Research Article

László Kovács*, András Bóta, László Hajdu, and Miklós Krész

Networks in the mind – what communities reveal about the structure of the lexicon

Abstract: The mental lexicon stores words and information about words. The lexicon is seen by many researchers as a network, where lexical units are nodes and the different links between the units are connections. Based on the analysis of a word association network, in this article we show that different kinds of associative connections exist in the mental lexicon. Our analysis is based on a word association database from the agglutinative language Hungarian. We use communities – closely knit groups – of the lexicon to provide evidence for the existence and coexistence of different connections. We search for communities in the database using two different algorithms, enabling us to see the overlapping (a word belongs to multiple communities) and non-overlapping (a word belongs to only one community) community structures. Our results show that the network of the lexicon is organized by semantic, phonetic, syntactic and grammatical connections, but encyclopedic knowledge and individual experiences are also shaping the associative structure. We also show that words may be connected not just by one, but more types of connections at the same time.

Keywords: networks, mental lexicon, associations, multilayered network, Hungarian, communities

1 Introduction

The mental lexicon – dictionary of mind – stores information in lexical units. These units do not stand alone: they are connected to each other, enabling the activation of the units. The lexicon is thus seen by many researchers as a network, where lexical units are nodes and the different links between the units are connections (cf. Aitchison 1987, Quillian 1968, on more recent approaches see the edited volumes of Mehler et al. 2016, Vitevitch 2020a). Although the network character of the lexicon is receiving more and more attention from researchers, almost all studies are done on databases of Indo-European languages.

It has been suggested in the literature before that this network of the mind is a multilayered network: different kinds of connections exist in the lexicon which can be assigned to different layers (Stella et al. 2018). The characteristics of the multilayered lexicon are not described in detail however. Kovács (2013)

* **Corresponding author: László Kovács,** Savaria Department of Business Administration, Faculty of Social Sciences, Eötvös Loránd University, ELTE Savaria University Centre, Károlyi Gáspár square 4, H-9700, Szombathely, Hungary, e-mail: kovacs.laszlo@sek.elte.hu

András Bóta: Lulea University of Technology, Department of Computer Science, Electrical and Space Engineering, Embedded Intelligent Systems Lab, 97187, Lulea, Sweden, e-mail: andras.bota@ltu.se

László Hajdu: Innorenew CoE, Livade 6, SI-6310, Izola, Slovenia; University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technologies, Glagoljaška 8, SI-6000, Koper, Slovenia; University of Szeged, Institute of Informatics, Árpád tér 2, H-6720, Szeged, Hungary, e-mail: laszlo.hajdu@innorenew.eu

Miklós Krész: Innorenew CoE, Livade 6, SI-6310, Izola, Slovenia; University of Primorska, Andrej Marušič Institute, Muzejski trg 2, SI-6000, Koper, Slovenia; University of Szeged, Faculty Education, Boldogasszony sgt. 6, H-6725, Szeged, Hungary, e-mail: miklos.kresz@innorenew.eu

assumed that the mental lexicon itself can be seen as a multilayered network, where several interacting layers form the lexicon.

In this article, our goal is to validate this earlier assumption and identify the possible layers of the lexicon by analyzing the communities – closely knit groups – inside the lexicon. Communities can reveal the organizing forces of the mental lexicon, thus enabling us to identify the coexisting layers of the lexicon. To demonstrate the existence of the layers, we examine a word association database collected in Hungary. Since the analyzed database is not monolingual – it contains associations in other languages – the database enables us to see the interaction of different languages. Our results allow us to hypothesize the multilayered network structure of the multilingual lexicon.

In the first part of this article, we give a short overview about networks and networked structures of the lexicon: what networks are and how networks exist in the lexicon. The second part of this article shows which communities are detected in the analyzed Hungarian word association database. We examine the database with different community detection methods. Based on the results we describe the multilayered structure of the lexicon. Results show that different layers of the lexicon exist and each layer can play a role by activating words in the lexicon.

2 Background

2.1 Networks

Networks are interconnected systems where the elements of the system – nodes – are linked to each other (cf. Barabási 2016, Menczer et al. 2020). Research in the end of the 20th century revealed that networked structures can be found all around us: they are captured and analyzed across all disciplines from medicine (Loscalzo et al. 2017) to geography (Barthélemy 2011) or economy (Bramoullé et al. 2016, Easley and Kleinberg 2010, Jackson 2008).

Networks consist of elements – nodes – and links connecting the nodes. Nodes and links can represent various things: in a social network, nodes are people while links represent, e.g., acquaintance; in a geographical or in a transportation network, nodes can represent cities, while links may indicate roads connecting the cities. Wikipedia can also be considered as a network, where the nodes are the pages of Wikipedia and the hyperlinks between these pages are the links (for an overview and more examples see Barabási 2016 and Menczer et al. 2020).

According to the links between the nodes, networks can be weighted or unweighted and directed or undirected. Weighted networks consider the strength of the connection between nodes, while in unweighted networks only the existence of the connection is important; its weight has no relevance. Take as an example a transportation network. In a weighted network two cities connected via an interstate highway have a stronger connection, as when they were connected only by a backroad. In an unweighted network, however, the two kinds of connections – highway and backroad – are not taken into account: the only important thing is that a road connects the cities.

Road systems are in most cases undirected networks: traffic can flow in both directions. Other networks are directed. In directed networks, the direction of the link is also analyzed: e.g., in a hierarchical network of people, such as a military organization, commands "flow" just in one direction.

In networks, three hierarchic levels are distinguished: micro-, meso- and macro-level (Vitevitch et al. 2014; Siew et al. 2019 use the notions microscopic, mesoscopic and macroscopic network measures).

At the micro-level, the degree of the nodes and the clustering coefficient are relevant. The degree of a node describes how many connections – links – a node has. In a directed network, we can speak of indegree and outdegree: indegree indicates the incoming links (connections toward the node), while outdegree specifies the links originating from the node (connections from the node). The clustering coefficient shows "the extent to which the neighbors of a given word are also neighbors of each other" (Vitevitch et al. 2014:

124; cf. Siew et al. 2019). Centrality measures also describe the micro level (Siew et al. 2019). On the macro-level degree distribution, average node measures (Siew et al. 2019) and mixing are important (Vitevitch et al. 2014). Nodes in most networks can be categorized into types according to their properties or attributes, and the probability of connections between nodes depends on these types (Newman 2003). In social networks for example, nodes representing people with similar interest or social-economic status are more likely to be connected. This is called assortative mixing. In a food web including carnivores, herbivores and plants with links denoting "eating," we would see many connections between plants and herbivores, and herbivores and carnivores, while few connections among plants and other plants, or herbivores and other herbivores, giving an example of disassortative mixing. Finally, mixing can be based on strictly graph theoretic properties of nodes. In some complex networks, nodes with similar degrees are more likely to be connected to each other, while in other complex networks, the opposite is true, indicating assortative and disassortative mixing, respectively.

In between these two the meso-level of networks exists, which is characterized by path length and communities (Vitevitch et al. 2014). Communities are closely knit groups in networks. The distribution of the edges is inhomogeneous both globally and locally, thus highly connected groups of nodes can be identified, creating communities (Fortunato 2010). One of the most natural examples of communities can be observed in social networks, where people tend to form groups according to their family, jobs, hobbies and other interests. Modularity is also important at the meso-level (Siew et al. 2019). The formula of modularity shows how the structure of edges within the communities relate to a random clustering; a higher value indicates a better quality of community detection (Newman 2006).

Path length – also part of the meso-level of networks – shows how many connections separate two nodes: the shortest possible way is called shortest path (cf. Barabási 2016). Siew and colleagues (2019) consider however, path length as part of the micro-level of networks. Advances in the analysis of networked structures lead to the emergence of network science (Newman 2003, Menczer et al. 2020).

The notion of multilayer networks is also relevant. In a multilayer network, different kinds of networks exist, which can be combined. For example, an air transportation network between cities can be defined according to different airlines flying to different cities, creating different networks belonging to specific airlines. On the other hand, we can analyze all the airline connections of the selected cities – in this case we want to see all connections including all airlines. In this case, the layers of the individual airline networks are combined and we can speak of a multilayer network (Menczer et al. 2020: 26). When these layers use the same set of nodes the network is called a multiplex network (Menczer et al. 2020).

2.2 Networks and language

Language also has its networks and networked structures (cf. Cong and Liu 2014, Solé et al. 2010, Kovács 2013, Mehler et al. 2016, Siew et al. 2019). Some of these structures – and the effects of the structures – were known and investigated before the emergence of network science, such as semantic networks and the spreading activation theory (cf. Bovasso et al. 1993, Collins and Loftus 1975, Figueroa et al. 1976, Quillian 1968), the networked structures of speakers (Milroy 1987), the mental lexicon (Aitchison 1987; Meara 2009) or the diverse Parallel Distributed Processing theories and connectionist models (cf. Rogers and McClelland 2003; Rogers and McClelland 2004).

These approaches all use the term network; however, in most cases the notion network was rather considered as a metaphor (Vitevitch 2020b). Accordingly, networks denoted hierarchical organized structures, containing several types of connections (Collins and Loftus 1975; Quillian 1968), or they explained cognitive processes with nodes representing neurons in PDP models (Vitevitch 2020b). Modern network approaches have their roots in graph theory, in physics (complex systems), social network analysis and computer science (Vitevitch 2020b, based on Newman). Previous network metaphors rather described processes, whereas modern network science analyzes complex real-word networks and exerts to describe and quantify their structure with mathematical methods (Castro and Siew 2020; cf. Vitevitch 2020b). The

propagation of network research however, enables a deeper understanding of different structures and processes connected to language. Describing language as a network allows us to see elements of language in their complex, interconnected structure. These new approaches lead to new explanations on several fields, e.g., phonetics (Siew and Vitevitch 2019), lexical processing (Castro et al. 2017, Goldstein and Vitevitch 2017), word learning (Goldstein and Vitevitch 2014), cognitive science (Baronchelli et al. 2013, Siew et al. 2019), syntactic structures (Liu and Xu 2011, Jingyang et al. 2019) or learning grammar (Ibbotson et al. 2019; for an overview see Cong and Liu 2014, Mehler et al. 2016 and Solé et al. 2010). The provided new methods and tools of network science are one cornerstone of research on linguistic networks; the other is the available linguistic data collected and/or analyzed by computer programs.

The communities introduced in the previous section – closely knit groups of nodes – also exist in linguistic networks. Community detection shows that both word length and word frequency contribute to community building in phonological networks in English (cf. Siew 2013), while community detection in syntactic networks in Chinese reveals that words of the same class are part of the same community and both word length and word frequency contribute to community building (cf. Cong and Liu 2014). Communities also enable a better understanding of the cognitive system (Siew et al. 2019).

One of the research areas in linguistics, which enormously benefits from the new tools and methods of network science, is psycholinguistics and the analysis of the mental lexicon. The mental lexicon stores words – lexical units – in the memory. These units are connected to each other (Jackendoff 2002, Klein 2015). The units of the lexicon are traditionally investigated with word associations. In word association experiments, a cue word is presented and subjects are asked to name as a response the first word which comes to their minds. The method has a long tradition dating back to Galton (1879, 1883) and is widely used in linguistics (Cramer 1968), in psychology and philosophy (Mandelbaum 2017) and also in marketing (Kovács 2019); for an exhaustive overview of the method see Cramer (1968).

The mental lexicon stores information in lexical units, which contain information about the meaning, morphology, phonology and syntax of an item, and additionally also pragmatic and stylistic information (Levelt 1993: 182–183, see also Bonin 2004 viii). Klein (2015: 937) describes the units in a similar way: they consist of phonological (pronunciation), graphematic (written form), morphosyntactic (word class, inflection) and semantic (meaning) properties with additional information such as frequency, semantic counterpart (equivalent) in other languages or encyclopedic knowledge.

Levelt (1993: 183–184) emphasizes that these properties of lexical units connect the units with each other with intrinsic or with associative connections. Intrinsic connections can arise, e.g., through semantic information, through the same morphological stem, through the same phonemes and he assumes also the relation between syntactical information, e.g., words of a word class. In associative connections. "the basis lies in the frequent co-occurrence of the items in language use" (Levelt 1993: 184). He rates antonymy as an associative connection. This interrelation of the units – e.g., as being the antonym or synonym of each other – is also highlighted by Klein (2015). The mental lexicon thus can be seized as a network, where lexical units are nodes, which are connected to each other (Figure 1).

Language is considered not only as a network but also as a multi-level system, where dynamic networks exist and function parallel to each other (Boccaletti et al. 2014, Cong and Liu 2014, Liu and Cong 2014).

Multilayer networks (syntax, co-occurrence, syllable and grapheme level) were analyzed on a corpus of English and Croatian sentences by Martinčić-Ipšić et al. (2016). Their results show that multilayer networks can be used to understand the networked structure of language better. Stella et al. (2017) applied multiplex networks on modeling early word learning, showing that multiplex linguistic networks help to predict word acquisition of young children. Recently, Siew and Vitevitch (2019) focused on orthography and phonology, creating phonographic networks, where links exist when they are present on both the phonological and the orthographical layer. Results show that phonographic degree helps in visual word recognition, while phonographic clustering coefficient is important for spoken word recognition.

In networks consisting of multiple layers, nodes can be connected on one layer (intralayer connections) and across layers (interlayer connections) (Vukić et al. 2020). Describing knowledge organization Vukić et al. (2020) identify a multidimensional knowledge network, consisting of four interacting layers: factual, conceptual, procedural and metacognitive layers.
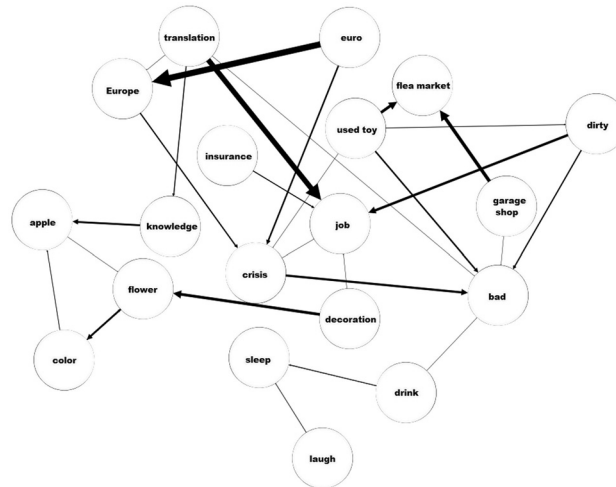
Figure 1: A subgraph of the analyzed dataset.

2.3 Networks in the lexicon

The networked structures of the lexicon were researched in the past few decades in several languages, on several corpora and in several contexts, starting with e.g., the properties and characteristics of semantic networks (Coronges et al. 2007, Griffiths et al. 2007, Steyvers and Tenenbaum 2005), the analysis semantic relations like homonymy or polysemy (Gravino et al. 2012), phonetic networks (Vitevitch 2008), second language acquisition (Schur 2007), networked properties of multiple associations (De Deyne and Storms 2008), resemblances between the collocation networks of mother and their children (Ke and Yao 2008) just to mention a few.

Several of these researchers describe the networked structure of the lexicon by analyzing word association data, where a connection is formed between two words when the cue word elicited the other word as a response. One of the often-analyzed word association databases is the South Florida Word Association Database (Nelson et al. 2004).

Assortative mixing – nodes are connected to similar nodes – can be observed in word association tasks, e.g., syntactic forms facilitate similar syntactic forms, concrete words are likely to connect to concrete words (for an overview see Van Rensbergen et al. 2015). Assortativity in the lexicon also may be elicited by shared features of concepts, semantic, phonological or orthographical similarity or co-occurrences (Siew et al. 2019). Although assortativity describes and explains structures and connections in the lexicon, the exact forces leading to assortativity – e.g., semantic similarity, learned co-occurrences or other influencing factors – are yet to be understood more clearly (Van Rensbergen et al. 2015).

The research on communities offers a new approach to the networks of the lexicon. In the lexicon, communities exist at the meso level of networks. In the mental lexicon, these communities can be detected by seeing which cue words and associations form closely knit groups (De Deyne et al. 2016, De Deyne et al. 2017).

The field of community detection contains a large number of detection algorithms, focusing on different types of networks, metrics and real-life scenarios. Detection algorithms are mostly defined for undirected and unweighted networks (Blondel et al. 2008, Newman and Girvan 2004, Peixoto 2014, Lancichinetti et al. 2009, Wu et al. 2012, Bóta and Krész 2015), but there are examples for finding directed (Fragkiskos and Vazirgiannis 2013) or weighted (Aicher et al. 2015) communities too. The largest dividing point in the literature is whether communities should be defined as non-overlapping (Aicher et al. 2015, Blondel et al. 2008, Newman and Girvan 2004, Peixoto 2014) or overlapping communities (Lancichinetti et al. 2009, Wu et al. 2012, Bóta and Krész 2015).

Non-overlapping community detection partitions the nodes of the network into disjunct (non-overlapping) sets. This way, each node is assigned to a single community. This approach offers a strict mathematical definition, and makes it easier to evaluate the quality of output of the detection methods (Newman 2006, Peixoto 2014). However, in real-life networks, nodes are often shared between communities. As such, the node sets corresponding to the communities may overlap. Overlapping detection methods support this definition. However, there is no consensus on the exact definition of overlapping communities, making the comparison of these methods difficult.

De Deyne et al. (2016) describe several possibilities for such communities, e.g., taxonomic structures for animals or frames, scripts and schemata. They used in an empirical study the Order Statistics Local Optimization Method community finding algorithm on a Dutch word association network and they show that communities obtained by community detection methods are comprehensible, however, more analysis of communities of the lexicon is necessary to see the exact role of communities.

In a recently published paper, Citraro and Rossetti (2020) introduced a novel approach of community detection which can be used for the analysis of the mental lexicon. They considered node attributed linguistic networks with semantic and phonological layers and applied a new clustering method algorithm – EVA – which considers both the topological structure and the node homogeneity criteria. The node attributes reflect different psycholinguistic variables (e.g., age of acquisition, semantic size) which makes it possible to use the clustering method with the same topology, but with different attributes.

3 Methods

3.1 Expanding the scope: new language, new cues

Most research on word association data is conducted on English (e.g., Gravino et al. 2012, Steyvers and Tenenbaum 2005) and Dutch (e.g., De Deyne and Storms 2008) corpora.

Our work examines another language: Hungarian. Hungarian being an agglutinative language uses various affixes which indicate grammatical relations. By analyzing the network structures of the Hungarian lexicon, a non-Indo-Germanic language is investigated, also showing how agglutination influences associations.

We use a smaller dataset than other extensive studies on the topic, in order to enable us to analyze the data not only in an automated way but also examine and categorize all the results (in this case communities) manually. This critical evaluation is important to verify that the communities are not meaningless (that the used algorithms function in the proper way), since results of automated processes are partly contradictory when examined closely (cf. Brookes and McEnery 2019).

The main specifics of the Hungarian database are the following: the database grew on a natural way, meaning apart from the initial word list all data was created as response to previous associations. Proper names are also included in the database, making it unique. Since Hungarian is an agglutinative language, it was decided that inflected forms can also be part of the database – however, the initial word list contained no inflected forms. The length of associations was not restricted, therefore, longer associations (e.g., collocations and proverbs containing more than two words) are also present in the database.

3.2 Data

Data for current research were collected via an online program – agykapocs.hu – designed for collecting associations on a website. The website was active between 2008 and 2014 and collected more than 182,000 associations, from 1,035 respondents, primarily in Hungarian. The project was a new approach to studying the lexicon in Hungary, and both the project and the results were described in detail – although only in

Hungarian – in a book (Kovács 2013). For analysis we used the data of this previous research; current manuscript generated no new data.

On the website cue words were presented to the participants, and they had to write down the first association which came to their mind reading the presented word. No restrictions were given – any word was accepted up to 255 characters. Next to the blank space for the answer word were two buttons: one button labeled "OK" and one labeled "No associations" – the latter could be clicked, if the test person could not (or did not want to) give any association. After giving the associations and pushing the "OK" (or the "No associations") button, the next stimulus appeared on the screen (Kovács 2013). Test persons conducted the test on a voluntary basis and they could any time abandon the test just by navigating away from the webpage.

The first 134 words appeared in a given order for every participant. After the initial 134 words additional cue words appeared in a random order. These random words came from the answers given by other participants. The first 134 cue words ensured that a large number of associations were collected for the given word, while the random cue words from the database ensured that a network can be created from the obtained word association data.

The cue words and the given associations were stored in a MySQL database. For each association pair, the database also recorded the timestamp (date) of the recording and the elapsed time from the appearance of the stimulus to the submission of the answer. The idea was that the associations given by the respondents can function as a cue word too. It was however, necessary, that the words given by the respondents do not appear automatically as a cue word. Thus, in the background software researchers could see the list of recently given associations, with the opportunity to take the following actions on the words (Kovács 2013):
- Delete non-words when test persons typed meaningless characters (e.g., *hfghsdf*).
- Modify words containing a typo (e.g., *föváros* [*capital*] – *Budapst* [wrong spelling] modified to Budapest).
- Modify the language associated with the given word.
- Allow answers to function as a stimulus word.

In some cases, the decision was made that the answers remain part of the database but they could not function as cue words. This decision was made in the following cases (Kovács 2013):
- Character sets containing only numbers.
- Swear words and sexual terms.
- Word parts having no meaning without being attached to other words.

From the cue words and associations a network was created, where cue words were connected to the associations given to the cue word. The created network is a weighted, directed network. Weight arises from the frequency of the associations: if one person gave the association *apple* → *pear*, the association has a strength of one. When 15 people give the same answer (15 people trigger to the cue word *apple* the association *pear*), the strength is 15 and so on. Direction arises from the cue word: *apple* → *pear* means that apple triggered pear.

Although the database is primarily in Hungarian, the initial word lists existed in the database in several languages and associations were not restricted to Hungarian only. Thus, several hundred English, German and Italian associations were collected, too. The language associated with the words was also stored.

3.3 Community detection

The two community detection methods discussed in this article are hub percolation (Bóta and Krész 2015, Hajdu et al. 2018) and modularity maximization (Blondel et al. 2008). The first one belongs to the category of overlapping detection method for both directed and undirected networks, while the second one finds non-overlapping methods on undirected networks. A short description of each method follows.

3.3.1 Hub percolation

The hub percolation method (Bóta and Krész 2015) is based on two intuitive ideas. In graph theory, cliques are defined as subgraphs of a graph, where all nodes are connected to all other nodes in the set of nodes corresponding to the subgraph. Cliques defined this way represent an ideal community. However, some nodes of the network are more important than others in holding communities together. They act as the hubs of the network.

The hub percolation method first finds the set of all maximal cliques, then selects the hubs of the network according to a customizable hub selection strategy. The algorithm finds the communities of the network by extending and merging cliques containing hubs. The hub selection strategy allows the user to adapt the results of the method to specific needs of the application. The method produces overlapping communities.

3.3.2 Directed hub percolation

The directed version of the hub percolation method (Hajdu et al. 2018) detects directed cliques instead of undirected ones as in the original hub percolation method. Directed hub percolation is similar to the undirected method. The main difference is that the hub selection strategies are modified in a way that the direction of the edges is taken into consideration during the selection of the hubs. As before, the algorithm finds the set of directed cliques in the network, then selects hubs based on the directed hub selection strategy. In the last step, due to the directed edges, an additional parameter is introduced that defines the merging strategy of the cliques. This method also produces overlapping communities.

3.3.3 Modularity maximization

Modularity maximization aims to maximize the value of modularity as defined in the study of Newman (2006). This value describes the quality of non-overlapping community detection by counting the difference between the actual number of edges and the expected number of edges between node pairs. The algorithm finds the community structure that maximizes the value of modularity. In this article we used the algorithm introduced by Blondel and colleagues (2008).

4 Results

Since individual community detection methods may produce different community structures, we decided to perform our analysis with each algorithm considered in this article separately. This allows us to look at the structure of the word association network from three different perspectives and avoid the potential short-comings of the individual approaches. In this section, we present our findings for each method separately and provide a comparison of the results in Section 5.

4.1 Results for undirected hub percolation

We started our analysis with the communities found by the hub percolation method. As we mentioned before, these are overlapping communities, while the network itself was considered to be undirected – only the existence of a connection was important, without the direction of the connection, therefore, in this case the cue → association direction is neglected.

The algorithm detected 2,717 communities, with 63,328 words in these communities. Since the communities are overlapping, words can be part of different communities at the same time. The communities have 3 to 255 members, from which 106 are three-word communities and 57 are four-word communities. We closely examined the communities seeking to identify some common organizing principle behind their structure. Based on De Deyne et al. (2016) and on Siew and Vitevitch (2019) we assumed that communities of semantical properties (e.g., domains, synonyms) and communities of phonetical similarities exist, while based on Kovács (2013) we assumed that communities with encyclopedic knowledge – "world knowledge" – can also be identified.

The communities identified by the algorithm are difficult to categorize. Of the three-word communities, 89 are semantical communities containing synonyms or words from the same domain, 9 are encyclopedic, containing information like Shakespeare – Rome – not to be (all examples except for Figure 3 are translated from Hungarian). There was one community of conjunctions (because – therefore – just) and also in one case there was no apparent connection between the words (I am not interested – VAT- nothing). In six cases, there existed at least two different kinds of connections in the three-word-communities, where a word functioned as a bridge connecting the two other words (Hun. teve (camel) [t v] – tv (tv) [te: ve:] – hiradó (news reports)). Some of the communities exhibit inflectional resemblances, e.g., in the community prepare – manufacture – make, all three forms are infinitive forms (gyártani – késziteni – csinálni), while in takes a meal – eats – driks (étkezik – eszik – iszik) all forms are third singular. Similarly, in the community wheat – barley – sows – harvests (uzát – árpát – vet – arat) sow and harvest are in third singular, while wheat and barely are singular accusative forms (ending with -t).

Out of the 57 examples of four-word-communities, in 33 cases at least two different relationships existed inside the community or the organizing force behind the community is hard to explain, e.g., customs – smile – country – politician, hour/watch – economy – friendship – school. The larger communities displayed a more heterogeneous character with even more types of relations inside the community.

4.2 Results for directed hub percolation

We also used the hub percolation method while considering the database as a directed network. The algorithm identified 58 communities containing 740 words. The communities have 4–29 members. All the identified communities are semantic domains, which organize around one or more key concepts. For example, the community of landscape/environment also contains cloth through the word clean (Figure 2).
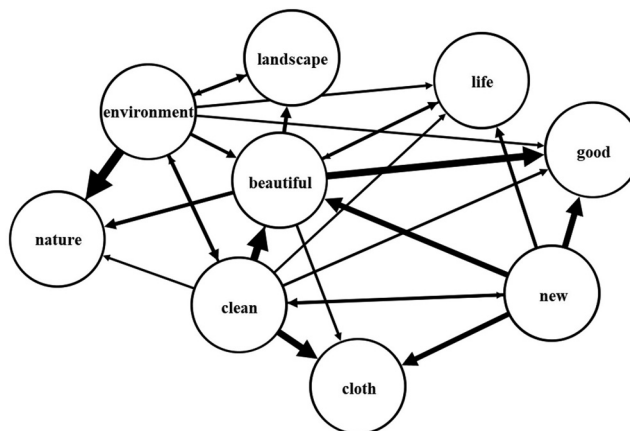

Figure 2: Community structure identified by directed hub percolation.

4.3 Results for modularity maximization

The heterogeneous nature of the communities found by both versions of the hub percolation method led us to try another approach. We selected a modularity maximization algorithm to see if it detects more homogenous communities. The communities identified by the modularity maximization algorithm are not overlapping. With this algorithm we identified more small communities, since larger communities are non-homogeneous and often incorporate bridging elements, which connect two communities.

First, we searched for communities containing associations named at least by two participants, to ensure that the found communities are not individual. The algorithm identified 692 communities containing 8,710 different words. The community sizes were from 2 to 290 members. There were 528 communities containing two words, 63 communities containing three words, 23 four-word and 10 five-word communities.

Next, we searched for communities in which all associations were named by at least four participants, to see even more stable structures and to see whether they differ from those named by two or more respondents. With this setting, the algorithm found 164 communities containing 2,801 different words with community sizes between 2 and 114 members. A total of 107 communities have only two and four have three members.

As with hub percolation, larger communities are primarily semantic communities; however, looking closer at the smaller (2-3-4 members) communities, the organizing force behind the community structure is more easy to identify (Table 1).

In Table 1, the first category names refer to cues and associations, which together form a name, like Wiener – Neustadt, Leonardo – da Vinci, Kevin – Costner, Rocky – Balboa, James – Bond. In these cases, the first organizing principle was names; but the knowledge of these names presupposes the existence of encyclopedic knowledge, therefore, in these cases the second organizing principle is encyclopedic knowledge. The table suggests that in several cases, the organizing force behind the community is not a single common characteristic of the communities: members share more than one common property. In this regard, the first and second categories represent no hierarchy – they only show that in several cases more than one common characteristic of the members is present at the same time. For example, in case of borotva (razor) + penge (blade) there is a semantic domain relation – however, put together borotvapenge (razor blade) is a

Table 1: Communities identified by the modularity maximization algorithm

Community category	Association by at least two subjects		Association by at least four subjects	
	Number of communities (first characteristics)	Number of communities (second characteristics)	Number of communities (first characteristics)	Number of communities (second characteristics)
Names	22	10	2	1
Semantic	607	0	159	—
-Synonym	240	—	49	—
-Domain	172	—	62	—
-Antonym	56	—	13	—
-Collocation	71	—	20	—
-Pair	27	—	5	—
-Categorization	17	—	2	—
-Abbreviation	1	—	0	—
-Associative	21	—	6	—
Encyclopedic	26	33	1	7
Phonetic	7	15	1	2
Translation	5	1	1	—
Grammatic	10	139	13	—
Comp. word	6		—	—
Miscellaneous	8	—	—	—

compound word in Hungarian. Likewise, in the case of lábbal – kézzel (with hand – with foot) one principle is semantic; the two words are part of the same domain, but in the same time their inflection is also the same. The same is true for nyugaton – keleten (in the west – in the east).

The complexity of the communities and the diverse kinds of connections inside them is shown in an English example of the database (Figure 3).

On three larger communities we also counted the different kinds of connections inside the communities (Table 2).

Syntactic connections represent words of the same word class (in line with Levelt 1993). Since syntactic connections – same word class – are not the primary reason for naming an association, in Table 2 we separated syntactic connections from other associations to get a more consistent picture. In the analyzed databases, there were no associations which were only syntactic, meaning no two words were connected only because they were in the same word class. In all cases other connections (e.g., semantic) were also present.

As shown in Table 2, the types of associations governing a larger community may vary on the community, e.g., in the case of the community bank almost 13% of the associations were connected to encyclopedic knowledge, because the respondents named several banks (e.g., bank – CIB).

5 Discussion

5.1 Communities in free association data

Our results show that different types of associations are found in the investigated free association network, and small communities are organized by these types of associations. Some communities share semantic information, others share grammatical information too and again others have a similar phonetic structure. According to our findings, the existence of the following types of associative communities are proposed:
- Semantic communities consisting of the following connections:
  - synonymy (*trezor – széf*; *vault – safe*)
  - antonymy (*aktív – passzív*; *active – passive*)
  - collocations (*követendö – példa*; *example - to follow*; *good – and evil*)
  - domains (*hotdog – virsli; hotdog – sausage*)
  - pairs (*apa – anya; father – mother*)
  - associative communities (*eltévedés – GPS, get lost – GPS*)
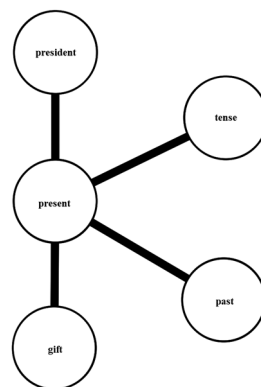- Phonetic communities (*pincér – cincér; present – president*)

Figure 3: An English community from the database. *Present* has different meanings (tense and gift) and both were named by subjects, *past* refers also to the domain of grammar, whereas *present* and *president* show phonetical resemblance.

Table 2: Different types of associations in the communities Health, Language and Bank

Community	Health		Languages		Bank	
Members	137		129		84	
Number of connections in the community	230		289		115	
Multiple connections – one syntactic	208		156		90	
Multiple connections – other than syntactic	16		63		34	
Types of associations	Number of items	%	Number of items	%	Number of items	%
Semantic	176	71.14	264	74.86	97	65.54
Encyclopedic	2	0.81	10	2.87	19	12.84
Phonetic	1	0.41	6	1.72	0	0
Translation	2	0.81	5	1.44	1	0.68
Grammatic	13	5.28	29	8.33	0	0
Associative	49	19.92	11	3.16	10	6.76
Compound word	4	1.63	27	7.76	20	13.51
Individual	0	0.00	0	74.71	2	1.35
Semantic detailed (semantic = 100%)	Number of items	%	Number of items	%	Number of items	%
- Synonymy	35	20.00	20	7.75	14	14.58
- Domain	104	59.43	192	74.42	63	65.63
- Pair	11	6.29	16	6.20	2	2.08
- Antonymy	1	0.57	2	0.76	0	0
- Collocation	6	3.43	4	1.55	5	5.21
- Categorizing	18	10.29	24	9.30	12	12.5
	Number of items		Number of items		Number of items	
Syntactic connections	208		156		90	

- Grammatic communities (*ültetni – fát; plant – a tree; kelt-fel; stood – up*)
- Syntactic communities (*gondoz – ápol; cherish – care*)
- Encyclopedic communities (*Mona Lisa – Leonardo, Harley – Davidson*).

We also found evidence for connections related to individual characteristics, created by life experiences, profession, gender, last experience etc. (testvér – Dezsö; brother – James). The individual connections were not present in the small communities found by modularity maximization, since in our analysis we considered only associations which were given by more than one/more than three respondents. Due to the relatively small number of respondents, there were just a few overlaps in clearly unique, personal associations, e.g., in associations regarding given names. These were present however, in the network with a connection strength of 1 (only one person gave a specific association), e.g., nagybátyám – Zoli (my uncle – Zoli), Petivel – szomszéd (with Peti – neighbour), szép – Eszter (beautiful – Eszter), szerelem – Eszter (love – Eszter) and unokatestvér – Viki (cousin – Viki). The existence of the individual connections is corroborated also by Siew and colleagues (2019), stating that aging process and personality also influence connections in the lexicon.

In this regard, some network structures seem to be shared by a smaller or larger group of individuals (e.g., speech community, speakers with the same experiences), whereas some of the structures and connections are present only in individuals resulting in unique associations.

These results show that besides the semantic relatedness of words, grammatical and syntactical structures also play a role in organizing the lexicon, building parts of sentences or grouping words inflected the same way. This is in line with the results of Stella (2019), who proved the existence of a multiplex lexical network by analyzing 30 months old children. He assumes that the multiplex network consists of a phonological and a semantic layer, the latter composed of a layer of free associations, of co-occurrences and of

feature sharing. The results of Stella (2019) and our work show that the mental lexicon contains more than simple semantic information, and that its structure is not only defined by semantic relations: other relations may also play a crucial role in the organization of the lexicon. Our result extends the results of De Deyne et al. (2016), who assume only lexico-semantic relationships in the lexicon. Results also show that assortativity is an organizing feature in the lexicon: similar words – having similar or the same meaning; similar form; similar inflection; similar pronunciation – tend to activate each other.

Our results show that communities can be kept together by different kinds of shared properties, and in some cases, communities share more than one type of characteristics. In this regard, relationships can be multidimensional: multiple kinds of connections can exist at the same time (pl. ázik – fázik; good – food). In addition, in some cases we observed that words from one register automatically prime a word of the same register. For example, the Hungarian words for sibling is testver; baty is the older brother, occs is the younger brother and hug is the younger sister. In the database, however, a cluster was made of the nickname versions of these words, so a cluster of teso – ocsi – hugi – batyus (sib – lil bro – lil sis – older bro) existed.

5.2 Interacting languages

Examining the data, we found that different languages also interact with each other, creating links between languages. The data shows the following kinds of interlanguage connections.
- Equivalence: hajó – ship – Schiff (hajó – "ship" (Hun) – ship (Eng) – Schiff "ship" (Ger))
- Phonetic: fut – foot (Hun. fut [fut] = "run").

Seeing languages as different, but connected networks, the translation process could be interpreted as the effort to match two networks, looking for connections between networks. In the interconnectedness of languages not only lexical meaning and equivalence but also cultural factors may play a crucial role (Szalay and Deese 1978).

In many cases, however, the networks of two languages cannot be matched directly when, e.g., grammatical structures differ (a structure does not exist in another language – e.g., Hungarian has no passive voice) or conceptual systems do not overlap (German Kreditfähigkeit and Kreditwürdigkeit are both translated with creditworthiness). Any situation where networks of the language systems differ may cause potential translation difficulties and/or problems.

Activated words in different languages in a free association task also put forward the question how activation is processed by bilinguals (De Bot and Schreuder 1993). Since we do not have a large multilingual database, we cannot give a general description of the networks and connections of bilinguals. However, the associations of the individuals are stored separately in the database, so by seeing the raw data we observed that some subjects tended to translate words (and give translations as associations) in the experiment. It is also known that several subjects were language teachers but only a few subjects translated the words – so we can formulate the assumption that the networks of different languages exist in bilinguals, but they are not connected the same way in each individual. Individual connection patterns emerge and make the networks of different languages easier or harder to access. Similar patterns – translations as associations – were also observed and reported by Szalay and Deese (1978).

Translations as associations are in line with De Bot (2004), who assumes the existence of language nodes in the lexicon, which control activation. We surmise, however, in contrast, that each lexical unit may itself carry the information of the languages it belongs to. The information is stored with the node (e.g., node x belongs to language L1), and this information connects the words of the same language, thereby creating a network of a given language, where all words belonging to L1 are stored in one network. At the same time, however, "false" connections and information are also allowed to be stored with the lexical units, resulting e.g., in activating words from another language or activating false friends.

Registers of a language may function the same way: information related to the register is stored with the lexical units, and the lexical units containing the same register are connected with each other.

All results are in line with the above described structure of lexical units of the mental lexicon (Bonin 2004, Klein 2015, Levelt 1993): information stored in lexical units is connected to information stored by other lexical units, thus enabling various kinds of connections (e.g., semantic, phonetic), but also multiple connections between the same units.

5.3 Possible implications for the layers of the lexicon

Our results showed that in a word association experiment small communities are governed by different types of associations.

Knowing that language and linguistic phenomena can be seen as a multiplex network (Vukić et al. 2020) and that lexical units of the lexicon store several types of information (Levelt 1993), we assume that the mental lexicon itself has a multiplex, multilayered network structure (cf. Stella et al. 2018).

In this structure different subnetworks (or layers of networks) may exist, which overlay each other and are responsible for priming and for the organizing structures of the lexicon. Based on the data of associations we assume that the findings of Stella and colleagues (2018) could be expanded with new layers of the lexicon. Assuming that the observed communities are in line with the network structure of the lexicon, we presume that these observed connections create a network, which has several layers storing different kinds of information (Figure 4, cf. e.g., Stella et al. (2018)).

These highly interconnected layers could be explanatory for several characteristics and processes of the lexicon, e.g., priming, individual associations, translation or even bilingual activation. Searching in the lexicon can thus mean searching for more resemblance and for assortativity of any kind. The assertive character of associations may result in more connections between layers, enabling shortcuts between layers (cf. Stella 2019).

As seen in the example of syntactic connections (connection according to same word class), they never stand alone. We may think of these connections as an additional governing force besides other types of associations. One possible way to explain this would be that the cognitive systems producing them work in a sequential way (e.g., layers activated one after another) and not in a parallel way (e.g., all layers active at once).

At the same time units are also assigned to a specific language, thus creating a network for the language. The language information could enable the layers of different languages to be connected to
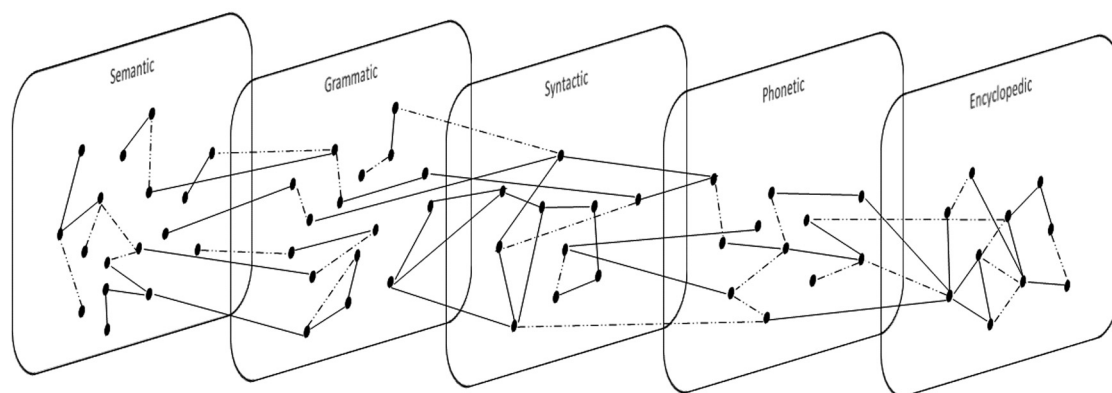

Figure 4: Possible layers of the lexicon. Dashed lines symbolize individual connections, while solid lines show connections shared by a group of speakers.
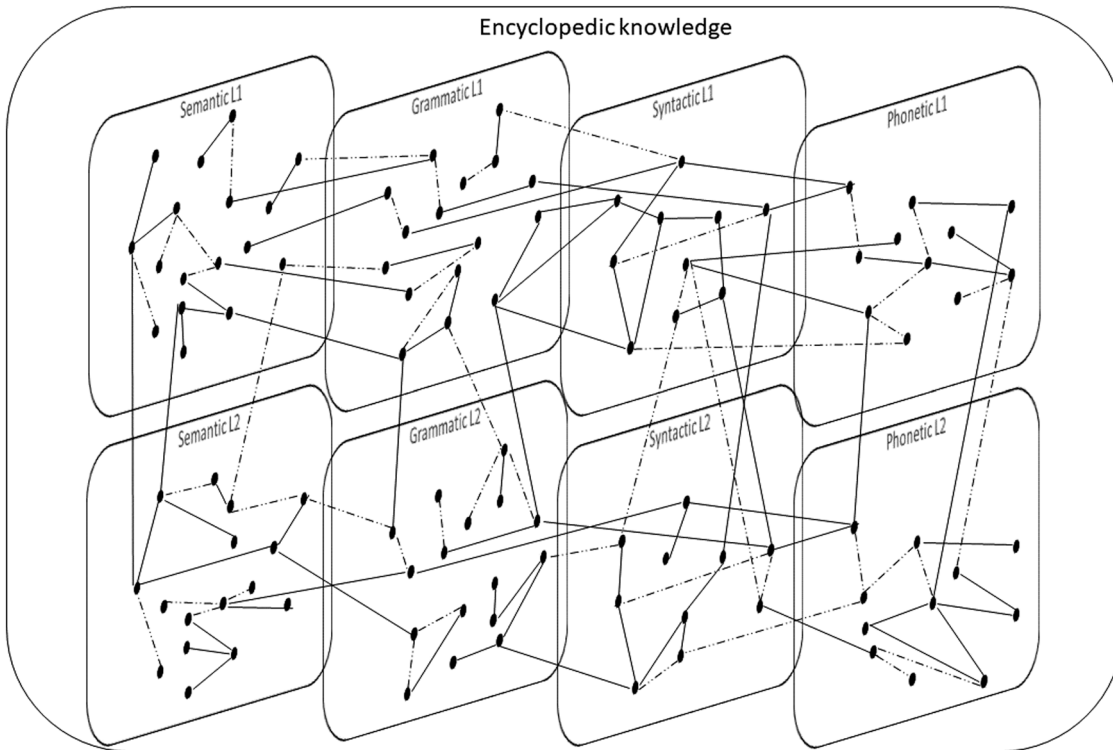
Figure 5: Different networks for different languages, where every language network consists of layers. Dashed lines symbolize individual connections, while solid lines show connections shared by a group of speakers.

each other, e.g., through equivalence (cf. Klein 2015). The between-language connections may play a role not only in translation processes but also by language learning (Mallikarjun et al. 2017).

The above described layers may exist in all the languages the subject speaks; and these language-specific multilayered networks may be connected to each other. In this regard, the individual network layer and the encyclopedic (knowledge) layer can be partly language-independent (e.g., the name Kevin – Costner or Mona – Lisa is the same in several languages), but in some cases the encyclopedic layer has also a language-specific component: Ärmelkanal (Ger) – La-Manche csatorna (Hun) – English Channel (Eng) (Figure 5).

All these layers and connections may not be static, but changing and evolving over time (Stella 2019, Siew et al. 2019)

The assumption of the described multilayered network of the lexicon is however, based only on a single word association experiment. It is not clear whether our results describe only the structure of associations, or the structure of the lexicon. It is also not clear whether we would find the same layers by other research designs or other tasks. Also the question remains open, whether these layers must be seen as different layers of one network or they are completely different networks existing simultaneously.

6 Conclusions

Examining the community structure of a Hungarian word association database we presented results for different kinds of associative connections and communities in the mental lexicon, where different connections coexist and organize the lexicon.

We analyzed the communities with two community detection algorithms and found evidence that communities can display at least two different properties together, e.g., synonyms were inflected the

same way. Based on the found communities we argued that communities in the lexicon are organized by various types of connections and that communities and connections represent in some cases multiple types of linguistic similarities at once.

The research of the networked structure of the lexicon has just begun – combined with characteristics of other human and artificial networks, this approach may prove extremely valuable for understanding not only the lexicon, but also the organization of human language itself.

References

Aicher, Christopher, Jacobs Abigail Z., and Clauset Aaron. 2015. "Learning latent block structure in weighted networks." *Journal of Complex Networks* 3(2): 221–48. doi: 10.1093/comnet/cnu026.

Aitchison, Jean. 1987. *Words in the mind*. Oxford: Basil Blackwell. doi: 10.1002/acp.2350030209.

Barabási, Albert-László. 2016. *Network science*. Cambridge: Cambridge University Press.

Barthélemy, Mark. 2011. "Spatial networks." *Physics Reports* 499(1): 1–101.

Baronchelli, Andrea, Ramon Ferrer-i-Chancho, Romualdo Pastor-Satorras, Nick Chater, and Morten H. Christiansen. 2013. "Networks in cognitive science." *Trends in Cognitive Sciences* 17(7): 348–60. doi: 10.1016/j.tics.2013.04.010.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 P10008. doi: 10.1088/1742-5468/2008/10/p10008.

Boccaletti, Stefano et al. 2014. "The structure and dynamics of multilayer networks." *Physics Reports* 544(1): 1–122. doi: 10.1016/j.physrep.2014.07.001.

Bonin, P. 2004. Introduction. In *Mental Lexicon: "Some Words to Talk about Words"*, ed. P. Bonin, (p. vii–xi.). New York: Nova Sciencia Publishers.

Bóta, András and Miklós Krész. 2015. "A high resolution clique-based overlapping community detection algorithm for small-world networks." *Informatica* 39(2): 177–87.

Bovasso, Gregory, Lorand Szalay, Vincent Biase, and Matthew Stanford. 1993. "A graph theory model of the semantic structure of attitudes." *Journal of Psycholinguistic Research* 22(4): 411–25. doi: 10.1007/bf01074344.

Bramoullé, Yan, Andrea Galeotti, and Brian W. Rogers (eds.). 2016. *The oxford handbook of the economics of networks*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199948277.001.0001.

Brookes, Gavin and Tony McEnery. 2019. "The utility of topic modelling for discourse studies: A critical evaluation." *Discourse Studies* 21(1): 3–21. doi: 10.1177/1461445618814032.

Castro, Nichol and S. Q. Cynthia Siew. 2020. "Contributions of modern network science to the cognitive sciences: revisiting research spirals of representation and process." In *Proceedings of the Royal Society A*. 476: 20190825.

Castro, Nichol, Kristin M. Pelczarski, and Michael S. Vitevitch. 2017. "Using network science measures to predict the lexical decision performance of adults who stutter." *Journal of Speech Language and Hearing Research* 60(7): 1–8. doi: 10.1044/2017_jslhr-s-16-0298.

Citraro, Salvatore and Rossetti Giulio. 2020. "Identifying and exploiting homogeneous communities in labeled networks." *Applied Network Science* 5:55. doi: 10.1007/s41109-020-00302-1.

Collins, Allan M. and Elizabeth F. Loftus. 1975. "A spreading-activation theory of semantic processing." *Psychological Review* 82(6): 407–28. doi: 10.1016/b978-1-4832-1446-7.50015-7.

Cong, Jin and Liu Haitao. 2014. "Approaching human language with complex networks." *Physics of Life Reviews*. 11(4): 598–618. doi: 10.1016/j.plrev.2014.04.004.

Coronges, Kathryn A., Alan W. Stacy, Thomas W. Valente. 2007. "Structural comparison of cognitive associative networks in two populations." *Journal of Applied Social Psychology* 37(9): 2097–129. doi: 10.1111/j.1559-1816.2007.00253.x.

Cramer, Phebe. 1968. *Word association*, London: Academic Press.

De Bot, Kees. 2004. "The multilingual lexicon: modelling selection and control." *International Journal of Multilingualism* 1(1): 17–32. doi: 10.1080/14790710408668176.

De Bot, Kees and Robert, Schreuder. 1993. "Word production and the bilingual lexicon." In *The Bilingual Lexicon*, eds. Schreuder Robert and Weltens Bert, (p. 191–214). Amsterdam/Philadelpiha: John Benjamins. doi: 10.1075/sibil.6.

De Deyne, Simon and Gert Storms. 2008. "Word associations: Network and semantic properties." *Behavior Research Methods* 40(1): 213–31. doi: 10.3758/brm.40.1.213.

De Deyne S., Kenett Y. N., Anaki, D., Faust M. and Navarro D. 2017. "Large-scale network representations of semantics in the mental lexicon.". In *Frontiers of cognitive psychology. Big data in cognitive science*, ed. M. N. Jones, (p. 174–202). London-New York: Routledge.

De Deyne, Simon, Steven Verheyen and Gert Storms. 2016. "Structure and organization of the mental lexicon: a network approach derived from syntactic dependency relations and word associations." *Understanding Complex Systems* 99:47–79. doi: 10.1007/978-3-662-47238-53.

Easley, David and Jon Kleinberg. 2010. *Networks, crowds and markets: reasoning about a highly connected world*. Cambridge: Cambridge University Press.

Figueroa, G. Jesus, Esther G. Gonzalez, and Victor M. Solis. 1976. "An approach to the problem of meaning: Semantic networks." *Journal of Psycholinguistic Research* 5(2): 107–15. doi: 10.1007/bf01067252.

Fortunato, Santo. 2010. "Community detection in graphs." *Physics Reports* 486(3–5): 75–174. doi: 10.1016/j.physrep.2009.11.002.

Fragkiskos, Malliaros D. and Michalis Vazirgiannis. 2013. "Clustering and community detection in directed networks: A survey." *Physics Reports* 533(4): 95–142. doi: 10.1016/j.physrep.2013.08.002.

Galton, Francis. 1879. "Psychometric experiments." *Brain* 2:149–62. doi: 10.1037/10913-022.

Galton, Francis. 1883. *Inquires into human faculty and its development*. London: Macmillan. doi: 10.1037/10913-000.

Goldstein, Rutherford and Michael S. Vitevitch. 2014. "The influence of clustering coefficient on word-learning: How groups of similar sounding words facilitate acquisition." *Frontiers in Psychology* 5:1307. doi: 10.3389/fpsyg.2014.01307.

Goldstein, Rutherford and Michael S. Vitevitch. 2017. "The influence of closeness centrality on lexical processing." *Frontiers in Psychology* 8:1683. doi: 10.3389/fpsyg.2017.01683.

Gravino, Pietro, Vito Servedio, Alain Barrat, and Vittorio Loreto. 2012. "Complex structures and semantics in free word association." *Advances in Complex Systems* 15(3–4): 1250054–1. doi: 10.1142/s0219525912500543.

Griffiths, Thomas L., Mark Steyvers and Alana Firl. 2007. "Google and the mind: Predicting fluency with PageRank." *Psychological Science* 18:1069–76. doi: 10.1111/j.1467-9280.2007.02027.x.

Hajdu, László, Miklós Krész, and András Bóta. 2018. "Community based influence maximization in the independent cascade model." *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems*. Poznań IEEE. Vol. 15, (p. 237–43). doi: 10.15439/2018f201.

Ibbotson, Paul, Vsevolod Salnikov, and Richard Walker. 2019. "A dynamic network analysis of emergent grammar." *First Language* 39(6): 652–80. doi: 10.1177/0142723719869562.

Jackendoff, Ray. 2002. *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198270126.001.0001.

Jackson, Matthew O, 2008. *Social and economic networks: models and analysis*. Princeton-Oxford: Princeton University Press.

Jingyang, Jiang, Wuzhe Yu, and Haitao Liu. 2019. "Does scale-free syntactic network emerge in second language learning?." *Frontiers in Psychology* 10:925. doi: 10.3389/fpsyg.2019.00925.

Ke, Jinyun and Yao Yao. 2008. "Analysing language development from a network approach." *Journal of Quantitative Linguistics*. 15(1): 70–99. doi: 10.1080/09296170701794286.

Klein, Wolfgang. 2015. "Lexicology and lexicography." In *International encyclopedia of the social & behavioral sciences*, ed. James Wright, second edition. Vol 13. Amsterdam: Elsevier. doi: 10.1016/b0-08-043076-7/03019-9.

Kovács, László. 2013. *Conceptual systems and lexical networks in the mental lexicon,* Budapest: Tinta (In Hungarian: Fogalmi rendszerek és lexikai hálózatok a mentális lexikonban).

Kovács, László. 2019. *Brand and brand name*. Budapest: Tinta. (In Hungarian: Márka és márkanév).

Lancichinetti, Andrea, Santo Fortunato and János Kertész. 2009. "Detecting the overlapping and hierarchical community structure in complex networks." *New Journal of Physics* 11(3): 033015. doi: 10.1007/978-3-642-31821-4_2.

Levelt, Willem J. M. 1993. *Speaking: from intention to articulation*. Cambridge-London: MIT Press.

Liu, Haitao and Jin Cong. 2014. "Empirical characterization of modern chinese as a multi-level system from the complex network approach." *Journal of Chinese Linguistics* 42(1): 1–38.

Liu, Haitao and Xu Chunshan. 2011. "Can syntactic networks indicate morphological complexity of a language?" *Europhysics Letters Volume* 93(2): 28005. doi: 10.1209/0295-5075/93/28005.

Loscalzo, Joseph, Albert-László Barabási, and Edwin K.and Silverman (eds.) 2017. *Network medicine*. Cambridge-London: Harvard University Press.

Mallikarjun, Amritha, Rochelle S. Newman, and Jared M. Novick. 2017. "Exploiting the interconnected lexicon: Bootstrapping english language learning in young Spanish speakers." *Translational Issues in Psychological Science* 3(1): 34–47. doi: 10.1037/tps0000105.

Mandelbaum, Eric. 2017. "Associationist theories of thought." In *The stanford encyclopedia of philosophy* (*Summer 2017 Edition*), ed. Edward N. Zalta https://plato.stanford.edu/archives/sum2017/entries/associationist-thought (25.06.2020).

Martinčić-Ipšić, Sanda, Margan Domagoj, and Ana Meštrović. 2016. "Multilayer network of language: A unified framework for structural analysis of linguistic subsystems." *Physica A* 457:117–28.

Meara, Paul. 2009. *Connected words*. Amsterdam & Philadelphia: John Benjamins.

Mehler, Alexander, Andy Lücking, Sven Banisch, Philippe Blanchard and Barbara Job (eds.) 2016. *Towards a theoretical framework for analyzing complex linguistic networks*. Berlin-Heidelberg: Springer.

Menczer, Filippo, Santo Fortunato, and Clayton A. Davis. 2020. *A first course in network science*. Cambridge: Cambridge University Press.

Milroy, Lesley. 1987. *Language and social networks*. Oxford: Blackwell.

Nelson, Douglas L., Cathy L. McEvoy, and Thomas A. Schreiber. 2004. "The university of south florida word association, rhyme, and word fragment norms." *Behavior Research Methods, Instruments & Computers* 36:402–7. doi: 10.3758/BF03195588.

Newman, Mark E. J. 2003. "The structure and function of complex networks." *SIAM Review* 45(2): 167–256. doi: 10.1016/S0010-4655(02)00201-1.

Newman, Mark E. J. 2006. "Modularity and community structure in networks." *Proceedings of the National Academy of Sciences USA* 103(23): 8577–82. doi: 10.1073/pnas.0601602103.

Newman, Mark E. J. and Michelle Girvan. 2004. "Finding and evaluating community structure in networks." *Physical Review E* 69(2): 026113. doi: 10.1103/physreve.69.026113.

Peixoto, Tiago P. 2014. "Efficient Monte Carlo and greedy heuristic for the inference of stochastic block models." *Physical Review E* 89(1): 012804. doi: 10.1103/physreve.89.012804.

Quillian, Ross M. 1968. "Semantic memory." In *Semantic information processing*, ed. M. Minsky, (p. 227–70). Cambridge: MIT Press. doi: 10.21236/ad0641671.

Rogers, Timothy T. and James L. McClelland. 2003. "The parallel distribued processing approach to semantic cognition." *Nature Reviews Neuroscience*. 4:310–22. doi: 10.1038/nrn1076.

Rogers, Timothy T. and James L. McClelland. 2004. *Semantic cognition*. Cambridge: MIT Press.

Schur, Ellen. 2007. "Insights into the structure of L1 and L2 vocabulary networks: intimations of small worlds." In *Modelling and Assessing Vocabulary Knowledge*, eds. H. Daller, J. Milton and J. Treffers-Daller, (p. 182–203). Cambridge: Cambridge University Press. doi: 10.1017/cbo9780511667268.013.

Siew, Cynthia S. Q. 2013. "Community structure in the phonological network." *Frontiers in Psychology* 4:553. doi: 10.3389/fpsyg.2013.00553.

Siew, Cynthia S. Q. and Michael S. Vitevitch. 2019. "The phonographic language network: using network science to investigate the phonological and orthographic similarity structure of language." *Journal of Experimental Psychology: General* 148(3): 475–500. doi: 10.1037/xge0000575.

Siew, Cynthia S. Q., Dirk U. Wulff, Nicole M. Beckage, and Yoed N. Kenett. 2019. "Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics." *Complexity* 2019(5915): 1–24. doi: 10.31234/osf.io/eu9tr.

Solé, Richard, Bernat Corominas-Murtra, Sergi Valverde, and Luc Steels. 2010. "Language networks: Their structure, function, and evolution." *Complexity* 15(6): 20–6. doi: 10.1002/cplx.20305.

Stella, Massimo. 2019. "Modelling early word acquisition through multiplex lexical networks and machine learning." *Big Data and Cognitive Computing* 3(1): 10. doi: 10.3390/bdcc3010010.

Stella, Massimo, Nicole M. Beckage, and Markus Brede. 2017. "Multiplex lexical networks reveal patterns in early word acquisition in children." *Scientific Reports* 7:46730.

Stella, Massimo, Nicole M. Beckage, Markus Brede, and Manlio De Domenico. 2018. "Multiplex model of mental lexicon reveals explosive learning." *Scientific Reports* 8:2259. doi: 10.1038/s41598-018-20730-5.

Steyvers, Mark and Joshua B. Tenenbaum. 2005. "The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth." *Cognitive Science* 29:41–78. doi: 10.1207/s15516709cog2901_3.

Szalay, Lóránd B. and James E. Deese. 1978. *Subjective meaning and culture: An assessment through word associations*. Hillsdale: Lawrence Erlbaum Associates.

Thedchanamoorthy, Gnana, Mahendra Priaveenan, Dharshana Kasthuriratna and Upul Senanayake. 2014. "Node assortativity in complex networks: An alternative approach," *Procedia Computer Science* 29: (p. 2449–61). doi: 10.1016/j.procs.2014.05.229.

Van Rensbergen, Bram, Gert Storms, and Simon De Deyne. 2015. "Examining assortativity in the mental lexicon: Evidence from word associations." *Psychonomic Bulletin & Review* 22(6): 1717–24. doi: 10.3758/s13423-015-0832-5.

Vitevitch, Michael S. 2008. "What can graph theory tell us about word learning and lexical retrieval?" *Journal of Speech, Language, and Hearing Research* 51:408–22. doi: 10.1044/1092-4388(2008/030).

Vitevitch, Michael S. (ed.) 2020a. *Network Science in Cognitive Psychology*. New York-London: Routledge.

Vitevitch, Michael S. 2020b. "Introduction." In *Network Science in Cognitive Psychology*, ed. M.S. Vitevitch, (p. 1–9). New York-London: Routledge.

Vitevitch, Michael S., Rutherford Goldstein, Cynthia Siew, and Nichol Castro. 2014. "Using complex networks to understand the mental lexicon." *Yearbook of the Poznań Linguistic Meeting* 1:119–38. doi: 10.1515/yplm-2015-0007.

Vukić, Durdica, Sanda Martinčić-Ipšić, and Ana Meštrović. 2020. "Structural analysis of factual, conceptual, procedural, and metacognitive knowledge in a multidimensional knowledge network." *Complexity* 2020:1–17. doi: 10.1155/2020/9407162.

Wu, Zhi-Hao, Lin You-Fang, Steve Gregory, Huai-Yu Wan, and Sheng-Feng Tian. 2012. "Balanced multi-label propagation for overlapping community detection in social networks." *Journal of Computer Science and Technology* 27(3): 468–79. doi: 10.1007/s11390-012-1236-x.