected Components. Connected component refer to nodes that are connected in some way to each other. The giant component is the component with the largest number of nodes connected to each other in some way. In the context of language networks, smaller components are sometimes referred to as "lexical islands" (Vitevitch, 2021).

Number of communities. A community is a group of nodes that are more connected to each other than they are to nodes in other communities. Communities are not the same thing as a connected component. Within a connected component, there may still be subgroups of (connected) nodes that tend to be more connected to each other than to other nodes in the component, thereby forming a community.

Modularity (Q). Modularity is a measure of how strong the division between components is in a network. Values closer to 0 indicate that the community structure is weak, whereas values closer to 1 indicate that the community structure is better defined.

Average Clustering Coefficient. The average clustering coefficient measures the extent to which nodes that are neighbors to a given nodes are also neighbors of each other. Values close to 0 indicate that the neighboring nodes tend to just be connected to a given node, whereas values close to 1 indicate that the neighboring nodes connect to the given node and have a high probability of being connected to each other as well.

Average Path Length. Path length refers to the shortest number of steps it takes to get from one node to another node. The average path length is the mean value for all possible pairings of nodes in the network.