The Influence of Semantic Network Structure of English on Word Processing

Student a věda – lingvistika 2024

Tomáš Savčenko

Anglický jazyk – Obecná lingvistika

3. ročník navazujícího magisterského studia

ÚAJD

1. Introduction

There has been a long tradition of implicitly conceptualizing language as a network, with linguistic theories from structuralism to construction grammar focusing on links between dyadic entities. Network theory has been applied to various research field and increasingly often to the study of language (Barabási & Pósfai, 2016). While still a relatively young method in linguistics, it is a promising versatile tool whose spread to other research fields can facilitate interdisciplinarity and follow on the tradition of implicit "networkness" of linguistic theories. Network science's universal principles allow for generalizations to other disciplines such as psychology or computer science to study the cognitive underpinnings of language. The recent development of chatbots with convincing language skills raises questions of if, why, and how computers can understand language (Gastaldi, 2021) and if it can tell us something about how language works in the human mind (Vitevitch, 2008). The present study aims to contribute to this discussion by modelling word meaning in English with network science and recent machine learning techniques to examine word processing in the mental lexicon.

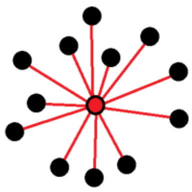
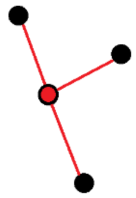
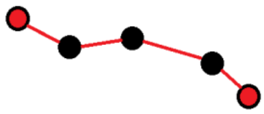
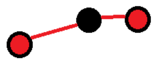
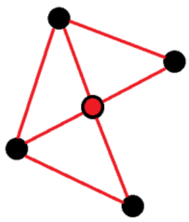
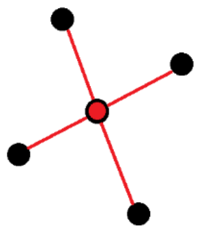
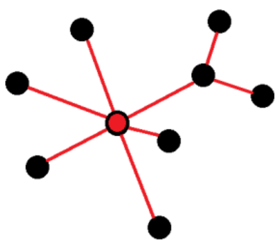
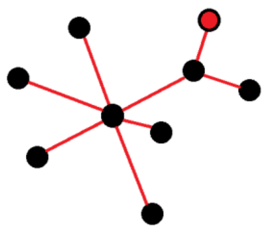
2. Theoretical Background

2.1. Network Science

Network science applies mathematical principles to explore the structure of complex systems, including the internet, power grids, or language, by studying nodes and their links (Barabási & Pósfai, 2016). The language network science formalizes units of language as nodes and the relationships between them as links, but it moves beyond traditional structuralist frameworks like Saussure’s, which also implicitly imagined language as a network. It expands the idea by applying rigorous formalism used across different research fields and by focusing on the statistical properties of such networks. The use of network science tools allows linguists to conduct quantitative empirical analyses of language agnostic to the assumptions of specific linguistic theories. It can also be used to test hypothesis about language within a broader perspective of cognitive science (Vitevitch, 2008, 2019).

Key network science measures include degree centrality, shortest path, clustering coefficient, and closeness centrality, quantifying properties of individual nodes and broader network structures. They are illustrated in table 1.

Table 1 Summary of relevant basic network measures

Measurement	Definition	Comparison between Higher and Lower Values	
Degree centrality	The number of links of a node relative to the overall network size.		
Shortest path length	The shortest route between two nodes which equals the number of links it contains.		
Clustering coefficient	A measure of local clustering which signals how often the neighbours of a node tend to be neighbours of each other.		
Closeness centrality	A measure signalling how close a node is to all other nodes.		

These measures help identify relevant network structures. For example, hubs which are prominent nodes that are characterised by high degree and closeness centrality. Local clustering within the network, being a feature of the mesoscopic level, can be described by clustering coefficient of specific nodes. We can compute average values of these measures across a whole network which shifts the view to the macroscopic level of analysis. Average values describe the overall features of a network. Many cognitive and social phenomena modelled with networks show small-world structure (SWS) (Watts & Strogatz, 1998). In general, SWS is characterized by short average shortest path length and high local clustering. Figure 1 contrasts a network with SWS and so-called scale-free network which has less local clustering and longer average shortest path length. Finally, the information contained in the network can be enriched by weighted links, which add information about the strength of connections between nodes.

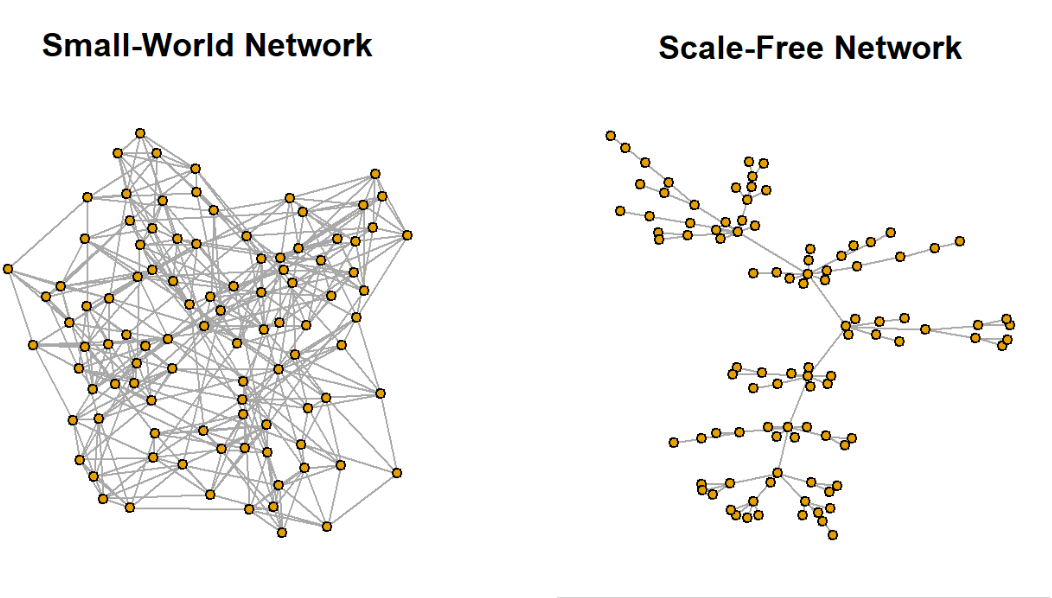


Figure 1 A comparison between small-world network and scale-free network

2.2. Psycholinguistics and Word Processing

Combining network science with psycholinguistics provides a robust framework for exploring the structure and properties of mental lexicon, a suggested network-like cognitive repository crucial for word processing. Word processing involves identifying and retrieving language units to produce or understand speech. Despite the ongoing discussion about the nature of language units of the mental lexicon, e.g., concepts, morphemes, words, phrases, syntactic rules, etc. (Aitchison, 2012), this work conceptualizes the basic units of the mental lexicon as words in accord with the previous practice of language network science (Vitevitch, 2019).

Some of the experimental designs in psycholinguistics to study word processing include word identification task, lexical decision task, and naming task. These tasks, adaptable to auditory and visual modalities, examine how different variables, like word frequency, imageability, length, etc., affect word processing, reflected by reaction times and accuracy rates of participant's responses. For instance, word identification tasks might assess the impact of noise on auditory word recognition, while lexical decision tasks involve participants distinguishing real words from similar pseudo words. Quick and correct answer is interpreted as efficient word processing.

Integrating network science with psycholinguistic studies, researchers can leverage network measures and analyse them as independent variables to investigate their influence on the dependent variables, such as reaction time that reflect word processing. This approach has yielded useful insights, particularly within the study of phonological networks (Vitevitch, 2019). The following chapter introduces the previous research combining network science with psycholinguistics conducted on phonological networks.

2.3. Phonological Networks

The use of network science in linguistics is still relatively young but notable body of research has been done on phonological networks, where nodes represent words linked by phonological similarity. In the work of Michael S. Vitevitch, this relationship is defined by 'substituting, adding, or deleting a single phoneme in a given word to form a "phonological neighbor". For example, the words hat, cut, cap, scat, and _at were considered phonologically similar to the word cat' (Vitevitch, 2008: 3). Such networks, exhibit SWS, indicated by high clustering coefficient, a few words functioning as hubs, and short average shortest path length (Vitevitch, 2008). An example of a phonological network with SWS is on figure 2 below.

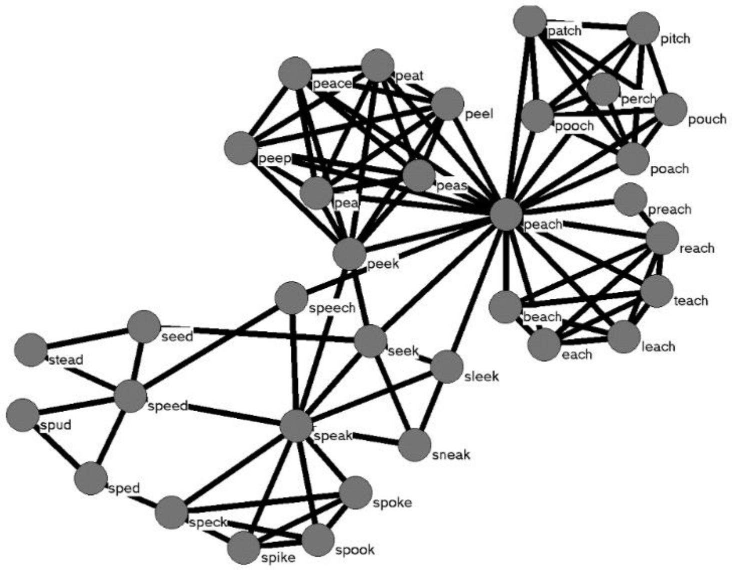


Figure 2 Example phonological network from Vitevitch et al. (2023)

It contains the phonological neighbourhood of words *peach* and *speak.* There are at least three discernible clusters around the hub node *peach*.

Vitevitch and colleagues have investigated the influence of network measures like the clustering coefficient on word retrieval. In a 2009 study, Chan and Vitevitch conducted two lexical retrieval experiments. The first experiment used a perceptual identification task where participants identified noise-distorted words, grouped by high and low clustering coefficient, with accuracy rates significantly higher in the high clustering coefficient group. The second experiment, a lexical decision task, required participants to distinguish real words from non-words, showing slower response times for words with high clustering coefficient, although the effect size was small.

Subsequent research, including a 2011 study by Vitevitch et al., used simulations to show that lower clustering coefficient led to greater activation in a phonological network, suggesting faster and more efficient word retrieval. This appears to contradict the slower response times for high-clustering words found in the earlier study, suggesting a complex role of clustering coefficient in word processing. These studies demonstrate that clustering coefficient and other network measures play role in word retrieval and processing.

This section laid the groundwork for subsequent discussions of semantic networks, the main focus of this study. Ultimately, the aim is to probe if the influential network measures identified in phonological networks hold similar significance in semantic networks during word processing. At the same time, this study aims to take advantage of the recent advancement of machine learning and large language models (LLMs), which can represent and model language and meaning in unprecedented ways. The following section provides an overview of different kinds of semantic networks.

2.4. Semantic Networks

Semantic relationships conceptualized as links between word nodes in a network have been constructed in different ways in language network science. The main step in constructing any kind of network is acquiring a list of all the pairs of nodes that are linked in the network. Researchers have used different sources of data and methods to produce these lists.

Some semantic networks are built on perceptual or functional word features, where links denote semantic proximity based on shared attributes (e.g., “is large”, “has fur”) or roles (e.g., “used in cooking”, “is a vehicle”). These networks are structured according to semantic feature norms established by researchers such as McRae et al. (2005), Vinson & Vigliocco (2008), and more recently Preininger, Brand & Kříž (2022). These norms are generated through experiments where participants list or evaluate features along certain semantic dimensions for specific words. For instance, in the work by Preininger, Brand, and Kříž (2022) on Czech, participants rated words like “pes” (dog) or “vousy” (beard) on Lickert scale assessing dimensions such as urbanity, femininity, positivity, etc.

Another semantic network approach uses links based on the norms from free association tasks, reflecting the spontaneous connections between words experimentally elicited from participants. The creation of free association norms involves tasks with participants reacting to a cue word with their first instinctive word, a technique demonstrating reliable outcomes despite its spontaneous nature (Nelson, McEvoy, & Schreiber, 2004). The norms reflect the associative strength and context-free priming within the mental lexicon.

Semantic-conceptual networks are based on structured lexical databases like WordNet or Roget’s Thesaurus which categorize words by conceptual similarities, synonymy, antonymy, and hierarchical super-subordinate categorizations, etc. These networks can be better interpreted than those based on word co-occurrences, despite being constrained by the theoretical and language-specific nature of the underlying databases. Figure 3 is an example of a visualization of such network. [1]

[1] An example of an online tool visualizing relationships from Wordnet as a network: https://visuwords.com/

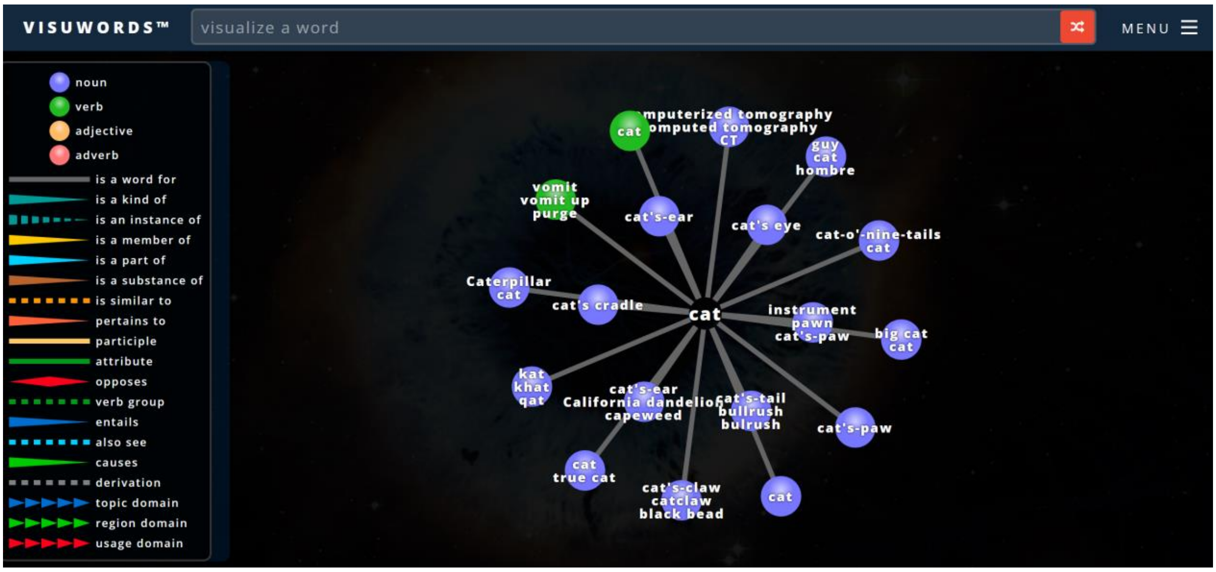


Figure 3 An example of a visualization of Wordnet's network for the word 'cat' made by Visuwords educational project

Networks based on natural language corpora represent semantic relationships through co-occurrence and usage pattern measures in large text corpora. This data-driven approach, using semantic space models and word co-occurrence measures, captures implicit semantic connections inferred from linguistic context. Semantic space models, analysing extensive text corpora, use intricate algorithms to convert words into vectors within a high-dimensional space, allowing semantic similarities to be deduced from the vectors' spatial relationships (Lund & Burgess, 1996; Gastaldi, 2021). Cosine similarity is used to quantify the semantic distances between vectors. Word co-occurrence measures and semantic space models differ in their approach to modelling word meaning and their computational methods. Word co-occurrence relies on the frequency of words appearing together within a specific context, typically represented in matrices with frequencies or statistics like pointwise mutual information (PMI) (Church & Hanks, 1990). However, these matrices, constrained by their context windows, often overlook broader semantic relationships. Vector spaces, in contrast, abstract from narrow co-occurrence data to capture semantic similarities even among words with potentially broader usage patterns. This principle has informed the development of large language models (LLMs) such as Word2Vec and Bidirectional Encoder Representations from Transformers (BERT), primarily for commercial applications, with limited use in linguistic research (Gastaldi, 2021). This paper aims to construct a semantic network of English using BERT, examining the impact of network measures, such as the clustering coefficient, on word processing.

Before reviewing the methods and results of constructing my semantic network and correlating its values with the results of word retrieval experiments, the following section will briefly introduce and demonstrate the basics of word vectors, the main mechanism behind LLMs and the way in which they can represent meaning.

2.5. Large Language Models and Word Vectors

The introduction of machine learning techniques like Word2Vec by Mikolov et al. (2013a, b, c) and later BERT by Devlin et al. (2019) has significantly advanced the field of natural language processing (NLP) by transforming words into high-dimensional vectors based on their contextual usage within large text corpora. These techniques, grounded in the distributional hypothesis (Firth, 1957; Harris, 1954), posit that words in similar contexts have similar meanings, hence their relative distance in vector space. The following figures aim to demonstrate the different kinds of semantic relationships that word vectors can capture.

Multidimensional word vectors are arrays of numbers assigned to words. A common way to visualize them is by applying dimension reducing statistical methods such as t-distributed stochastic neighbour embedding (t-SNE) or principal component analysis (PCA) and plotting them in 2D space[2]. Figure 4 shows such a visualization which is able to cluster together similar categories of words such as numerals or inflected forms of the same word. In a 2D visualization of word vectors, the x and y axes are abstract dimensions that represent the compressed information from the original high-dimensional space, with the aim of preserving the relative distances between words to reflect their similarities in meaning or context. The axes themselves do not have a specific meaning beyond this spatial representation.

[2] An on-line tool for visualizing 2D and 3D word vectors can be found at https://projector.tensorflow.org.

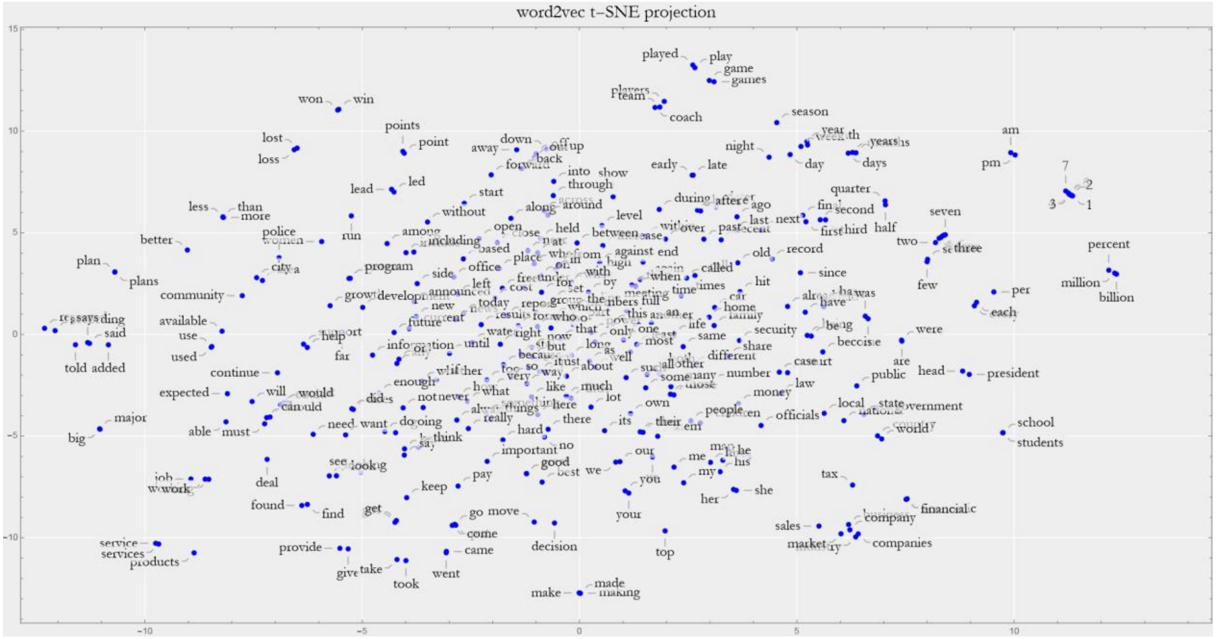


Figure 4 T-SNE two-dimensional projection of the word2vec vectors representing a selection of the most frequent words in its training corpus adopted from Gastaldi (2021)

An interesting property of word vectors is that the relative distances between them appear to reflect more than just semantic similarity. A famous example by Mikolov et al. (2013) is analogies through "word arithmetics". Mikolov and colleagues noticed that they could take a word vector for "man", subtract it from the vector for "king", add the vector for "woman" and the resulting closest vector in the vector space would be "queen".

1) Man – King + Woman = Queen

Geometrically speaking, this means that the same distance and direction between pairs of word vectors can be interpreted as similar kind of semantic relationship, e.g. gender in example 1. This logic is illustrated in figure 5.

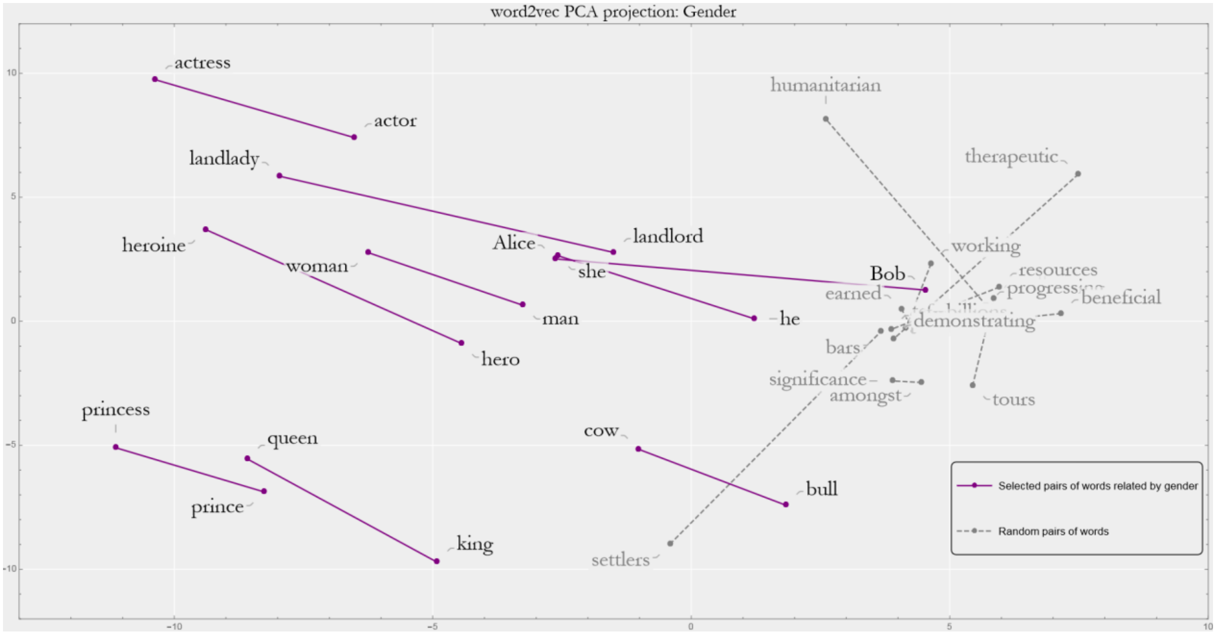


Figure 5 Distance and direction representing the gender relation, revealed in the vector space by a PCA projection. Adopted from Gastaldi (2021).

Different kinds of mainly semantic analogies have been discovered within word vectors made with Word2Vec but also syntactic patterns such as verb tenses or adjectival comparatives plotted in figure 6 and 7.

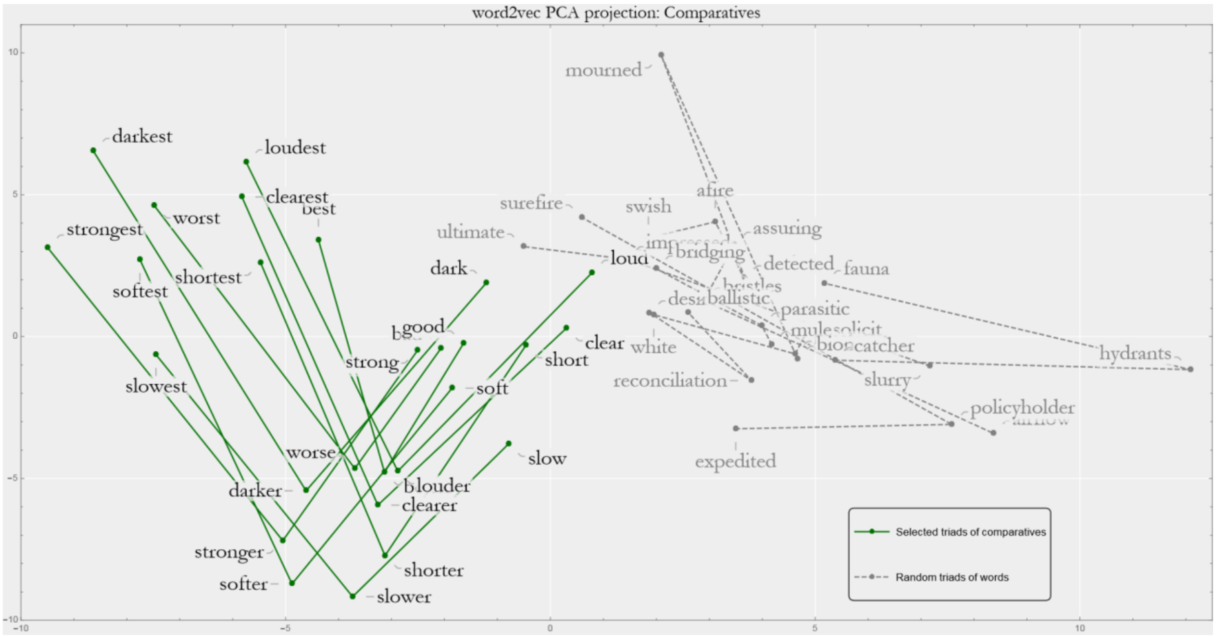


Figure 6 Pattern in the vector space (word2vec) corresponding to the comparative category (base, comparative and superlative forms). Adopted from Gastaldi (2021).

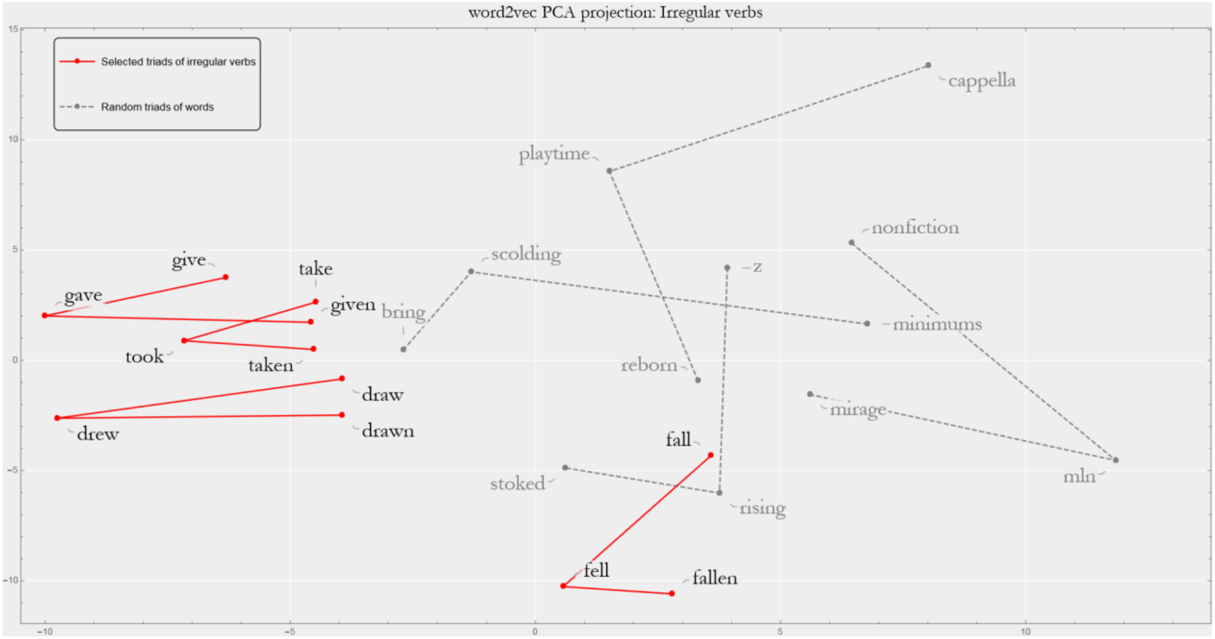


Figure 7 Pattern in the vector space (word2vec) corresponding to conjugation of irregular verbs. Adopted from Gastaldi (2021).

Word vectors are also able to track diachronic language change when computed on texts from different times. Figure 8 shows a plot from Hamilton et al. (2018) in which the development of the words “gay”, “broadcast”, and “awful” was studied.

Figure 8 Visualization of semantic change based on word2vec.

These examples provide evidence that word vectors can capture various kinds of semantic relationships from almost raw text. Figures 4 – 8 present visualizations of word vectors using statistical techniques that reduce the dimensions of the original vectors which also leads to loss of some of the information contained in them. The following section describes methods of this study of constructing a semantic network based on word vectors without reducing their dimentionality, thus conserving the full information about semantic similarity.

3. Method and Material

3.1. Network data

This work integrates computational methods to produce a semantic network of English based on a LLM, network science to analyse it and compare it with the results from psycholinguistic lexical retrieval experiments. One aim is to find meaningful correlations between these independent sources of information to follow on the research done on phonological networks. Another aim is to assess the relevance of LLMs for linguistic research and cognitive science more broadly. The practical part of this work includes the following steps:

1. Training BERT on a sample from The TV Corpus
2. Outputting word vectors from BERT
3. Creating a list of links by computing cosine similarity between every pair of word vectors
4. Visualizing the network in GEPHI
5. Obtaining reaction times from lexical retrieval experiments in MALD
6. Correlating reaction times with network measurements

Python programming language offers access to the open-source language model BERT via the *transformers* library. BERT is a state-of-the-art open source LLM. It can be trained on tokenized raw-text data to output word vectors. In this study, BERT was trained on a random sample of 1000 sentences from the TV Corpus, which is freely available for downloading as a plain text at english-corpora.org. The TV Corpus contains subtitles of tv shows from the 1950's to the present time. Subtitles are a valuable source of textual data that is more similar to spoken language than other sources of written text. The advantage of subtitle corpora over spoken corpora is its greater size.

The output of training BERT was a list of word vectors for each token. In total, there was 9683 tokens each of which was represented by 768 dimensional vectors. To ensure a meaningful representation of semantic similarity between word vectors, I excluded vectors containing non-alphabetic characters, thus eliminating punctuation, special symbols, and numbers. This reduced the number of word vectors to 6664. Table 2 contains a snapshot of the table with the filtered word vectors.

Table 2 A part of the table containing 6664 word vectors with 768 dimension computed by BERT

	0	0.1	1	2	3	
1	i	0.4131588	-0.5164368	-0.070106514	0.24670188	0.1
2	used	0.6167732	-0.935629	0.5481038	0.20566644	1.
3	to	0.33126172	-0.43050405	0.4024777	-0.4328461	0.7
4	date	0.6792053	-0.105997495	0.5854818	-0.86545646	0.2
5	a	-0.25574097	-0.69546574	0.51769054	-0.17579953	0.7
6	girl	0.3497564	-1.0572797	0.2570811	-0.24540454	0.0
7	who	-0.43124136	-0.5059501	0.3511485	-0.17387336	0.3
8	was	-0.036158808	-1.056186	0.3266946	-0.4658766	0.3
9	the	0.22171763	-1.2748561	0.6104466	-0.22556755	0.2
10	captain	0.6563785	-0.41294754	0.09312493	-0.5945709	-0.2
11	of	-0.8298847	-0.3123703	0.19391257	-0.1716561	-0.02
12	the	-0.20958453	-0.3160304	-0.034724604	-0.3607289	-0.02
13	netball	0.5075988	0.30638695	0.93784165	-0.5134297	-0.1
14	team	-0.2594932	-0.09756132	0.19902855	-0.76846117	-0.5
15	is	0.38111046	-0.4069852	0.18678449	-0.48942044	0.6
16	he	0.11544245	-0.497555	0.71277905	0.4922282	0.
17	okay	0.0743646	-0.3843242	0.7288899	-0.2971244	1.
18	but	0.3264922	0.12335851	-0.17650524	0.09791387	-0.

The first column contains word labels, the rest are values of the multidimensional vectors with 768 dimensions. The rows are the individual words.

Next, I calculated cosine similarity between every pair of word vectors which reflects how similar the vectors are based on the angle between them and the origin point of the vector space. The cosine similarity value ranges from -1 to 1. 1 indicates that the two vectors are identical in orientation. In the context of word vectors, this would mean that the two tokens are semantically very similar or nearly identical. 0 implies orthogonality of vectors, suggesting no semantic similarity between the two tokens. -1 is less common in word vectors but theoretically would indicate completely opposite meanings (Sidorov et al., 2014). However, interpreting these values can be somewhat opaque and context-dependent as there are no universally agreed-upon thresholds. Generally, higher values (closer to 1) indicate greater similarity. Often-used values for considering two vectors similar include 0.5 and 0.7 (Zhou et al., 2021). Figure 9 displays a histogram of the distribution of the cosine similarity values calculated between word vectors computed by BERT for the sample from the TV Corpus.

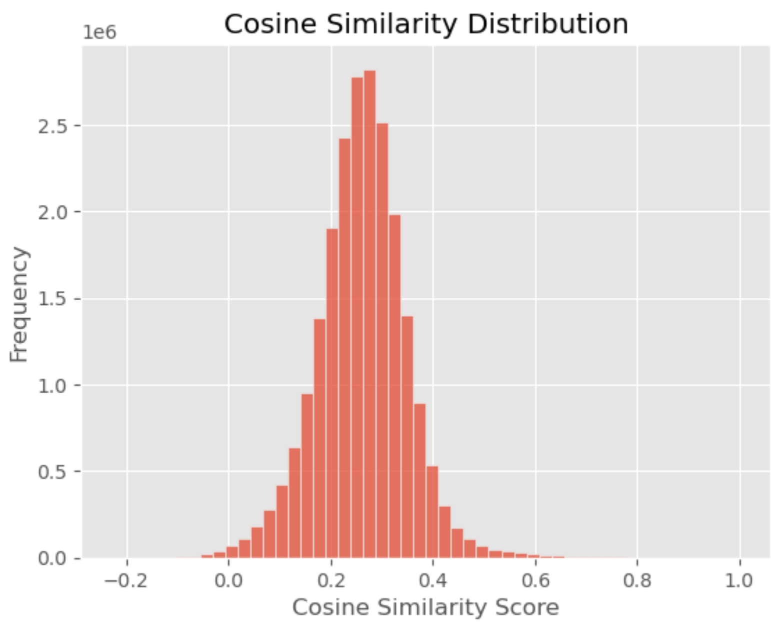


Figure 9 Cosine similarity distribution

It follows a normal distribution with the most common values around cosine similarity of 0.3. As values of cosine similarity above 0.7 are very sparse, the value of 0.5 is used as the basis for constructing the links between the nodes in the network. Therefore, nodes in the following semantic network will represent tokens and a weighted link is placed between each pair where the value of cosine similarity is greater than 0.5. The weight of the link is always equal to the value of cosine similarity for the respective pair of word vectors. This method produces the final edge list that was used for the network construction. This list was further reduced by only keeping nouns (without proper names) and verbs (without abbreviated verb forms, e.g. 're').

3.2. Word Processing Data

The final step was to include the data from lexical decision task from The Massive Auditory Lexical Decision (MALD) database. MALD database is a freely available auditory and production data set for speech and psycholinguistic research that contains time-aligned stimulus recordings for 26,793 words and 9592 pseudowords, and response data for 227,179 auditory lexical decisions from 231 unique monolingual English listeners. It is a valuable source of reaction times for different words from the auditory lexical decision task. The word reaction times from MALD database were extracted for words in the network. Reaction time values under 200ms and over 4000ms were excluded to avoid erroneous data and outliers. The rest of the reaction time values were averaged for each unique word. The final step of this study is to do a regression analysis of the word reaction times and network measures (closeness centrality, clustering coefficient, and degree centrality) to test whether there are any effects of network structure and measures on word processing. While the whole network contained 869 unique words, 237 were not in MALD database so the regression analysis was conducted on 632 unique words for which there is both the mean reaction time and network measures. Next section presents the results.

4. Research

Figure 10 shows a visualization of this network produced in the network visualization software GEPHI. It contains a total of 869 nodes (i.e. unique words) connected by 3196 edges. Edge thickness reflects its weight – the thicker, the higher the cosine similarity signalling semantic similarity. Furthermore, node size reflects its degree, therefore; the bigger the node, the higher its degree is. The layout of nodes does not reflect anything. The network appears to have SWS with clusters of high-degree hub nodes and more scarcely connected low-degree nodes.

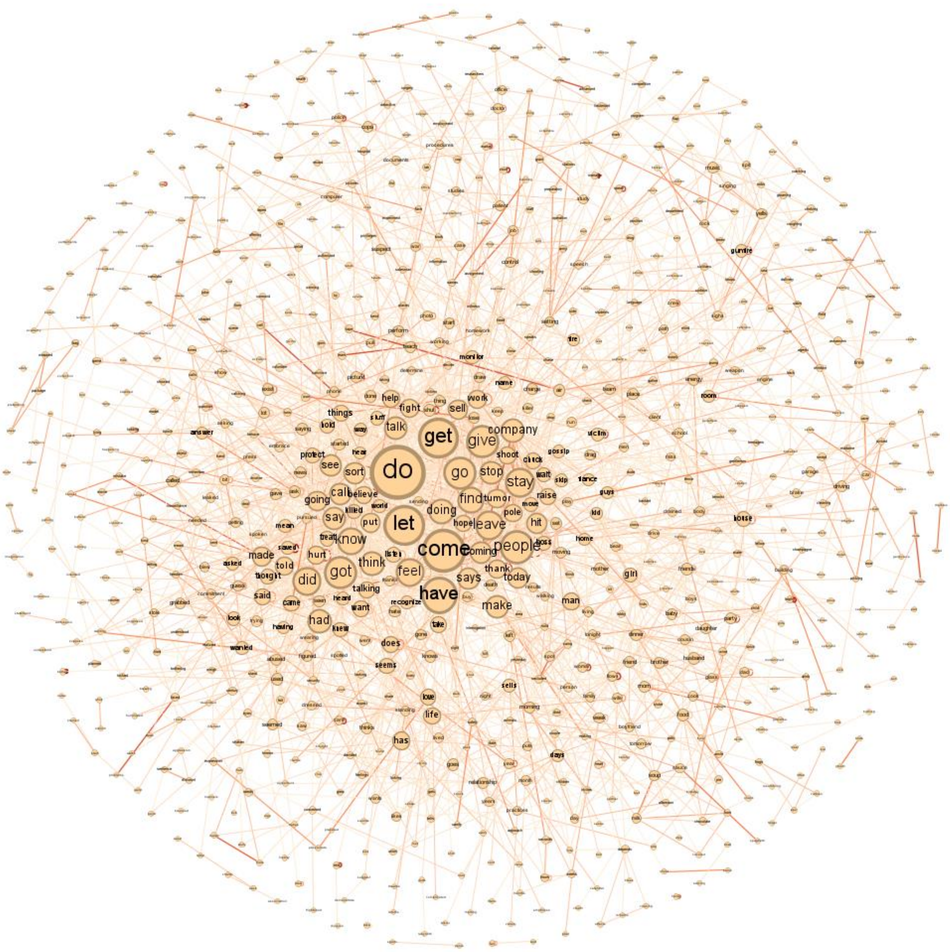


Figure 10 Semantic network based on cosine similarity measures of word vectors computed by BERT from a sample of The TV Corpus

Table 3 displays basic macroscopic measures of this network.

Table 3 Relevant macro measures of the semantic network

Mean degree	7.36
Network diameter	11
Mean clustering coefficient	0.15
Mean path lengths	4.15

Other network measures were computed for individual word nodes; namely, degree centrality, closeness centrality, and clustering coefficient. Figure 11 shows box plots of the distribution of these measures together the distribution of average word reaction times.

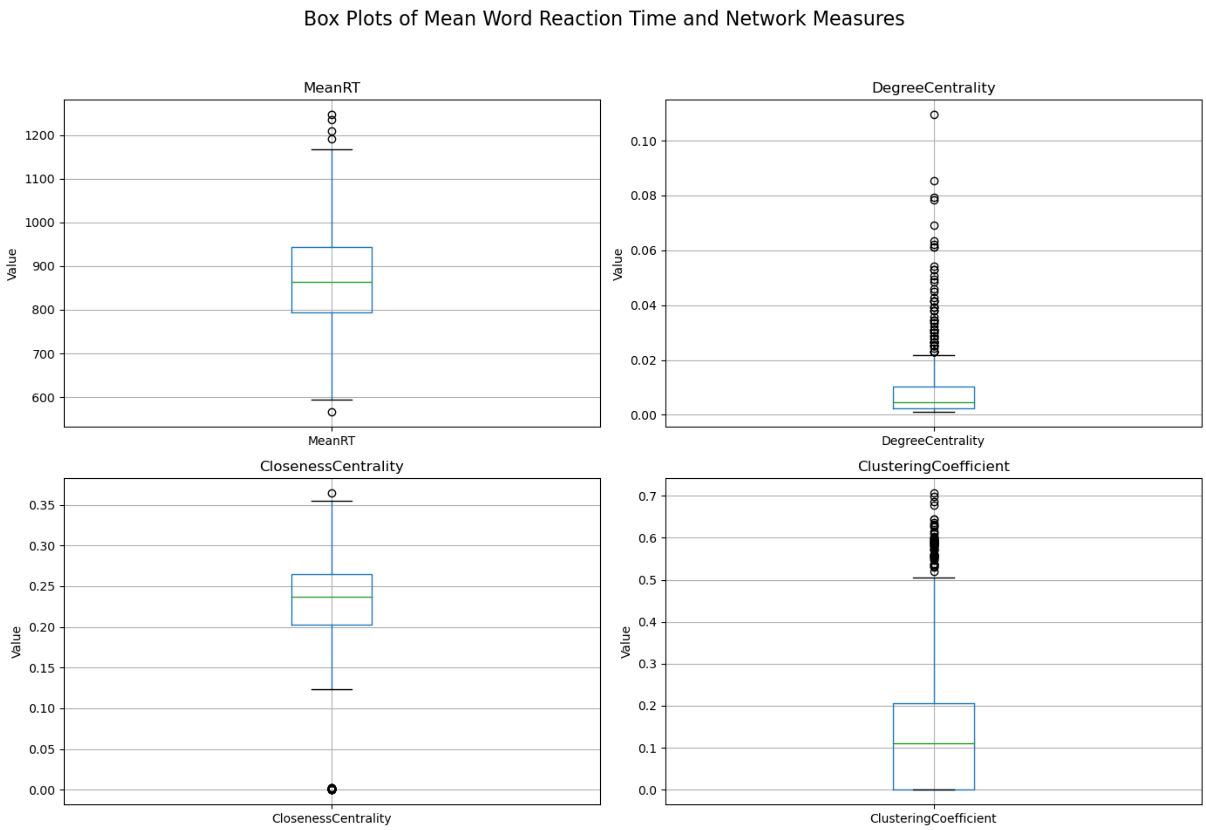


Figure 11 Box Plots of Mean Word Reaction Time and Network Measures

A multiple linear regression model was computed to analyse how the three network measures relate to the mean reaction time (MeanRT), with all variables being transformed using the natural logarithm. The model predicts the natural logarithm of MeanRT as a function of the natural logarithms of *degree centrality + 1*, *closeness centrality + 1*, and *clustering coefficient + 1*. The "+1" in the formula ensures that there are no issues with taking logarithms of zero. The model suggests there may be a negative relationship between degree centrality and MeanRT, meaning that nodes with higher degree centrality scores tend to have lower reaction times, possibly indicating more efficiency or priority in word processing, but it is not statistically significant. The overall fit of the model is quite weak, as indicated by the low R-squared value, meaning that other variables not included in the model might be influencing MeanRT. Table 4 shows summary of the regression model.

Table 4 Summary of the regression model

```
Call:
lm(formula = log(MeanRT) ~ log(DegreeCentrality + 1) + log(ClosenessCentrality +
    1) + log(ClusteringCoefficient + 1), data = stats_reg)

Residuals:
     Min       1Q   Median       3Q      Max
-0.42548 -0.08355 -0.00469  0.08331  0.36327

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                     6.75980    0.02222 304.214   <2e-16 ***
log(DegreeCentrality + 1)      -0.96550    0.51105  -1.889   0.0593 .
log(ClosenessCentrality + 1)    0.02516    0.11829   0.213   0.8316
log(ClusteringCoefficient + 1)  0.06403    0.04050   1.581   0.1144
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1251 on 624 degrees of freedom
Multiple R-squared:  0.01156,	Adjusted R-squared:  0.006803
F-statistic: 2.432 on 3 and 624 DF,  p-value: 0.0641
```

Added-variable plots are used to show the relationship between a given independent variable and the dependent variable, while accounting for the presence of other independent variables in the model. From these plots in figure 12, we can conclude that degree centrality has a slight negative impact on MeanRT, even after accounting for the other variables in the model. Meanwhile, closeness centrality and clustering coefficient do not seem to have a strong independent effect.

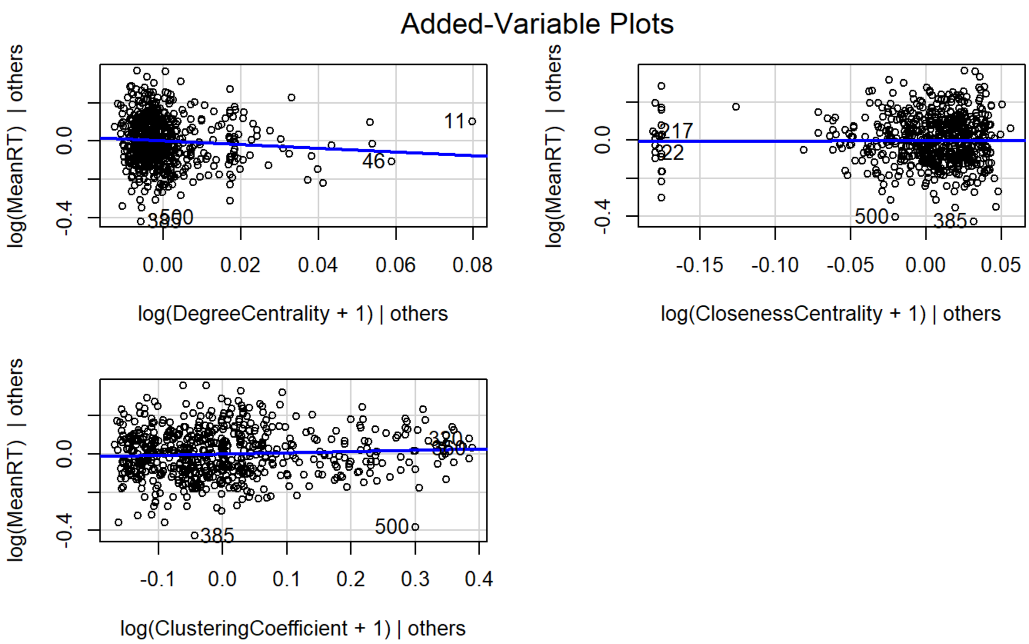


Figure 12 Added-variable plots

5. Conclusion

The results allow careful optimism for combining LLMs with network science to construct semantic networks for psycholinguistic research. First, the network based on semantic vectors from BERT had SWS which is in accord with other networks mapping cognitive and social phenomena. Second, degree centrality shows influence on word retrieval that approaches statistical significance with the results suggesting that when the number of neighbours increases, the reaction time decreases. It could be understood that more neighbours tend to speed up word processing, suggesting that mental lexicon leverages semantic vectors in a sense that it is able to recognize words faster when they are semantically similar to many others. Clustering coefficient and closeness centrality did not show a clear relationship with reaction time. Although the overall fit of the model was weak, work is being done on considering number of factors to improve it. This includes conceptualizing and calculating several control variables but also probing other kinds of training data for the semantic space in future semantic networks. This work used a sample from the TV Corpus that mimics spoken language, but other sources should be considered such as spoken corpora or experimentally elicited narration.

Nevertheless, considering that psycholinguistic effects are usually influenced by many variables and that the present results come from independent sources of information where BERT is designed for commercial purposes without an aim to capture a cognitively plausible representation of semantic, there is a potential for their future use in linguistic research. The present semantic network was constructed on a sample of 1000 sentences, a relatively small number that might be further decreased, which makes BERT and other state-of-the-art LLMs a powerful tool for representing meaning with various applications.

There is a growing body of research in both phonological (Kennet & Faust 2019) and semantic (Colunga & Sims 2017, de Boer et al. 2018, Hadley et al. 2019) networks that shifts towards more applied research, for example, of language acquisition and language impairments. Semantic networks capturing succinct semantic structure and its dynamics could contribute to the development of supplementary tools for the diagnosis of cognitive impairments that involve minute incremental changes in language in the early stages such as dementia or delayed language development.

6. References

Aitchison, Jean. *Words in the Mind: An Introduction to the Mental Lexicon, 4th Edition*. 4th edition. Chichester, West Sussex; Malden, MA: Wiley-Blackwell, 2012.

Barabási, Albert-László, and Márton Pósfai. *Network Science*. Cambridge: Cambridge university press, 2016.

Chan, Kit Ying, and Michael S. Vitevitch. 'The Influence of the Phonological Neighborhood Clustering Coefficient on Spoken Word Recognition.' *Journal of Experimental Psychology: Human Perception and Performance* 35, no. 6 (2009): 1934–49. https://doi.org/10.1037/a0016902.

Church, Kenneth Ward, and Patrick Hanks. 'Word Association Norms, Mutual Information, and Lexicography'. *Computational Linguistics* 16, no. 1 (1990): 22–29.

Colunga, Eliana, and Clare E. Sims. 'Not Only Size Matters: Early-Talker and Late-Talker Vocabularies Support Different Word-Learning Biases in Babies and Networks'. *Cognitive Science* 41, no. S1 (2017): 73–95. https://doi.org/10.1111/cogs.12409.

Boer, J. N. de, A. E. Voppel, M. J. H. Begemann, H. G. Schnack, F. Wijnen, and I. E. C. Sommer. 'Clinical Use of Semantic Space Models in Psychiatry and Neurology: A Systematic Review and Meta-Analysis'. *Neuroscience & Biobehavioral Reviews* 93 (1 October 2018): 85–92. https://doi.org/10.1016/j.neubiorev.2018.06.008.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 'BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding'. arXiv, 24 May 2019. https://doi.org/10.48550/arXiv.1810.04805.

Engelthaler, Tomas, and Thomas T. Hills. "Modeling Early Word Learning Through Network Graphs." In Network Science in Cognitive Psychology, edited by Michael S. Vitevitch, p. 166-183. 1st ed. Routledge, 2019. https://doi.org/10.4324/9780367853259.

Firth, John Rupert. *A Synopsis of Linguistic Theory, 1930-1955*, 1957.

Gastaldi, Juan Luis. 'Why Can Computers Understand Natural Language?: The Structuralist Image of Language Behind Word Embeddings'. *Philosophy & Technology* 34, no. 1 (March 2021): 149–214. https://doi.org/10.1007/s13347-020-00393-9.

Hadley, Elizabeth B., David K. Dickinson, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 'Building Semantic Networks: The Impact of a Vocabulary Intervention on Preschoolers' Depth of Word Knowledge'. *Reading Research Quarterly* 54, no. 1 (January 2019): 41–61. https://doi.org/10.1002/rrq.225.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 'Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change'. arXiv, 25 October 2018. https://doi.org/10.48550/arXiv.1605.09096.

Harris, Zellig S. 'Distributional Structure'. *Word* 10 (1954): 146–62. https://doi.org/10.1080/00437956.1954.11659520.

Jarmasz, Mario, and Stan Szpakowicz. 'Roget's Thesaurus and Semantic Similarity'. In *Current Issues in Linguistic Theory*, edited by Nicolas Nicolov, Kalina Bontcheva, Galia Angelova, and Ruslan Mitkov, 260:111. Amsterdam: John Benjamins Publishing Company, 2004. https://doi.org/10.1075/cilt.260.12jar.

Kenett, Yoed N., Faust, Miriam. 'Clinical Cognitive Networks A Graph Theory Approach'. In Network Science in Cognitive Psychology, edited by Michael S. Vitevitch, p. 136-165. 1st ed. Routledge, 2019. https://doi.org/10.4324/9780367853259.

Kipfer, Barbara Ann. *Roget's International Thesaurus*. HarperCollins Publishers, 2022.

Lund, Kevin, and Curt Burgess. 'Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence'. *Behavior Research Methods, Instruments, & Computers* 28, no. 2 (1 June 1996): 203–8. https://doi.org/10.3758/BF03204766.

McRae, Ken, George S. Cree, Mark S. Seidenberg, and Chris Mcnorgan. 'Semantic Feature Production Norms for a Large Set of Living and Nonliving Things'. *Behavior Research Methods* 37, no. 4 (November 2005): 547–59. https://doi.org/10.3758/BF03192726.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 'Efficient Estimation of Word Representations in Vector Space'. arXiv, 6 September 2013a. https://doi.org/10.48550/arXiv.1301.3781.

Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever. 'Exploiting Similarities among Languages for Machine Translation'. arXiv, 16 September 2013b. https://doi.org/10.48550/arXiv.1309.4168.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 'Distributed Representations of Words and Phrases and Their Compositionality'. arXiv, 16 October 2013c. https://doi.org/10.48550/arXiv.1310.4546.

Miller, George A. 'WordNet: A Lexical Database for English'. *Communications of the ACM* 38, no. 11 (autumn 1995): 39–41. https://doi.org/10.1145/219717.219748.

Nelson, Douglas L., Cathy L. McEvoy, and Thomas A. Schreiber. 'The University of South Florida Free Association, Rhyme, and Word Fragment Norms'. *Behavior Research Methods, Instruments, & Computers* 36, no. 3 (1 August 2004): 402–7. https://doi.org/10.3758/BF03195588.

Nelson, Douglas L., and Nan Zhang. 'The Ties That Bind What Is Known to the Recall of What Is New'. *Psychonomic Bulletin & Review* 7, no. 4 (December 2000): 604–17. https://doi.org/10.3758/BF03212998.

Preininger, M., Brand, J., & Kříž, A. (2022). Quantifying the Socio-semantic Representations of Words. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).

Sidorov, Grigori, Alexander Gelbukh, Helena Gómez-Adorno, and David Pinto. 'Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model'. *Computación y Sistemas* 18, no. 3 (29 September 2014): 491–504. https://doi.org/10.13053/cys-18-3-2043.

Siew, Cynthia S. Q., and Michael S. Vitevitch. 'Spoken Word Recognition and Serial Recall of Words from Components in the Phonological Network.' *Journal of Experimental Psychology:*

Learning, Memory, and Cognition 42, no. 3 (2016): 394–410. https://doi.org/10.1037/xlm0000139.

Siew, Cynthia S. Q., and Michael S. Vitevitch. 'An Investigation of Network Growth Principles in the Phonological Language Network.' *Journal of Experimental Psychology: General* 149, no. 12 (December 2020): 2376–94. https://doi.org/10.1037/xge0000876.

Vinson, David P., and Gabriella Vigliocco. 'Semantic Feature Production Norms for a Large Set of Objects and Events'. *Behavior Research Methods* 40, no. 1 (1 February 2008): 183–90. https://doi.org/10.3758/BRM.40.1.183.

Vitevitch, Michael S., ed. *Network Science in Cognitive Psychology*. 1st ed. Routledge, 2019. https://doi.org/10.4324/9780367853259.

Vitevitch, Michael S., Gunes Ercal, and Bhargav Adagarla. 'Simulating Retrieval from a Highly Clustered Network: Implications for Spoken Word Recognition'. *Frontiers in Psychology* 2 (2011). https://doi.org/10.3389/fpsyg.2011.00369.

Vitevitch, Michael S. 'What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval?' *Journal of Speech, Language, and Hearing Research* 51, no. 2 (April 2008): 408–22. https://doi.org/10.1044/1092-4388(2008/030).

Vitevitch, Michael S., Nichol Castro, Gavin J. D. Mullin, and Zoe Kulphongpatana. 'The Resilience of the Phonological Network May Have Implications for Developmental and Acquired Disorders'. *Brain Sciences* 13, no. 2 (23 January 2023): 188. https://doi.org/10.3390/brainsci13020188.

Watts, Duncan J., and Steven H. Strogatz. 'Collective Dynamics of "Small-World" Networks'. *Nature* 393, no. 6684 (June 1998): 440–42. https://doi.org/10.1038/30918.

Zhou, Kaitlyn, Kawin Ethayarajh, Dallas Card, and Dan Jurafsky. 'Problems with Cosine as a Measure of Embedding Similarity for High Frequency Words'. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* 401–23. Dublin, Ireland: Association for Computational Linguistics, 2022. https://doi.org/10.18653/v1/2022.acl-short.45.