Author(s): Lahti, Lauri

Title: Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language ability levels and frequency ranking

Year: 2015

Version: Post print

**Please cite the original version:**

Lahti, Lauri. 2015. Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language ability levels and frequency ranking. E-Learn - World Conference on E-Learning. Proc. E-Learn - World Conference on E-Learning, 19- 22 October 2015, Kona, Hawaii, USA. 6.

Lahti, L. (2015e). Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language ability levels and frequency ranking. Accepted to appear in Proc. E-Learn - World Conference on E-Learning, 19-22 October 2015, Kona, Hawaii, USA. Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, USA. Available online at http://www.editlib.org. Open access in Aaltodoc publication archive at http://aaltodoc.aalto.fi.

# Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language ability levels and frequency ranking

Lauri Lahti
Department of Computer Science
Aalto University School of Science, Finland

**Abstract**: We propose a new computational method to support learning that relies on adaptive exploration of the shortest paths in conceptual networks that have been formed based on co-occurrences of concepts in suitable text samples and selecting concepts corresponding to desired language ability levels and frequency ranking. Relying on Google Web 1T 5-gram database we have built a conceptual co-occurrence network reaching the coverage of 3018 unique concepts and 54 610 unique pairs of co-occurring concepts thus approximately matching with a vocabulary size suggested to be enough for sufficient comprehension and with the highest language ability levels of English Vocabulary Profile. Our method offers to the learner recommendations about suitable exploration paths along the shortest connecting paths between the concepts belonging to a desired learning topic vocabulary, computed with Yen's algorithm. By indicating for each concept the language ability level and the frequency ranking position in British National Corpus enables to prioritize such shortest paths of concepts that most best match the current suitable comprehension level of the learner. Our preliminary experiment showed that the method can pedagogically intuitively support cumulative adoption of knowledge in the context of study entities belonging to a core curriculum. Relying on our research we are opening a free educational resource at http://www.freelearningpath.org that enables learners and educators to get adaptive guidance for exploring a desired educational content.

## Introduction

We propose a new computational method to support learning that relies on adaptive exploration of the shortest paths in conceptual networks that have been formed based on co-occurrences of concepts in suitable text samples and selecting concepts corresponding to desired language ability levels and frequency ranking. In our previous work we have identified how the learner's exploration in conceptual networks built based on hyperlink network of Wikipedia online encyclopedia (http://www.wikipedia.org) can be used to support adoption of new knowledge and we have generated conceptual networks reaching sizes that correspond to vocabularies of various language ability levels of English Vocabulary Profile and Oxford Wordlist (Lahti 2015a).

Each article in Wikipedia can be considered to define a concept that is depicted by the title of the article and the hyperlink network connecting Wikipedia articles can be considered to define relationships between concepts. As of August 2015 the biggest language edition, English edition, of Wikipedia contains about 4.9 million articles whereas it has been estimated that a native adult possibly has a vocabulary of about 20 000 word families and that there are well over 54 000 word families in English (Nation & Waring 1997). Knowing 3000-5000 word families or just 2000-3000 word families can be enough for a sufficient 95-percent level comprehension of text (Nation & Waring 1997).

It has been found that a person typically responds faster to high-frequency words than low-frequency words of a language (Duyck et al. 2008) and that he recognizes and produces faster words that he has learned earlier than later in the life (Izura & Ellis 2002). According to a spreading activation theory the information in a memory is encoded in a network of interconnected cognitive units so that the units can spread activation to related units thus causing activation patterns representing specific conscious experiences (Anderson 1983; Dix et al. 2010). Small network properties have been identified in Wikipedia (Ingawale et al. 2009) and in human brain networks (Wang et al. 2010) and scale-free properties have been identified in Wikipedia (Masucci et al. 2011) and possibly in human brain networks (Bullmore & Sporns 2009). It has been identified that co-occurrence of concepts in sentences can be represented with networks that have scale-free small-world properties and in which the average distance between any two concepts is about 2-3 link steps (Ferrer i Cancho & Solé 2001). In individual associative networks that were grown during six weeks and reaching 1358-9429 nodes and 3729-27124 directed links had a small-world structure and the shortest paths in a range 5.65-7.05 links (Morais et al. 2013). Concerning one-hop replication networks, it was found that a small-world network containing 10 000 nodes has an average search length of about 950 hops for an average degree of 10 and an average search length of about 200 hops for an average degree of 30 (Redero-Merino et al. 2010).

## Previous work

In our previous work, in respect to the online database of English Vocabulary Profile that defines six cumulative language ability levels A1-C2 ranging from a beginner to an experienced language user (Capel 2013), it appeared in our experiments that the linkage between Wikipedia articles, which are considered concepts, offers varied degrees of conceptual coverage, as of June-July 2013 (Lahti 2015a). At the language ability level C1 we identified 3198 unique nouns and there were 21 410 unique Wikipedia hyperlinks connecting those 3198 unique nouns, and furthermore there were 2470 unique nouns in 21 410 unique Wikipedia hyperlinks connecting those 3198 unique nouns. At the language ability level C2 we identified 3710 unique nouns and respectively 25 153 unique connecting Wikipedia hyperlinks, and furthermore 2878 unique nouns in these unique Wikipedia hyperlinks. This means on average about 8.7 unique Wikipedia hyperlinks per each concept of a noun vocabulary of language ability level of C1 or C2 (i.e. dividing the number of unique hyperlinks by the number of unique nouns in these unique hyperlinks).

Vocabularies for all levels A1-C2 form together a vocabulary A1&A2&B1&B2&C1&C2. For each of 2878 unique nouns of the vocabulary A1&A2&B1&B2&C1&C2 there were on average 8.7 departing unique hyperlinks (a median value 5) and an average 8.7 arriving unique hyperlinks (a median value 5) linking to other unique nouns of the vocabulary A1&A2&B1&B2&C1&C2. Among 25 153 unique hyperlinks for 4824 hyperlinks (about 19 percent) there was another hyperlink going also into an opposite direction, thus 2412 connections were bidirectional.

In our previous work we have identified some correspondences between conceptual networks built based on Wikipedia hyperlink network and based on co-occurring concepts in text (Lahti 2015c). For a selected set of concepts, that we referred to as a *learning topic vocabulary*, we identified for each corresponding Wikipedia article the set of conceptual relationships based on links leading from the concept of the title of the article to all occurrences of concepts of the learning topic vocabulary in this article. All these gathered conceptual relationships together formed a *vocabulary co-occurrence network*.

For a collection of 175 unique concepts we contrasted the linkage of the vocabulary co-occurrence network and the linkage of Wikipedia hyperlink network. It appeared that the vocabulary co-occurrence network included 1338 unique conceptual pairs, containing 173 unique concepts, whereas the hyperlink network included 816 unique conceptual pairs, containing 167 unique concepts. In the vocabulary co-occurrence network in links each start concept had on average 7.73 (median 6) different unique end concepts and each end concept had on average 9.10 (median 4) different unique start concepts. In the hyperlink network in links each start concept had on average 4.92 (median 4) different unique end concepts and each end concept had on average 7.78 (median 5) different unique start concepts.

## Method

Motivated by our previous work that for a collection of 175 unique concepts contrasted the linkage of the vocabulary co-occurrence network and the linkage of Wikipedia hyperlink network (Lahti 2015c) we wanted to continue our research on a larger vocabulary scale to generate a *conceptual co-occurrence network* that corresponds to those conceptual networks that we had generated in our previous work based on Wikipedia hyperlink network for various language ability levels (Lahti 2015a). Thus we decided to build a conceptual co-occurrence network reaching the coverage of about 3000 unique concepts thus approximately matching with a vocabulary size suggested to be enough for a sufficient comprehension (Nation & Waring 1997) and with vocabulary sizes for language ability levels of C1 or C2 of English Vocabulary Profile (Capel 2013).

In our previous work (Lahti 2015c) we proposed a method for generating a *pedagogic co-occurrence network*. The purpose was to recommend for the student *learning paths* that are pedagogically rewarding exploration paths in a conceptual network. This was done by first identifying the highest-ranking traversed concepts and the highest-ranking traversed links in a randomly generated exploration path and then selecting in a decreasing order of ranking and with an increasing spacing some representative concepts and links to form the pedagogic co-occurrence network. Relying on the pedagogic co-occurrence network it was then possible to form a *randomness-based learning path* that tried to offer such an exploration routing that most optimally followed the principles of the theory of spaced learning.

Based on our previous work and its findings it seems to us that for relatively specific contexts and clearly constrained learning topics it can be practical to use the method relying on random explorations as proposed in our publication (Lahti 2015c). However, to address broader contexts and less clearly constrained learning topics it can be practical to use a method that considers more diverse and extensive semantic relatedness criteria. Thus we now want to propose a new computational method to support learning that relies on adaptive exploration of the shortest paths in conceptual networks that have been formed based on co-occurrences of concepts in a set of suitable text samples and selecting concepts corresponding to desired language ability levels. This leads us to propose using an extensive linguistic resource that defines co-occurrences of concepts in a large corpus that is Google Web 1T 5-gram database (FAU Erlangen-Nürnberg 2015) and a respected categorization of concepts into language ability levels that is English Vocabulary Profile (Capel 2013).

Resembling our previous work (Lahti 2015c), in our new method a set of essential concepts for the current learning task needs first to be identified, for example based on some of the most frequent words in a representative text sample, thus forming a learning topic vocabulary. Relying on a conceptual co-occurrence network that we have generated based on Google Web 1T 5-gram database the method offers to the learner recommendations about suitable exploration paths along the shortest connecting paths between the concepts belonging to the learning topic vocabulary. To compute top k shortest loopless paths in the network we suggest using Yen's algorithm (Yen 1971). Among alternative shortest paths shown in a decreasing order of the length of the path there typically appears such concepts in paths that belong to different language ability levels of English Vocabulary Profile and thus indicating for each concept the language ability level as well as the frequency ranking position in British National Corpus enables to prioritize such shortest paths of concepts that most best match the current suitable comprehension level of the learner. The method enables the learner to explore the shortest paths between the concepts of the learning topic vocabulary and thus to become cumulatively exposed to the modular knowledge structure of the learning topic and how its components are interconnected. Please note that different kinds of shortest paths may be gained for opposite directions between a pair of concepts as well as if all links of the network are enabled to be explored in both directions. It is also possible to show to the learner a partial listing about the concepts belonging to the shortest paths and then the learner should try to complete himself the listing based on recalling the shortest paths that have been shown to him earlier.

## Experiment

For each of 3018 highest-ranking nouns of British National Corpus available from lemmatized word lists of an online database (Leech et al. 2001) we queried an online database for Google Web 1T 5-gram database (FAU Erlangen-Nürnberg 2015) to identify all other nouns belonging to the same set of 3018 nouns that co-occur at the distance of at the most four words left or right, with the association measure of t-score and considering at most 50 highest-ranking nouns having a frequency of at least 40. When identifying nouns we relied on just matching spelling and thus some non-nouns may have possibly become unintentionally considered as nouns and some nouns may have become unintentionally excluded. For all 3018 unique nouns we gained together a set of 54 610 unique pairs of nouns (i.e. each pair corresponding to a co-occurrence of two nouns, both belonging to the set of 3018 nouns). There were 2994 unique nouns in 54 610 unique pairs of nouns.

For each of 2994 unique nouns there were on average 18.1 co-occurring other unique nouns of 2994 unique nouns (a median value 18) and each of 2994 unique nouns was co-occurring for an average 18.1 other unique nouns of 2994 unique nouns (a median value 11). Among 54 610 unique pairs of nouns for 26 406 pairs of nouns (about 48 percent) there was co-occurrence in both directions, thus 13 203 connections were bidirectional. If we coarsely theoretically estimate how the number of reachable concepts increases by traversing links in a network so that at each step the next link to traverse can be selected from about 18 alternative links, then $18^2$=324 concepts are reachable at the second step, $18^3$=5832 at the third step and $18^4$=104 976 at the fourth step, that can be contrasted with vocabulary sizes suggested for a sufficient comprehension (Nation & Waring 1997) and defined for language ability levels of English Vocabulary Profile (Capel 2013).

We wanted to carry out a preliminary experiment to evaluate how our proposed method can support cumulative adoption of knowledge in the context of study entities belonging to a core curriculum. In the official Finnish national core curriculum defined for compulsory basic education in primary school there are various formal study entities that are expected to become learned by the students (Finnish National Board of Education 2014). For class levels 1-2 in mathematics the titles of four major study entities include: skills of thinking, numbers and calculation, geometry and measurement, and computer science and statistics. For class levels 1-2 in environmental science the titles of six major study entities include: growth and development, acting at home and school, observation of changes in the neighbourhood, researching and experimenting, thinking about basic requirements for life and practicing a sustainable way of living. For a set of conceptual pairs identified in these ten titles of study units we generated the shortest paths in a conceptual co-occurrence network with our proposed method and evaluated the results (see Table 1); the pair of the concepts geometry and measurement was excluded since the concept geometry was not found in the conceptual co-occurrence network. Table 2 shows some of the shortest paths in a conceptual co-occurrence network for three conceptual pairs.

| *Conceptual pair, in parenthesis is a frequency ranking position in BNC (in range 1-314s) and a language ability level in EVP (in range A1-C2), "-s" indicates a shared position* | | *The number of steps in the shortest paths (the number of alternative routes for this path length)* | | |
|---|---|---|---|---|
| *Concept A* | *Concept B* | *Direction from concept A to concept B* | *Direction from concept B to concept A* | *When all links enabled to be traversed in both directions* |
| skill (198s/B1) | thinking (281s/-) | 4 (12 alt.) | 4 (1 alt.) | 3 (2 alt.) |
| number (18/A1) | calculation (301s/B2) | 4 (11 alt.) | 2 (3 alt.); 3 (32 alt.) | 2 (3 alt.); 3 (54 alt.) |
| computer (153/A1) | statistics (292s/B2) | 3 (8 alt.) | 2 (3 alt.); 3 (41 alt.) | 2 (3 alt.); 3 (54 alt.) |
| growth (193s/B2) | development (47s/B1) | 1 (1 alt.); 2 (3 alt.) | 1 (1 alt.); 2 (2 alt.) | 1 (1 alt.); 2 (8 alt.) |
| home (42s/A1) | school (27s/A1) | 2 (2 alt.); 3 (23 alt.) | 2 (2 alt.); 3 (30 alt.) | 2 (4 alt.); 3 (94 alt.) |
| observation (274s/B2) | neighbourhood (307s/A2) | 4 (1 alt.) | 4 (1 alt.) | 3 (6 alt.) |
| research (92/B1) | experiment (267s/B1) | 4 (7 alt.) | 2 (4 alt.); 3 (27 alt.) | 2 (4 alt.); 3 (62 alt.) |
| requirement (231s/B2) | life (12/A1) | 3 (7 alt.) | 4 (6 alt.) | 3 (20 alt.) |
| way (4/A2) | living (278s/B2) | 2 (1 alt.); 3 (2 alt.) | 2 (1 alt.); 3 (5 alt.) | 2 (2 alt.); 3 (20 alt.) |

**Table 1**: Some properties of the shortest paths in a conceptual co-occurrence network for nine conceptual pairs identified in the titles of study units of mathematics (in rows 1-3) and environmental science (in rows 4-6).

## Discussion and future work

In our previous work (Lahti 2015a; Lahti 2015b) at the language ability level C2 having the vocabulary A1&A2&B1&B2&C1&C2 of 3710 unique nouns these nouns were connected by 25 153 unique Wikipedia hyperlinks which contained 2878 unique nouns with the following distribution of new nouns at each level: 283 at A1, 483 at A2, 706 at B1, 718 at B2, 328 at C1 and 360 at C2. Now in the conceptual co-occurrence network having 2994 unique nouns in 54 610 unique pairs of nouns there was the following distribution of new nouns at each language ability level: 261 at A1, 393 at A2, 639 at B1, 689 at B2, 282 at C1 and 226 at C2, altogether 2490 nouns of the vocabulary A1&A2&B1&B2&C1&C2. In 25 153 unique Wikipedia hyperlinks the nouns having the highest number or departing links were human (121), food (93), water (85), nature (79) and entertainment (74), and the nouns having the highest number or arriving links were animal (108), human/water (106), earth (101), mammal (98) and psychology (92). In 54 610 unique pairs of nouns of the conceptual co-occurrence network the nouns having the highest number or departing links were management/reduction/road (34), development/holder/maximum/planning/prevention/venue (33) and chief/collar/conservation/department/maintenance/production/science/transmission (32), and the nouns having the highest number or arriving links were management (157), black (144), health (137), american (126), human (124), red (119), water (118), no (117), law (115) and service (110). In the lemmatized word lists of British National Corpus (Leech et al. 2001) the nouns having the highest number of occurrences per million words were time (1833), year (1639), people (1256), way (1108) and man (1003). In the previous expression the departing links refer to co-occurring other nouns for the current noun and the arriving links refer to the current noun co-occurring for other nouns.

When the shortest paths include alternative paths having an equal number of steps we suggest that the student can be supported in selecting the most pedagogically suitable path to be explored next by indicating the different roles of intermediary concepts along the path. This means that in paths such concepts can be highlighted that represent certain language ability levels (possibly lower or higher levels than the levels of the concepts at the endpoints of the path) or which occur in those paths going into both traversal directions between endpoints or in those paths explored when all links are enabled to be traversed bidirectionally. In addition, it is possible to highlight concepts which have frequently occurred in just previously explored paths during the current learning session. We suggest that if the shortest paths contain only two or three steps it can be useful to provide more diverse contexts for the student by recommending a bit longer paths, and possibly a priority should be given for such paths that contain such concepts that frequently occur in shorter/longer paths than the current path.

| *Some of the shortest paths in a conceptual co-occurrence network, in parenthesis is a frequency ranking position in BNC (in range 1-314s) and a language ability level in EVP (in range A1-C2), "-s" indicates a shared position* |
|---|
| <B: computer (153/A1) ¤ crime (235s/B1) ¤ statistics (292s/B2) |
| <B: computer (153/A1) ¤ mathematics (304s/C2) ¤ statistics (292s/B2) |
| <B: computer (153/A1) ¤ department (117s/A2) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ engineering (273s/B1) ¤ mathematics (304s/C2) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ department (117s/A2) ¤ justice (249s/B2) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ department (117s/A2) ¤ mathematics (304s/C2) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ system (14/B1) ¤ justice (249s/B2) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ department (117s/A2) ¤ economics (295s/B1) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ network (233s/B2) ¤ traffic (257s/A2) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ lab (311s/B1) ¤ crime (235s/B1) ¤ statistics (292s/B2) |
| >(B): computer (153/A1) ¤ science (197s/A2) ¤ mathematics (304s/C2) ¤ statistics (292s/B2) |
| ><B: growth (193s/B2) ¤ development (47s/B1) |
| >B: growth (193s/B2) ¤ future (168s/B1) ¤ development (47s/B1) |
| ><B: growth (193s/B2) ¤ human (291s/B1) ¤ development (47s/B1) |
| <B: growth (193s/B2) ¤ fund (206s/C1) ¤ development (47s/B1) |
| <B: growth (193s/B2) ¤ strategy (235s/B2) ¤ development (47s/B1) |
| B: growth (193s/B2) ¤ spur (312s/C2) ¤ development (47s/B1) |
| B: growth (193s/B2) ¤ poverty (293s/B2) ¤ development (47s/B1) |
| B: growth (193s/B2) ¤ sector (214s/C1) ¤ development (47s/B1) |
| B: growth (193s/B2) ¤ innovation (300s/C1) ¤ development (47s/B1) |
| B: observation (274s/B2) ¤ deck (307s/B2) ¤ strategy (235s/B2) ¤ neighbourhood (307s/B1) |
| B: observation (274s/B2) ¤ classroom (296s/A1) ¤ management (117s/B2) ¤ neighbourhood (307s/B1) |
| B: observation (274s/B2) ¤ period (78/B1) ¤ transition (297s/C2) ¤ neighbourhood (307s/B1) |
| B: observation (274s/B2) ¤ period (78/B1) ¤ renewal (313s/C1) ¤ neighbourhood (307s/B1) |
| B: observation (274s/B2) ¤ wildlife (304s/B1) ¤ management (117s/B2) ¤ neighbourhood (307s/B1) |
| B: observation (274s/B2) ¤ wildlife (304s/B1) ¤ fund (206s/C1) ¤ neighbourhood (307s/B1) |
| >(B): observation (274s/B2) ¤ plant (178s/A1) ¤ maintenance (284s/B2) ¤ renewal (313s/C1) ¤ neighbourhood (307s/B1) |
| <(B): observation (274s/B2) ¤ prediction (308s/B2) ¤ storm (297s/A2) ¤ watch (292s/A1) ¤ neighbourhood (307s/B1) |

**Table 2**: Some of the shortest paths in a conceptual co-occurrence network for the conceptual pairs of computer and statistics, growth and development, and observation and neighbourhood. A notation > indicates a direction from the first to the last concept, a notation < indicates the opposite direction (from the last to the first concept) and a notation B indicates the case when all links of the network are enabled to be explored in both directions. A notation (B) indicates that when all links are enabled to be explored bidirectionally there are more bidirectional paths of this length than can be shown here (i.e. 64 paths for the pair computer and statistics and 198 paths for the pair observation and neighbourhood).

We suggest that a learning process can be supported with cumulative and modular network structures having properties motivated by previous results, including vocabulary levels, the shortest paths and sufficiently spaced repetition in diverse routings. There is a great need for further research to develop efficient computational modeling methods for better identifying various semantic dependencies that govern formation of meaningful pedagogical knowledge structures and educational stories. Anyway, our preliminary experiment shows how in our proposed method already relatively simple computational and linguistic approaches can enable generating useful adaptive educational resources that can be applied in diverse ways to enhance learning opportunities for students. We do not know any previous work similar to our proposed method especially in respect to generating a diverse set of alternative exploration paths so that selecting the most pedagogically suitable paths is supported by indicating the frequency ranking position and the language ability level of intermediary concepts along the path. It seems that for each concept also many other linguistic properties could be usefully indicated for the student to help finding the optimal exploration routing in educational knowledge to adopt it efficiently. We suggest that even on a broader scale than ordinary schooling the conceptual shortest paths generated by our proposed method could offer tools for initiating inspiration in creative work and brainstorming.

Relying on our research we are opening a free educational resource at http://www.freelearningpath.org that enables learners and educators to get adaptive guidance for exploring a desired educational content. Full listings of conceptual relationships and co-occurring conceptual pairs concerning the conceptual networks we have generated based on Wikipedia linkage and based on Google Web 1T 5-gram database are freely available online in our supplementing publications (Lahti 2015b; Lahti 2015f).

## References


Lahti, L. (2015a). Computer-assisted learning based on cumulative vocabularies, conceptual networks and Wikipedia linkage. Doctoral dissertation. Department of Computer Science, Aalto University School of Science, Finland. Unigrafia Oy, Helsinki, Finland. ISBN 978-952-60-6163-4 (printed), ISBN 978-952-60-6164-1 (pdf). http://urn.fi/URN:ISBN:978-952-60-6164-1.

Lahti, L. (2015b). Supplement to Lauri Lahti’s doctoral dissertation "Computer-assisted learning based on cumulative vocabularies, conceptual networks and Wikipedia linkage". Unigrafia Oy, Helsinki, Finland. Print ISBN 978-952-60-3707-3 and online ISBN 978-952-60-3708-0. http://urn.fi/URN:NBN:fi:aalto-201503182047.

Lahti, L. (2015c). Generation of learning paths in educational texts based on vocabulary co-occurrence networks in Wikipedia and randomness. Proc. Global Learn 2015: Global Conference on Learning and Technology, 16-17 April 2015. Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, USA. http://www.editlib.org/p/150943/.

Lahti, L. (2015f). Supplement to Lauri Lahti's conference article ”Educational exploration along the shortest paths in conceptual networks based on co-occurrence, language ability levels and frequency ranking”. Available online at http://aaltodoc.aalto.fi.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In Schmitt, N., & McCarthy, M. (eds.), Vocabulary: Description, Acquisition, Pedagogy. Cambridge University Press, New York, USA, 6-19.

Duyck, W., Vaderelst, D., Desmet, T., & Hartsuiker, R. (2008). The frequency effect in second-language visual word recognition. Psychonomic Bulletin & Review, 15(4), 850-855. http://users.ugent.be/~wduyck/articles/DuyckVanderelstDesmetHartsuiker2008.pdf

Izura, C. & Ellis, A. (2002). Age of acquisition effects in word recognition and production in first and second languages. Psicológica, 23, 245-281. www.uv.es/revispsi/articulos2.02/4.IZURA%26ELLIS.pdf

Anderson, J. (1983). A spreading activation theory of memory. Journal of verbal learning and verbal behavior. 22(3), 261-295.

Dix, A., Kafifori, A., Lepouras, G., Vassilakis, C., & Shabir, N. (2010). Spreading activation over ontology-based resources: from personal context to Web scale reasoning. International Journal of Semantic Computing (IJSC), 4(1), 59-102.

Ingawale, M., Dutta, A., Roy, R., & Seetharaman, P. (2009). The small wordls of Wikipedia: implications for growth, quality and sustainability of collaborative knowledge networks. Proc. Americas Conference on Infromation Systems (AMCIS 2009).

Wang, J., Zuo, X., & He, Y. (2010). Graph-based network analysis of resting-state functional MRI. Frontiers in Systems Neuroscience, 4:16.

Masucci, A., Kalampokis, A., Eguíluz, V., & Hernández-García, E. (2011). Wikipedia information flow analysis reveals the scale-free architecture of teh semantic space. Public Library of Science ONE (PLoS ONE), 6(2), e17333.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186-198.

Ferrer i Cancho, R., & Solé, R. (2001). The small world of human language. Proc. of the Royal Society of London, B., 268, 2261-2265.

Morais, A., Olsson, H., & Schooler, L. (2013). Mapping the structure of semantic memory. Cognitive Science, 37, 125-145.

Redero-Merino, L., Fernández Anta, A., López, L., & Cholvi, V. (2010). Perforamance of random walks in one-hop replication networks. Computer Networks, 54, 781-796.

Capel, A. (2013). Completing the English Vocabulary Profile: C1 and C2 vocabulary. English Profile Journal, 3, e1. Online database of English Vocabulary Profile provided by Cambridge University Press available at http://vocabulary.englishprofile.org/dictionary//word-list/uk/a1_c2/A.

Leech, G., Rayson, P., & Wilson, A. (2001). Word frequencies in written and spoken English: based on the British National Corpus. Longman, London, UK. ISBN 0582-32007-0. (A companion web site: http://ucrel.lancs.ac.uk/bncfreq/flists.html; http://ucrel.lancs.ac.uk/bncfreq/lists/5_1_all_rank_noun.txt)

FAU Erlangen-Nürnberg (2015). The Google Web 1T 5-Gram Database - SQLite Index & Web Interface - Collocations. http://corpora.linguistik.uni-erlangen.de/demos/cgi-bin/Web1T5/Web1T5_colloc.perl

Yen, J. (1971). Finding the k shortest loopless paths in a network. Management Science, 17(11), 712-716.

Finnish National Board of Education (2014). Finnish national core curriculum for compulsory basic education (in Finnish). http://www.oph.fi/download/163777_perusopetuksen_opetussuunnitelman_perusteet_2014.pdf