

# Network Analysis of Korean Word Associations

**Jaeyoung Jung**

Tokyo Institute of Technology  
2-12-1, O-okayama, Meguro-ku  
Tokyo, 152-8552, Japan  
jung.j.aa@m.titech.ac.jp

**Li Na**

Tokyo Institute of Technology  
2-12-1, O-okayama, Meguro-ku  
Tokyo, 152-8552, Japan  
li.n.ad@m.titech.ac.jp

**Hiroyuki Akama**

Tokyo Institute of Technology  
2-12-1, O-okayama, Meguro-ku  
Tokyo, 152-8552, Japan  
akama.h.aa@m.titech.ac.jp

## Abstract

Korean Word Associations (KorWA) were collected to build a semantic network for the Korean language. A graphic representation approach of applying coefficients to complex networks allows us to discern the semantic structures within words. A semantic network of the KorWA was found to exhibit the scale-free property in its degree distribution. The growth of the network around hub words was also confirmed through two experimental phases. As an issue for further research, we suggest that the present results may yield insights for computational neurolinguistics, as a semantic network of word association norms can bridge the gap between information about lexical co-occurrences derived from a corpora and anatomical networks as a basis for mapping out neural activations.

## 1 Introduction

Language is an intricate cognitive system. The mental system, called a grammar by linguists, allows human beings to form and interpret the sounds, words, and sentences of their language. The system is often broken down into several components, such as phonetics, phonology, morphology, syntax, and semantics (O'Grady et al. 2005). Depending on one's concerns, the basic elements of each level (i.e. phones, syllables, morphemes, words, or sentences) become the constituents of linguistic networks of sound patterns, morphological structures, or syntactic organizations. Parse trees, for instance, which are often used in analyzing the syntactic structures of sentences, employ links to represent the *syntagmatic* relationships between words. However, focusing

on the processes of conceptualizing feelings, experiences, and perceptions and of encoding them in words (namely, lexicalization), linguists have frequently drawn another kind of linguistic network substantiated as a map of words projecting semantic structures and relations onto an Euclidian space from a *paradigmatic* perspective. In that sense, word association data is attractive in terms of ease of data manipulation, especially when making a graph from a list of word pairs. Moreover, the tools for analyzing complex networks have been often applied to analyzing the structural features within large-scale word association data and to mining lexical knowledge from them.

Since Galton (1880), word association has been used as an empirical method for observing thought processes, memory, and mental states within clinical and cognitive psychology (Deese, 1965). From a linguistic perspective, word associations are undoubtedly valuable language resources because they are rich sources of linguistic knowledge and lexical information. The data has some unique characteristics that are very interesting and useful for cultural studies, reflecting the life styles, social, cultural and linguistic backgrounds of the native speakers who contributed to the data collections. Such information could be particularly useful for further applications not only within semantic studies but also for intelligent information retrieval, brain research, and language learning.

In short, so-called word association norms are crucial as large-scale paradigmatic corpora. They consist of word pair data based on psychological experiments where the participants are typically asked to provide a semantically-related response word that comes to mind upon presentation of a stimulus word. Two well-known word association data for English are the University of South Florida word association, rhyme and word fragment norms

(Nelson et al., 1998) and the Edinburgh Word Association Thesaurus of English (EAT; Kiss et al., 1973). For Japanese there are Ishizaki's Associative Concept Dictionary (IACD) (Okamoto and Ishizaki, 2001) and the Japanese Word Association Database (JWAD) (Joyce, 2005, 2006, 2007). Utilizing computational linguistic techniques that aim to mathematically analyze their structures, raw association data is often transformed into some form of graph or complex network representation, where the vertices stand for words and the edges indicate an associative relationship (Miyake et al., 2007). Such techniques of graph representation and their analysis allow us to discern the patterns of connectivity within large-scale resources of linguistic knowledge and to perceive the inherent relationships between words and word groups.

However, despite a long history of word association studies and the valuable contributions of such data to cognitive science, comprehensive, large-scale databases of Korean word association norms have been seriously inadequate. In one study, Lee (1970) surveyed word associations based on 30 adjectives and 29 words representing colors, targeting 40 university students and analyzed the response words for associative tendencies in terms of gender and grammatical word classes. More recently, Shin (1998) attempted to categorize words by conceptual systems in order to construct a lexical dictionary supporting foreign language learners. Although her data differs from word association norms and is not available as an accessible digital database for academic purpose, the semantic classification of the words can be exploited in complementing the analysis of Korean semantic networks.

A collection of Korean word associations (for short, KorWA) was planned and conducted with the strong motivation of constructing a worthwhile database of Korean word associations as a kind of resource that has multiple applications in a number of areas such as lexicography, language education, artificial intelligent, natural language processing, and cultural study. Moreover, we intend to share the database on the web to foster these various potential utilities.

In this paper, KorWA is represented into semantic networks and examined by some combinatorial methods in linguistics. The details are presented from the whole process of collecting the data to the results of the analysis based on the the-

ory of complex networks. Furthermore, this paper briefly discusses another important characteristic, dynamics in scale-free networks, which has recently attracted much attention in this research field. Finally we will mention the applicability of the graph-based analysis developed here to the future potential researches of the computational neurolinguistics.

## 2 Korean word associations

### 2.1 Design of Experiment

Preparation of an association experiment begins with the selection of a stimulus word set that is to be presented to the respondent in order to initiate their association process. Determining the stimulus word set is a crucial part in designing the experiment, as associative responses are greatly influenced by the characteristics of the presented words, in particular, the stimulus word familiarity influences response heterogeneity, variability, relational categories, and reaction times (Deese 1965). For the experiment of Korean word associations, we referred to a list of 5,000 Korean basic words (Seo et al. 1998), which was derived from the Yonsei Corpora consisting of 42,644,891 words as of 1998 [ILIS]. From the list, we compiled a list of 3,951 words, consisting of 2,628 nouns, 1,006 verbs, and 317 adjectives.

One hundred and thirty-two native Korean students (71 males and 61 females) at Daejeon University, South Korea voluntarily participated in the experiment. The students were mainly from the departments of Korean language and literature (54%), physical therapy (30%), and philosophy (11%). More than 70% of the students had educational background in the humanities. Most of the students (93%) were in their 20s; 82% between 20 and 25 years old and 11% between 26 and 30 years old. 14% of students answered that on average they read more than five books in a month. The task was conducted on the campus of Daejeon University from September 2007 to February 2008. It was a traditional pen-and-paper based task. Under the control of an instructor, each session of the task lasted for 30 minutes.

In the task, participants were instructed to write down the response words that came to mind when they looked at the presented words. We asked the subjects to write down all the words that they

could think of from the presented words. This procedure is called the continuous free association task, differing from the discrete association task where the subject is asked to only write down their first response (Cramer 1968).

As a means of naturally displaying continuous associations, the respondents were asked to map out their responses. That is, they drew a kind of associative map for a given word, by adding a line when they made an association and numbering the responses according to the order in which they came to mind from the stimulus word. In the experiment, an A5 size booklet was distributed to the participants. The booklet had 66 pages printed on one-side, including 2 front-back cover pages.

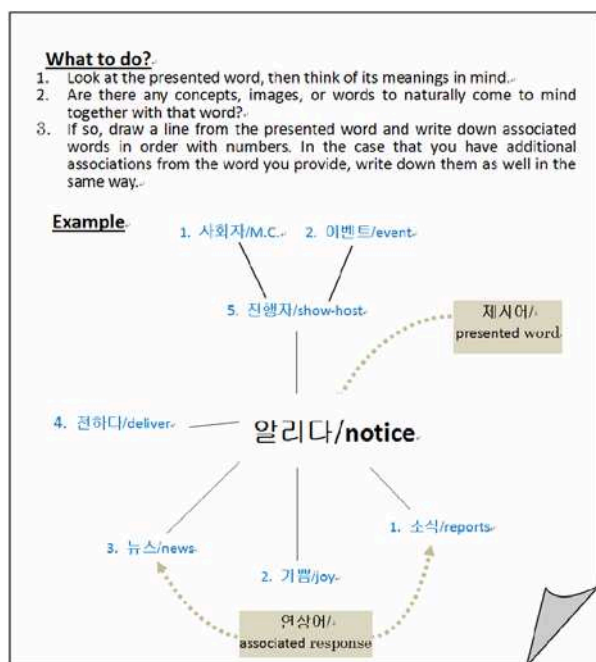


Figure 1. Instructions about the task and example

The first 4 pages contained instructional information; (1) a brief description of the experiment's purpose and its method, (2) a short survey for basic respondent information (gender, age, major, and number of books read in a month), (3) an example illustrating what to do in the task (shown in Figure 1), and (4) one practice before the task. Then, the remaining 60 pages were for the word association task, printed with one word per page. Thus, each participant was asked to provide word association responses for 60 words.

In total, 132 booklets were prepared for the task. A list of 60 words for each booklet was randomly

extracted from the 3,951 stimulus set. Apart from six of the 132 sets, the lists included 40 nouns, 15 verbs, and 5 adjectives. The others had slightly different numbers of the syntactic categories. However, eventually each stimulus word was planned to be presented to up to two subjects.

As the result of approximately 6-month period of data collection, we obtained 28,755 responses in total for the 3,942 stimulus words (from the original stimulus set, nine words failed to elicit any word associations). The 28,755 responses (tokens) consisted of 11,275 distinct words (types). Each item was presented to two respondents.

The KorWA database (Figure 2) was constructed from the collected word association responses. The data is arranged into six fields; (1) the part of speech of the stimulus word, (2) the stimulus word, (3) the part of speech of the response, (4) response order, (5) raw form of the response, and (6) response word in standard form.

Stimulus word		Response word			
#1	#2	#3	#4	#5	#6
part of speech	stimulus	part of speech	order	raw form	dictionary form
1	과목	1	1	학교	학교
1	과목	1	1	선생님	선생님
1	과목	1	2	수학	수학
1	과목	1	2.1	짜증	짜증
1	즐음	1	1	열정	열정
1	즐음	1	2	도전	도전
1	아버지	1	1	사랑	사랑
1	아버지	1	2	어머니	어머니
1	아버지	1	2.1	사랑	사랑
1	아버지	1	2.2	논란	논란
1	열매	1	1	성공	성공
1	열매	1	1.1	열쇠	열쇠
1	열매	1	1.2	노력	노력
1	열매	1	2	사과	사과
1	열매	1	2.1	과일	과일
1	열매	1	2.2	비타민	비타민
1	목적	1	1	지향점	지향점
1	목적	1	1.1	목표	목표
1	목적	2	2	이루다	이루다

Figure 2. Contents of the KorWA database

## 2.2 Basic Analysis

The Korean word association data collected is briefly summarized here in terms of the relations between the stimulus words and the responses together with some basic statistics. The participants produced on average 218 responses (standard deviation = 63.8, ranging from 98 to 482) for the complete set of 60 words in the free association task, which corresponds to 3.6 responses per stimulus word. Because each stimulus item was presented to two respondents, each stimulus has on average 7.3 association responses.

As already mentioned, our task was the continuous free association task where the respondent was allowed to provide more than one response, so there is the possibility of chaining responses where some association responses are elicited by prior responses. The response set of 28,755 tokens includes all such responses. Furthermore, it is possible to extract the primary responses that were given as the first response produced for each stimulus word. In doing that, it is possible to convert the continuous free association task condition to the discrete association task employed in other existing data. The primary associates for the 3,942 stimulus are 7,550 word tokens (4,197 types). The associations seem to be related to the grammatical classes of the stimulus. Ervin (1961) reports that many associations tend to have the same grammatical class as the stimulus word. Similarly, Jenkins (1954) and Saporta (1955) provide an interesting way of classifying association structures into two modes, i.e. *paradigmatic* associations and *syntagmatic* ones. In the former mode, the stimulus and response fit a common grammatical paradigm. For example, the word ACTION yields the associates of WORDS, LIFE, MOVEMENT, MOTION, GAME, and so on, which are not likely to occur as sequences in everyday English. In the latter case, the stimulus and response are generally contiguous, occupying different positions within phrases or sentences. Namely, they often form sequences, as in the relations between the stimulus word of ADMINISTRATIVE and its common associates of DUTY, JOB, CONTROL, DISCIPLINE, POWER, BUREAUCRATS, POSITION, AGENCY, ENTITY, SCHOOL, BOSS, GOVERNMENT, RULE, etc. (Deese 1965). Deese (1962) clarified the relative frequencies of paradigmatic and syntagmatic associations among the grammatical classes of English, especially with nouns, verbs, adjectives, and adverbs in his study. He observed that the tendency towards paradigmatic or syntagmatic association varied with word class; nouns are dominantly paradigmatic, while adjectives and verbs tend to be both paradigmatic and syntagmatic. In the case of adjectives, it is a particularly interesting tendency for the association types to have a strong correlation with frequency of usage. That is to say, for common adjectives, associations are more likely to be paradigmatic (e.g. for HOT, associates such as COLD, WARM, and COOL

more frequently occur than WOMEN, WEATHER, and the like), while uncommon adjectives are more syntagmatic (e.g. for ADMINISTRATIVE, associates such as DUTY, GOVERNMENT, and RULE are more often produced than SUPERVISORY, EXECUTIVE, and so on). What is more, most paradigmatic associates to adjectives are either synonymous with the stimulus (COLD-COOL) or the opposite of the stimulus (COLD-HOT). Common adjectives overwhelmingly have more antonyms as their response, but relatively low-frequency adjectives have more synonym associations.

A similar tendency is observed in our data, which included three types of grammatical class among the stimulus items, with nouns, verbs, and adjectives, covering 66.5%, 25.5%, and 8% of the stimulus set respectively. The different proportions of the word classes reflects their frequencies within the Yonsei corpora, i.e. among the 5,000 most frequent words, there is a much larger number of nouns, compared to verbs and adjectives. By tagging the responses with parts of speech data during the course of constructing the database, we can analyze the distributions of grammatical categories among the responses. The responses were overwhelmingly nouns (78%), followed by adjectives (7%), proper nouns (4.5%) and verbs (4.4%) in descending order. Within the primary response list, the distributions of word class are not greatly different, with 79% nouns, 6.7% adjectives, 4.8% verbs, 3.9% proper nouns, and around 6% others.

Corresponding to the grammatical class of the stimulus specifically, nouns are also the dominant responses. When considering just the primary responses, noun stimulus elicited mostly noun responses (80%), followed by adjectives (6%), proper nouns (5%), and verbs (3%); verb stimulus produced around 80% noun associates, 10% verbs and 4% adjectives; while for adjective stimulus, there were 70% noun responses, 19% adjectives, and 2% verbs. In short, we found a majority of noun-noun, verb-noun, and adjective-noun combinations within the stimulus-response relations. This demonstrates the association tendency for nouns to strongly elicit paradigmatic associations, as seen from the principal noun-noun relations, while verbs and adjectives tend to yield more syntagmatic associations, as seen from the major relations of verb-noun and adjective-noun.

### 2.3 Network Analysis (1)

**Degrees:** Recently, a number of studies have applied graph theory approaches in investigating linguistic knowledge resources. For instance, instead of word frequency based computations, Dorow, et al (2005) utilize graph clustering techniques as methods of detecting lexical ambiguity and of acquiring semantic classes. Steyvers and Tenenbaum (2005) conducted a noteworthy study that examined the structural features of three semantic networks (free association norms of Nelson et al., Roget's thesaurus, and WordNet). By calculating a range of statistical features, including the average shortest paths, diameters, clustering coefficients, and degree distributions, they observed interesting similarities between three networks in terms of their *scale-free* patterns of connectivity and *small-world* structures. Following their basic approach, we analyze the characteristics of the semantic network representation of KorWA by calculating the statistical features of the graph coefficients, such as degree and degree distribution.

The semantic network representation of the word association network is constructed by representing the words as nodes and associative pairing information for words as edges. The degree ( $D$ ) of a node denotes the number of edges that a node has. An undirected graph is structured by the edges, while a directed graph is structured by arcs that include the associative direction. The numbers of incoming and outgoing arcs from a node are referred to as the in-degree and out-degree of a node, respectively. The sum of the in-degree and out-degree values of a node is equal to its total degree.

This concept of graph analysis allows us to categorize the total words in the data into three types; one being words only found in the stimulus set (S-type), one being words occurred as both stimulus and responses (SR-type), and the last being words only observed among the response set (R-type). The proportion of S-type, SR-type, and R-type words in the total word set corresponds to 12.2% (1,568 words), 18.5% (2,374 words), and 69.3% (8,901 words) respectively. Here, it is worth focusing on the SR-type of words. These are words selected as the most frequent ones through a large-scale corpus covering various fields. At the same time, they also are produced by people in the free association task. This may indicate, in some sense, the high usability or commonness of those words. Indeed, the most frequent words in this data all belong to the SR-type.

## 2.4 Network Analysis (2)

**Scale-free:** The most frequent words belonging to the SR-type play the role of hubs in semantic networks made from word association data. These hubs can be represented as nodes that have not only outgoing links but also possess ingoing links, which leads us think of a scale-free graph, such as that incorporated within the Barabási-Albert (BA) model. It is widely known that Barabási and Albert (1999) have suggested that the degree distributions of scale-free network structures correspond to a power law, expressed as  $P(x = d) = d^{-\gamma}$  (where  $d$  stands for degree and  $\gamma$  is a small integer, such as 2 or 3). This type of distribution is also known as Zipf's law, which describes the typical frequency distributions of words in a document and plots on a log scale as a falling diagonal stroke. The degree distribution of nodes in the KorWA network also exhibits this scale-free property, which has also been observed in word association data for different languages.

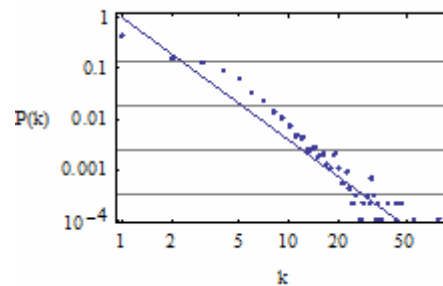


Figure 3. Degree distribution on log-log scales for the KorWA semantic network.  $P(k)$  is the probability that a node has  $k$  degrees in the network.

However, we should stress the importance of network dynamics and of microscopically examining the ongoing process of data accumulation to determine whether the scale-freeness observed for word association data is derived from the same mechanism as the BA model. Rather than being static, networks are recognized as evolving over time, with the adding or pruning of nodes and edges (Barabási and Albert, 1999; Watts, 1999). Indeed, we can easily identify such networks in a number of areas, from the World Wide Web to the internet connections on a physical level, co-authorships, friendships, and business transactions.

According to the BA model, the probability that a node receives an additional link is proportional to its degree. The probability that a new vertex will be connected to a vertex (node)  $i$  depends on the connectivity of that vertex. Barabási and Albert (1999) explain with this idea of *preferential attachment* in terms of the scale-free property and the presence of hubs within the network. Networks as dynamical systems which grow over time and have topological properties produce dynamical behaviors as well. In particular with research on the diffusion of a new trend or technology or the spread of a disease and virus, the structural properties of the network have presented a new approach to understanding epidemical behaviors over a network, including issues about why contagion occurs in certain cases, how it spreads, and what is the most efficient and effective way to prevent it. Many researchers have tried to address and analyze such behaviors with small-world models (Ball et al., 1997; Watts and Strogatz, 1998) and scale-free models (Pastor-Satorras and Vespignani, 2001).

The semantic networks that we have examined to date have similar structural properties to many other networks. So, it is also possible to explain the scale-free feature of semantic networks in terms of preferential attachment? How can such dynamic behavior be interpreted for semantic networks? In the next section, we would like to briefly discuss those questions a little further.

## 2.5 Network Analysis (3)

**Network Dynamics:** It is a matter of fact that language evolves; especially from a lexical perspective, where new vocabularies are generated and old senses sometimes disappear over time. However, tracing and observing such changes is rather difficult because such natural language evolution occurs over long periods of time. When considering the evolution of semantic networks, therefore, we assume that the growth of a semantic network may correspond to the increases in the numbers of words (nodes) and semantic relations (edges) in as more data is added in the construction of the network. Particularly, for our semantic networks which are built from word association data, the networks grow as more word association data is added.

In this sense, we can attempt to observe the growth process for semantic networks here. To that

aim, the KorWA network is particularly suitable, as it is constructed from KorWA data collected from two sessions that used exactly the same task. We may see how the network evolves by taking the sessions as two separate points in time.

From the beginning, the KorWA network starts with the 3,951 nodes that correspond to the set of stimulus words. It cannot be called a network at this stage because there are no links between these nodes. Then, as the word associations are collected, a network starts to appear by adding edges between the initial nodes and new nodes corresponding to the association responses. When the first session of data collection was complete, we found that the initially disconnected 3,951 nodes forming a large, well-connected network, as presented in Table 1.

**Table 1. Growth of the KorWA semantic network.**

	over time →		
	initially.	After 1 <sup>st</sup> session.	After 2 <sup>nd</sup> session.
Num. of nodes.	3,951.	9,054.	12,844 (Δ 3,790).
Num. of edges.	~.	13,669.	26,931 (Δ 13,262).
Average degree.	~.	3.02.	4.19.
Range of degree.	~.	1 - 87.	1 - 198.
Num. of components.	~.	127.	14.
Num. of nodes in the largest component (%).	~.	8,641 (95%).	12,807 (99.7%).
Pseudo diameter of the largest component.	~.	18.	13.

The number of nodes had increased to 9,054, and 13,669 edges were generated between them. 8,641 nodes corresponding to 95% of the total nodes are connected to each other, being the largest component in the network, but, at the same time, there were also 126 small partitions with 2 to 3 nodes connected to each. The pseudo diameter, the longest distance, of the largest component is 18, which indicates that the nodes within it are well connected to each other. In this network, a node has three links on average and the distribution of degrees in the network shows a power law distribution ( $P(k) \sim K^{-\gamma}$  with a degree exponent  $\gamma=2.42$ ), as in Figure 3 above.

Then, additional word associations were collected for the same set of stimulus words in the same manner as in the first session. When the new data was added to the first network, we obtained a larger network, as described in Table 1. The network grew by 12,844 nodes and 26,931 edges. Through this process, more than 99.7% of nodes (12,807) became interconnected, leaving on 37 words as elements disconnected from the whole graph. Moreover, the pseudo diameter of the larg-

est component became smaller despite the increase in its size. The discrepancy in the degrees of words became larger than before, with a degree range from 1 to 198.

**Table 2. Top 20 words with the highest degrees before and after growth of the KorWA network.**

Before growth	After growth
돈 ('money')/ 87	돈 ('money')/198
사랑 ('love')/ 79	사랑 ('love')/ 146
친구 ('friend')/ 56	친구 ('friend')/ 114
사람 ('human')/ 48	사람 ('human')/ 106
물 ('water')/ 48	마음 ('mind')/ 85
꿈 ('dream')/45	여자 ('woman')/80
군대 ('army')/ 45	물 ('water')/ 80
마음 ('mind')/ 44	공부 ('study')/ 74
집 ('house')/ 43	눈물 ('tear')/ 73
눈물 ('tear')/ 43	나 ('myself')/ 73
영화 ('movie')/39	꿈 ('dream')/ 70
공부 ('study')/39	군대 ('army')/ 69
눈 ('eye/snow')/ 36	집 ('house')/ 69
책 ('book')/ 35	술 ('alcohol')/ 69
술 ('alcohol')/ 34	책 ('book')/ 68
여자 ('woman')/ 34	눈 ('eye/snow')/ 65
나 ('myself')/ 33	싸움 ('fight')/64
물건 ('thing')/ 32	전쟁 ('war')/ 64
자동차 ('car')/ 32	영화 ('movie')/ 63
가족 ('family')/ 32	학교 ('school')/ 63

Note. The number after the slash indicates the degree for the word.

Over time (as reflected in the first and second sessions of data collection), 3,790 nodes and 13,262 edges newly appeared in the KorWA network. Through this growth, the network became much more interconnected, as clearly evidenced by the size of the largest component and the pseudo diameter. What is particularly salient is the number of links that a word has through the growth process. Interestingly, regardless of the double increase in the connections within the network, around 60% of the total nodes were still poorly connected, having a degree of only 1 or 2. On the other hand, some of nodes that already had plenty of links became much richer, becoming linked to even more other nodes; with the average degree for 1% of the total nodes being over 60. Table 2 lists the top 20 words in terms of highest degree values before and after growth. The first four words do not change in order, while the shifts for the other top items are not so

significant. However, for most of these items, the degree value roughly doubled.

From these observations, we can assume that there are some words that attract more links from other nodes, while most of these other words have just a few connections. This phenomenon appears even more conspicuously through the growth process. The scale-free nature of semantic networks also seems to reflect a kind of preferential attachment. What kinds of words always attract links from new nodes? As suggested already, these seem to be basic concept words, closely related to daily life and culture, and these hubs form a kind of bridge between several different conceptual domains.

Such words contributing to the connectivity of the network are central to the dynamic behavior of across the networks, and are likely to be key concept words for understanding a culture and for learning language within the contexts of semantic networks. Further study and exploration in the structural and dynamic characteristics within semantic networks may open up a new approach to semantics, cultural studies, and language learning from a cognitive perspective.

### 3 Conclusion and Further study

This paper has described our dataset to represent human language in the form of a network. With much interest in language as a communication and thinking tool, we have sought to build a semantic network representing lexical knowledge and the conceptual relations between words. To that aim, word association data is particularly suitable in terms of its data format and its abundant and useful content. We presented a project to collect Korean word association norms given the high utility and urgent need of data of this kind. We have detailed the project from the design of free association experiment to the basic analysis of the data collected.

The application of the word association data to computational neurolinguistics is an issue for our future work. We believe that our study could potentially represent a breakthrough for this research field. The methods of Mitchell et al. (2008), for example, suggest to us strong connections between neural activation data and lexical co-occurrence information, obtained from text corpora which plays a role of intermediating within linguistics

embodiment theory with a sensory-motor basis and amodal theory with computational models. According to Mitchell et al., the techniques of natural language processing combined with neural linguistics can enable us to predict the patterns of neural activation for word stimuli for which fMRI data are not yet available. In short, the neural associations within firing patterns turn out to be correlated with word associations within co-occurrence patterns.

However, the similarity coefficient or the distance between any two words might be computed not only from a set of documents but also from graphic representations of associative concepts, such as the one presented in this paper. If it is true that a word can be represented not only by a three-dimensional array of cerebral activation, but also in terms of the lexical relatedness that is incorporated as a linear combination of these patterns, it may not be an overstatement to say that there might be a structural homology between natural neural networks in the brain and semantic networks built from word association norms. This kind of meta-network perspective within cognitive science has become all the more important because attempts to fill the gaps in the modeling of neural pathways are increasingly attracting wide interest. Sporns et al. (2004), for instance, have tried to apply the conceptual methods of complex networks, such as *small word-scale free*, to cortical networks and to the more dynamic, functional and effective connectivity patterns that underlie human cognition. Similarly, Stam and Reijneveld (2007) have introduced a graph analysis applied to multi-channel recordings of brain activity, by setting up vertices at the anatomical loci within a neural circuit and linking some that elicit high correlation patterns to the same stimulus. Also within the experiment paradigms used by Mitchell et al, some techniques for constructing a network model could be effective for the distributional representation of cortical responses handled at the same level as meaning proximity, even though Mitchell et al. treated each voxel (volumetric pixel value in a 3-dimensional regular grid) independently. If such models of network settings could be applied to images of neural activation across all the voxels for a set of stimulus nouns, it is possible to assume, by a reverse process of parameter estimation, the existence of hidden semantic layers composed of unknown semantic features. These intermediate factors could

be compared with real vocabulary data, such as basic verbs (as in the experiment conducted by Mitchell et al.) taking the stimulus nouns as subjects or targets.

Moreover, the merits of introducing graph analysis techniques to computational neurolinguistics could possibly be found in the evolutionary dynamics of networks, to the extent that the degree of word nodes (or, more simply, their frequencies) could be weighted for the neural connectivity deduced from fMRI responses. The data formats of neural activation patterns could then assimilate diachronic data to represent how a network grows over time around the key concepts or hub words, in accordance with the learning processes of particular individuals. Future research from this perspective could also support the high accuracy of similar experiments regardless of distributional bias in word frequencies. Briefly, semantic networks constructed from word association data could convey the lexical co-occurrence of words within documents to a visual map of the human brain reacting to those words.



## References

- D. Mollison, F. Ball, and G. Scalia-Tomba. 1997. *Epidemics with two levels of mixing*, *Annals of Applied Probability* 7, pp. 46-89.
- A.-L. Barabási, and R. Albert. October 15, 1999. *Emergence of scaling in random networks*. *Science*, 286:509-512.
- P. Cramer. 1968. *Word association*. New York and London: Academic Press.
- J. Deese. 1962. "Form class and the determinants of association", *Journal of verbal learning and verbal behavior*, vol. 1, pp. 79-84.
- B. Dorow, D. Sergi, D. Widdows, E. Moses, K. Ling, and J. Eckmann. 2005. "Using Curvature and Markov Clustering in Graphs for Lexical Acquisition and Word Sense Discrimination", in MEANING-2005, 2nd Workshop organized by the MEANING Project.
- F. Galton. 1880. "Psychometric experiments", *Brain* 2, pp. 149-162.
- S. Saporta, 1955. *Linguistic structure as a factor and as a measure in word association*, in J. J. Jenkins (Ed.), *Associative process in verbal behavior: A Report of Minnesota Conference*, Minneapolis: University of Minnesota.
- T. Joyce. 2005. *Lexical association network maps for basic Japanese vocabulary*, in *Words in Asia cultural contexts*, V. B. Y. Ooi, A. Pakir, I. Talib, L. Tan, P. K. W. Tan, and Y. Y. Tan (Eds.). Singapore: National University of Singapore, pp. 114-120.
- G. R. Kiss. 1968. *Words associations and networks*, *Journal of Verbal Learning and Verbal Behavior*, vol.7, pp. 707-13.
- Y. J. Lee. 1970. *Comparative studies on word associations by male and female university students: based on adjectives and color-referring words (translated from 남녀 대학생의 언어형상에 관한 비교연구: 형용사와 색채어를 중심으로)*, *아시아 여성 연구 제 9 집*. 학술지 344.
- C. L. McEvoy and D. L. Nelson. 1982. *Category name and instance norms for 106 categories of various sizes*, *American Journal of Psychology* 95, pp. 581-634.
- H. Akama, J. Jung, M. Miyake, and T. Joyce. 2007. *Hierarchical Structure in Semantic Networks of Japanese Word Associations*, PACLIC21 (Pacific Asia Conference on Language, Information, and Computation-2007), pp.321-328.
- A. Carlson, K. Chang, M. Just, R. Mason, S. Shinkareva, T. Mitchell, and V. Malave. 2008. *Predicting human brain activity associated with the meanings of nouns*. *Science*, 320:1191-1195.
- C. L. McEvoy and D. L. Nelson. 2005. *Implicitly activated memories: The missing links of remembering*, In *Human learning and memory: Advances in theory and application*, C. Izawa and N. Ohta (Eds.). Mahwah, NJ and London: Lawrence Erlbaum Associates, pp 177-198.
- J. Archibald, J. Rees-Miller, M. Aronoff, and W. O'Grady. 2005. *Contemporary Linguistics: An introduction, 5th ed.* Boston & New York: Bedford/St. Martin's.
- J. Okamoto and S. Ishizaki. 2001. "Construction of associative concept dictionary with distance information, and comparison with electronic concept dictionary (translated from 概念間距離の定式化と既存電子辞書との比較)", *自然言語処理*, vol. 8, pp. 37-54.
- A. Vespignani, and R. Pastor-Satorras. 2001. *Epidemic spreading in scale-free networks*, *Physical Review Letter*. 86. pp. 3200-3203.
- H. S. Shin. 1998. "Korean vocabulary teaching and semantic dictionary (translated from 한국어 어휘교육과 의미사전)", *한국어교육* vol. 9, no.2, pp. 85-104.
- G. H. Jin, N. J. Nam, and S. G. Seo. 1998. "Determination of basic vocabulary for Korean language education as a foreign language (translated from 외국어로서의 한국어 교육을 위한 기초 어휘선정)", 1st year of annual report (December 14, 1998), Internationalization of Korean language promotion committee, Ministry of Culture, Sports, and Tourism.
- C. C. Hilgetag, D. R. Chialvo, M. Kaiser, and O. Sporns. 9 September 2004. "Organization, development and function of complex brain networks", *TRENDS in Cognitive Sciences* Vol.8 No.
- C. J. Stam, J. C. Reijneveld. 2007. "Graph theoretical analysis of complex networks in the brain". *Non-linear Biomedical Physics*.
- J. B. Tenenbaum and M. Steyvers. 2005. "The large-scale Structure of Semantic Networks: Statistical Analysis and a Model of Semantic Growth", *Cognitive Science* 29, pp.41-78.
- D. J. Watts and S. Strogatz. June 1998. "Collective dynamics of 'small-world' networks". *Nature* 393, pp.440-442.