Author(s): Lahti, Lauri

Title: Educational framework for adoption of vocabulary based on Wikipedia linkage and spaced learning

Year: 2012

Version: Post print

Please cite the original version:

Lahti, Lauri. 2012. Educational framework for adoption of vocabulary based on Wikipedia linkage and spaced learning. Global Learn 2012: Global Conference on Learning and Technology, online conference on 6 November 2012. P. 8-13. ISBN 1-880094-99-1 (electronic).

Lahti, L. (2012). Educational framework for adoption of vocabulary based on Wikipedia linkage and spaced learning. Proc. Global Learn 2012: Global Conference on Learning and Technology, online conference on 6 November 2012 (eds. Bastiaens, T., & Marks, G.), 8-13. Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, USA. ISBN 1- 880094-99-1. http://www.editlib.org/p/42033/

Educational Framework for Adoption of Vocabulary Based on Wikipedia Linkage and Spaced Learning

Lauri Lahti
Department of Computer Science and Engineering
Aalto University School of Science, Espoo, Finland

Abstract: We propose a new educational framework supporting learners in adoption of vocabulary and knowledge for various personalized and engaging pedagogic purposes. The most essential concepts of learning topic are linked based on the shortest connecting paths in hyperlink network of Wikipedia encyclopedia assisted with Wiktionary dictionary. In resulting pedagogic conceptual network learner traverses links along learning path generated by method in sequential process having tailored variation and repetition computed based on theory of spaced learning. The learning path is shown to learner as sequence of compact relation statements extracted from sentences surrounding hyperlinks in Wikipedia articles, supplied with set of visualizations based on main verb identified in them. Based on framework we have implemented prototype tool and promising experiments in educational setting indicated valuable support for adoption of new knowledge.

Introduction

To develop new computational methodology for education there is a need to address the latest understanding about cognitive processes and language acquisition. We hope that our new educational framework based on Wikipedia linkage and spaced learning can serve as a pioneering work by introducing a scalable conceptual grid for well timed personalized adoption of knowledge. It has been estimated that the vocabulary of a 5-year-old contains 4000-5000 word families and of an adult 20000 word families for native English speakers (Nation and Waring 1997), and for secondary English speakers highly educated adult manages well with 8000-9000 word families (Nation 2006). Yearly adoption for students is about 1000 new word families (Nation and Waring 1997) or 600-1200 new root words meaning weekly roughly 40-70 new words but of which only perhaps 8-10 words are learned at school (Lehr et al. 2004). It has been suggested that there are well over 54000 word families in English (Nation and Waring 1997) and children are exposed to 88500 word families in school texts for grades 3-9 (Nagy and Anderson 1984). British National Corpus (BNC) consisting of 100 million words from samples of texts and speech has 13994 unique lemmatized words occurring at least 100 times (Chujo 2004). We suggest to parallel all these knowledge entities and evolving cross-linking of them with 4,0 million articles networked in the English Wikipedia encyclopedia (http://en.wikipedia.org), each article defining a concept and enabling generation of personalized learning paths.

Previous work

Understanding 95 percent of general texts - omitting one in every 20 words - has been considered sufficient for reasonable comprehension (Laufer 1989), thus requiring a vocabulary of 3000-5000 or just 2000-3000 word families (Nation and Waring 1997). 95 percent understanding of English-for-second-language senior high school textbooks required about 3000 highest-ranking lemmatized BNC words (Chujo 2004). English-for-second-language college textbooks have a vocabulary level of 4000-4500 high-frequency word families and beyond highest-ranking 2000 word families a typical book supplies student with 162-2001 new word families (Hsu 2009). Hunt and Beglar (Hunt and Beglar 2005) concluded that chance of retaining meaning of a word is 5-20 percent and learning effectiveness benefits from combined distributed adopting and retrieval of knowledge at the longest delay that still maintains correct recall, for example gradually increasing delay for repeated retrieval up to 30 days. It was estimated that children did know and understand words they had encountered twelve times but not those encountered four times (McKeown et al. 1985). Repeated retrieval of information has been shown as key factor for long-term retention (Karpicke and Roediger 2007). In meta-analysis of 317 experiments, Cepeda et al. (Cepeda et al. 2006) concluded that when compared to non-spaced learning, spaced learning of items consistently showed benefits regardless of retention interval, and learning benefits increased as time lags increased between learning presentations. They also concluded that interstudy intervals producing maximal retention increased as retention interval increased. For sufficient comprehension reading rate of at least about 200 words per minute (Anderson 1999) and average sentence length below 20 words (DuBay 2004) was suggested, thus resulting at least 10 sentences per minute.

According to Hebbian learning theory associative learning relies on simultaneous activation of neural cells that increases synaptic strengths between them (Doidge 2007). Memory traces are stabilized by synaptic

consolidation within minutes to hours of learning and by system consolidation within weeks, months or even years. Harvey and Svoboda (Harvey and Svoboda 2007) showed with mice and rats that when a spine of synapse is stimulated to action potential also surrounding spines in distance of 10 micrometers are more sensitive for stimulus for about 10 minutes. Kandel (Kandel 2001) showed that stimulation of synapses of a marine snail can be successfully triggered by 4-5 spaced puffs of serotonin leading to activation of genes establishing long-term memory. Fields (Fields 2005) showed that to activate a gene for long-term memory formation in a synapse of mouse there is a need for at least three action potentials at least 10 minutes apart, and once the gene is activated it produces required proteins for about 30 minutes. Thus when learning a new knowledge item also human brain might benefit from 3-5 short distinct exposures separated by 10 minutes and then additional 30 minutes for continuous exposures. Marine snail exposed to four brief trains for four days could generate memories that lasted weeks (Kandel 2001).

Spacing lessons in time can promote children's ability for generalization of science concepts (Vlach and Sandhofer 2012). Incongruities between individual's knowledge and wiki's information leading to equilibration activities can enable learning (Cress and Kimmerle 2008). Machine learning techniques can be successfully used to classify semantic relations from hyperlink structure of Wikipedia (Bauer 2007). Tools based on exploring wiki knowledge have been developed to offer shallow learning curve and expressiveness of natural language (Kuhn 2009; Lahti 2011). Wozniak and Gorzelanczyk (Wozniak and Gorzelanczyk 1994) proposed computational method generating repetition intervals optimized so that 5 percent of to-be-remembered items are not remembered at the moment of repetition. Pavlik and Anderson (Pavlik and Anderson 2008) showed algorithm dynamically increasing and decreasing temporal spacing of items to improve recall and recall latency and to minimize failure-related time.

Method

Motivated by previous research we propose a new educational framework based on method that adjusts sequential ordering and spaced repetition to support adoption of new knowledge. For each learning topic it is possible to define a *learning topic vocabulary*, a set of concepts covering its essential pedagogical knowledge in respect to learner's needs, that can be selected manually by learner or teacher, or be a high-frequency wordlist extracted for example from course book or lecture slideshow. To avoid semantic challenges we currently accept only nouns to vocabulary. Learning takes place in series of sessions, for example one session per day or per week, each one focusing on learning a *session vocabulary* that is a subset of the learning topic vocabulary. A *pedagogic conceptual network* is generated by linking concepts of session vocabulary based on the shortest paths in hyperlink network connecting corresponding articles of English edition of Wikipedia encyclopedia (http://en.wikipedia.org). Each Wikipedia article represents a concept depicted by its title entry and all departing hyperlinks in this article define its relationships to other concepts. To find satisfactory definitions and redirects in cases of disambiguation and synonyms we used Wiktionary dictionary (http://en.wiktionary.org). A compact *relation statement* - containing main verb with some adjacent words - is extracted from sentence surrounding the departing hyperlink in article text to depict semantic relationship of linked concepts. In each learning session, learner cumulatively strengthens adoption of concepts belonging to pedagogic conceptual network as method shows step by step a sequence of chained relation statements based on routing generated to traverse conceptual linking of network. After reading currently shown relation statement, learner presses button "Next" to proceed to seeing following one. While traversing each link learner becomes fruitfully exposed to associative mixture of old and new knowledge in a sequential process having tailored variation and repetition computed based on theory of spaced learning. We call this traversed route as a *learning path*. (Fig. 1) shows an excerpt of a learning path based on traversing concepts in a small pedagogic conceptual network going through a link chain Family->Child->Parent->Birth. Dotted arrows indicate possible traversal routes in network and solid arrows route that forms learning path this time. Concepts traversed recently and requiring spacing before being traversing again are in parenthesis. The learning path is shown to the learner as a sequence of following relation statements extracted from Wikipedia articles: "*Family* helps in socialization process of child", "*Child* defines a relationship to parent or authority" and "Mother is a *parent* who performs the birth".

User interface of a prototype tool implementing the framework has three parts. One by one, *learning path illustration area* shows to learner the relation statement encountered next along learning path, supplied with static or animated visualization. *Concept map area* enables learner to draw concept maps during initialization and intermittent retention tasks. *Control panel* enables learner to adjust manually all parameters affecting learning session if needed. Learning session is constrained by parameters and values adjusted by learner's activity. Based on learner's needs and teacher's advice or earlier testing, learner manually sets parameters of current session: *session vocabulary size*, *degree of new content*, *session duration*, *learning speed*, *degree of required adoption*, *degree of exposure repetition*, *degree of retention repetition*, *interval of exposures* and *interval of retentions* (defined in (Tab. 1)). When starting new learning session method first evaluates learner's initial conceptualization level with following process. The

method asks learner to indicate desired new learning topic by naming one or more familiar concepts about it which defines initial form of session vocabulary. These concepts are cross-linked based on the shortest hyperlink chains in Wikipedia to create initial form of the pedagogic conceptual network. Method now automatically extends initial form of pedagogic conceptual network to cover as many concepts as defined by parameter "session vocabulary size" by progressively uniformly linking new concepts to it according to how Wikipedia articles corresponding to current concepts have the nearest hyperlinked articles. These new concepts are also added to session vocabulary. Next, method generates a random excerpt of learning path containing 20 steps and shows its relation statements to learner in a sequence and then method asks learner to draw a concept map representing her best recall and understanding about concepts she just saw and how they were linked. Method compares how much concepts and their links in concept map overlap with hyperlink structure of pedagogic conceptual network, and degree of matching between them defines parameter *degree of forgetting*, on scale 0,05-1,00 (overlap of 5-100 percent). Then method supplies each concept of pedagogic conceptual network with a value *measure of adoption* defined as (1 - degree of forgetting), estimating learner's probability to remember meaning of this concept in a random next future encounter.

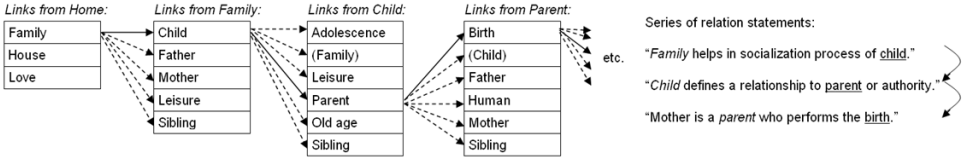

Figure 1: An excerpt of a learning path and sequence of extracted relation statements shown to the learner.

Parameter	*Definition*
Session vocabulary size	Amount of different concepts traversable in pedagogic conceptual network during learning session
Degree of new content	Percentage of previously unknown concepts in session vocabulary to be exposed to during learning session
Session duration	Available time for learning session
Learning speed	Aimed rate to traverse links (relation statements) per minute along learning path
Degree of forgetting	Probability that the meaning of an encountered concept is not remembered in a random next future encounter
Degree of required adoption	Minimum value of probability that meaning of an encountered concept is remembered in a random next future encounter and which is a probability value considered sufficient for a concept to be declared as learned well
Degree of exposure repetition	Minimum number of spaced exposures of a concept needed to learn it well
Degree of retention repetition	Minimum number of spaced retentions of a concept needed to learn it well
Interval of exposures	Minimum time between spaced exposures of a concept needed to learn it well
Interval of retentions	Minimum time between spaced retentions of a concept needed to learn it well

Table 1: Parameters of the framework affecting the learning during learning session.

Framework records evolution of learning path and learner can continue learning of previous session by loading from database learning paths and values achieved so far for each concept of learning topic vocabulary and pedagogic conceptual network. For each concept, framework keeps a record and updates five values. Besides "measure of adoption", they include *measure of exposure repetitions* (number of spaced exposures of the concept so far), *measure of retention repetitions* (number of spaced retentions of the concept so far), *time between exposures* (average time between spaced exposures of the concept so far) and *time between retentions* (average time between spaced retentions of the concept so far). At each step of proceeding to next concept along learning path, all five values of that concept are updated. "Measure of adoption" is updated by formula based on cumulative multiplication of probabilities of forgetting: *measure of adoption* $_{n+1}$ *= (1 - (1 - measure of adoption* $_{n}$ *)(degree of forgetting)).*

When generating learning path, framework guides learner to traverse in pedagogic conceptual network at each step from current concept next to a concept having now the lowest "measure of adoption", along the shortest connecting hyperlink chain. However, an additional restriction is that method aims to ensure fertile spacing between instances of traversing same concept again according to value "interval of exposures". If several concepts share the lowest value, framework guides learner to traverse to that concept which is encountered first with breadth-first search starting from current concept. Learning path is generated at each step to proceed next only to concepts directly linked from previous concept. Due to naturally emerging clustering hierarchy of hyperlink network connecting session vocabulary, framework can somewhat prioritize such routes that give additional probability for traversing hubs in pedagogic conceptual network to reach distant concepts and links. Learner should traverse concepts with defined learning speed within 10 percent margin or framework recommends learner to adjust her speed. If session vocabulary contains more unknown concepts than value "degree of new content" allows, framework first generates such learning path that traverses only inside a subset of session vocabulary having number of unknown concepts low enough to qualify "degree of new content". When unknown concepts of subset later gradually become learned due to spaced repetition, traversable vocabulary (subset) is cumulatively extended with additional unknown concepts.

With a spacing defined by parameter "interval of retention", method periodically interrupts proceeding along learning path with a *retention task* by asking learner to draw a concept map representing her best recall and understanding about concepts she just saw and how they were linked since the start of current learning session or

since the latest retention task. Method compares how much concepts and their links in concept map overlap with hyperlink structure of pedagogic conceptual network, and degree of matching between them redefines “measure of adoption” for all concepts involved and also “degree of forgetting”. All concepts belonging to session vocabulary need to become traversed in pedagogic conceptual network along learning path so many times and with sufficient spacing that finally - due to repeated cumulative exposure and retention - for each concept value “measure of adoption” reaches “degree of required adoption”, “measure of exposure repetitions” reaches “degree of exposure repetition” and “measure of retention repetitions” reaches “degree of retention repetition”. Now each concept of session vocabulary has reached enough exposures and retention to be declared as learned well. Then - or if session has lasted longer than “session duration” - session ends and method reports “measures of adoption” for each concept of session vocabulary and supplementing statistic about evolution of learning path, like number of traversals per each hyperlink. These results are stored so that adoption of vocabulary can flexibly continue in future learning sessions.

While the learner proceeds the links in pedagogic conceptual network, each shown relation statement is supplied with a visualization to help conceptualization of the relationship between the pair of concepts of the current link. Each visualization consists of two *concept pictographs* representing concepts, extracted from the images of corresponding Wikipedia articles or queried from Wikipedia Commons open image database (http://commons.wikimedia.org/), and a *transitional effect* (either static or dynamic) representing the relationship between concepts based on extracted relation statement. Static transitional effects consist of semitransparent still images placed over and between concept pictographs following visualization conventions of story-telling in comic strips and dynamic transitional effects consist of coordinating animation implementing movements and transformations of concept pictographs following visualization conventions of story-telling in cinema. A database of visualization conventions is collaboratively maintained for most frequently encountered key terms in relation statements and is queried to find most matching transition effect. To avoid semantic challenges transition effect is currently selected based on main verb identified in relation statement after it is supplied with part-of-speech tagging.

Experiment

Based on framework we have implemented a prototype tool and to verify its suggested educational gain we have carried out initial testing in real educational setting and tried to identify optimal measures for parameters and values. Since process of cumulatively gathering of knowledge along school years appears as an important area for research we have so far focused on analyzing learning topic vocabularies formed especially in respect to different learner characteristics based on age groups, gender, and written and spoken language. Since previous research indicates that vocabulary of just 2000 word families can be sufficient for reasonable comprehension in knowledge acquisition, we decided to define diverse set of unique tailored pedagogic conceptual networks that reach this critical size addressing various alternative characteristics of learner. Thus we decided to identify 2000 highest-ranking concepts for learners representing various combinations of characteristics and then define tailored pedagogic conceptual networks based on these concepts. We used three resources to gather the highest-ranking vocabularies of learners of various age groups. Oxford Wordlist (Bayetto 2010) provides high-frequency wordlists gathered from children of five school level (preparatory, 1st, 2nd, 3rd and 4th) as well as for boys and girls. British National Corpus (Leech et al. 2001) provides a high-frequency wordlist emphasizing adult’s language gathered from diverse samples. These two resources offer also separate wordlists for written and spoken language. To fill in the apparent age gap we gathered from written text of 70 teenagers high-frequency wordlists separately for five different ages (15, 16, 17, 18 and 19 years) as well as for boys and girls. Although bigger vocabularies generally seem to enable more diverse and fine-tuned educational results, our findings indicate that already pedagogic conceptual networks based on vocabulary of just 2000 concepts can offer valuable support for conceptualization of vocabulary and adoption of new knowledge.

With an experimental group of teenaged learners (n=49) we carried out knowledge acquisition tasks based on learning paths traversing 20 consecutive links inside a subset of 102 highest-ranking concepts in a pedagogic conceptual network generated for teenaged students. A control group (n=24) proceeded a predefined fixed series of 20 links providing same kind of sentences as for the experimental group but without semantically motivated continuity between linked concepts. In a following recall task, the experimental group managed to reproduce about 65 percent of adopted concepts thus clearly outperforming the control group reproducing only about 28 percent. 55 percent of learners considered traversing learning paths with prototype more useful or somewhat more useful than traditional learning from book, 31 percent equally useful and 14 percent less useful or somewhat less useful. Promisingly, learning paths suggested by framework seemed to match well learner’s needs when learning paths were generated by using such parameters that correspond to learning practices typical for successful spaced learning. In accordance with previous research our preliminary testing suggests using following approximate values for parameters. “Session duration” of about 30 minutes and “session vocabulary size” of about 100 concepts may enable

enough variation and spaced repetition. To avoid cognitive overload it seems promising to have about 5 percent as "degree of new content" and about 85 percent as the "degree of forgetting". To maintain continuity of comprehension, "learning speed" could be about 10 traversed concepts per minute. For each concept at least value 3-5 is suggested as "degree of exposure repetition" and "degree of retention repetition", and 10 minutes as "intervals of exposures" and "interval of retentions" to ensure it becomes learned. "Degree of required adoption" could be 95 percent so that with this probability the meaning of each concept is remembered in random next future encounter.

Discussion and future work

The proposed framework relies on a process in which the learner traverses hyperlink network of Wikipedia along a route generated based on verified notion that spacing exposures and retention at optimal intervals can support adoption of concepts and knowledge. As a part of our research we have already experimentally generated a variety of ready-to-use pedagogic conceptual networks for selected learning topic vocabularies concerning many popular conceptual themes of knowledge used actively in everyday life and encountered often in educational setting addressing diverse combination of characteristics of learner. Due to space constraints this collection is only available from the author's web site (http://www.cs.hut.fi/u/llahti/publ/lahti_2012a_data.pdf). We recognize risk of escalating ambiguity since many words have alternative meanings, even belonging to different word classes, and to simplify management of adoptable concepts of various types we currently accept just nouns but expect extending to other word classes to be promising. To solve ambiguity we suggest consulting complementing other wordlists and dictionaries to decide which occurrences of word in ranking order of wordlist can be expected to represent primarily a noun version and its most used meaning. Another challenge is that operating with wordlists is inherently vulnerable since accidental inclusion or exclusion of concept from wordlist on wrong basis can have dramatic biasing effect on the conceptual structure generated based on it and the structure's emphasis on vocabulary offered to convey adoption of new knowledge. Learners of different ages should benefit from having new knowledge introduced to them by relating them especially to their earlier existing and most actively used vocabulary that is most familiar to them.

We do not know any previous work to suggest similar solutions as we do. We suppose that besides proposing framework itself and freely distributable prototype software tool our main contribution is that we have generated and now publish online as open content a diverse collection of pedagogic conceptual networks that offer collaboratively defined key relationships between highest-ranking concepts of learning topic vocabularies and that especially we publish them on large scale reaching even critical level of coverage, i.e. vocabulary size suggested to suffice for general reading comprehension, and separately for different characteristics of learner (age group, gender as well as writing and speaking). To balance conceptual coverage and computational costs to get compact results we limited our analysis of pedagogic conceptual networks to contain at most 2000 highest-ranking word families from wordlists. Both pedagogic conceptual networks and learning paths can be used for various personalized and engaging pedagogic purposes. Operational principles of framework can be easily extended to generate learning paths based on various open web resources beyond Wikipedia, either automatically or on request by learner or educator. Instead of one linear learning path, alternative branching paths can be provided by letting learner to select at each step the link to be traversed next from a set of few alternatives thus contributing to evolution of routing according to her needs. Resulting complementing and interacting knowledge entities each represent an individual perspective to learning topic according to desired criteria based on selections along path, depending on for example moment in history, cultural background and domain of life. By adjusting breadth and depth of branching of learning paths, coverage of topic can be tailored. Diverse alternative branching learning paths as well as visualizations can be used in parallel, iteratively edited and ranked by community of learners thus indicating collaboratively preferred educational resources. The most preferred learning paths are recommendations by community members how to relate concepts about specific learning topic, thus defining local personalized ontologies. From a pre-existing collection of learning paths learner can select one or more paths according to her needs, possibly prioritizing those highly ranked by other learners or educators. This selection process can be used to enable complementing perspectives to learning topic and to define new forms of learning games relying on browsing and editing conceptual relationships.

Future work should heavily invest in rapid agile prototyping with diverse populations of learners in versatile real educational settings to gather large quantities of behavioral data for fine-tuned modeling of intuitive personalized learning practices when learners adopt vocabulary and new knowledge. With increasing penetration of smart phones and tablets through whole society we are living a critical period when the educational market becomes shared with long-lasting dominance by most innovative solutions and public education faces risks to become locked-in to proprietary commercial platforms. So academic community should be now actively involved in coordinating and defining standards that ensure support for sustainable development of educational tools and keep open access and open content on high level of research agenda. Like in our suggested framework, new systems should inherently

have flexible functionality supporting various kinds of educational needs and context, letting learner explore and express her creativity and personal identity. There is a need to develop easily tailorable user interfaces, plug-ins and input devices so that the learners themselves can adjust and select most motivating ways to process knowledge in various forms to be incorporated to learning activities and addressing her background. Adaptive visualization and exploration of knowledge structures should exploit pioneering technology for personalization, for example promising generic input solutions addressing eye/gesture tracking, touch response, EEG bio-feedback, geo-positioning, inertia sensing and image recognition aspects. New tools should promote easy ways to share and collaboratively cumulatively contribute to knowledge building process in learning communities with captivating and inspiring experiences. When learners intuitively invent, form and adopt new educational practices about how to link, agglomerate and traverse pieces of knowledge in their minds there needs to be ways to conveniently document and define these processes for future use and refinement. Like in our framework, new systems should offer recommendations for exploration in educational content on various levels of abstraction with such representation schemes that flexibly support chaining and looping in branching conceptual networks and capturing these exploration patterns into expressive reusable templates. Rich collection of automatically generated and updated templates should be instantly available for typical learning settings but they could be also modified and refined iteratively to address individual personal preference or collaboratively edited and ranked to form mutually agreed standards.

References

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.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? Canadian Modern Language Review 63, 1, 59-82.

Lehr, F., Osborn, J. & Hiebert, E. (2004). Research-based practices in early reading series: a focus on vocabulary. Regional Educational Laboratory, Pacific Resources for Education and Learning. http://www.prel.org/products/re_/ES0419.htm

Nagy, W., & Anderson, R (1984). How many words are there in printed school English? Reading Research Quarterly, 19(3).

Chujo, K. (2004). Measuring vocabulary levels of English textbooks and tests using a BNC lemmatized high frequency word list. In Nakamura, J., Inoue, N., & Tabata, T. (eds.), English corpora under Japanese eyes. Rodopi, Netherlands, 231-249.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In Lauren, C., & Nordman, M. (eds.), Special language: from humans thinking to thinking machines. Multilingual Matters, 316-323.

Hsu, W. (2009). College English textbooks for general purposes: a corpus-based analysis of lexical coverage. Electronic Journal of Foreign Language Teaching. 6(1), 42-62.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. Reading in a Foreign Language, 17(1).

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. Reading Research Quarterly, 20, 522-535.

Karpicke, J. & Roediger III, H. (2007). Repeated retrieval during learning is the key to long-term retention. Journal of Memory and Language 57, 151-162

Cepeda, N., Pashler, H., Vul, E., Wixted, J., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. Psychological Bulletin, 132(3), 354-380.

Anderson, N. (1999). Improving reading speed. English Teaching Forum, 37(2).

DuBay, W. (2004). The principles of readability. Impact Information, Costa Mesa, California, USA.

Doidge, N.(2007).The brain that changes itself: stories of personal triumph from the frontiers of neuroscience.Viking Press,USA.

Harvey, C., & Svoboda, K. (2007). Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature, 450(7173).

Kandel, E. (2001). The molecular biology of memory storage: a dialog between genes and synapses. Nobel Lecture, 8 December 2000. Bioscience Reports, 21(5).

Fields, R. (2005). Making memories stick. Scientific American, 292 (February 2005), 74-81.

Vlach, H., & Sandhofer, C. (2012). Distributing learning over time: the spacing effect in children's acquisition and generalization of science concepts. Child Development, 83, 1137-1144.

Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. The International Journal of Computer-Supported Collaborative Learning, 3(2), 105-122.

Bauer, D. (2007). Learning the semantics of Wikipedia hyperlinks. Bachelor thesis. University of Osnabrück, Germany.

Kuhn, T. (2009). How controlled English can improve semantic wikis. Proc. Fourth Semantic Wiki Workshop (SemWiki 2009).

Lahti, L. (2011). Educational concept mapping method based on high-frequency words and Wikipedia linkage. Proc. Fourth International Conference on Internet Technologies and Applications, Glyndwr University, UK. alldcbblxlbzy wtp vlcyptybl.

Wozniak, P., & Gorzelanczyk, E. (1994). Optimization of repetition spacing in the practice of learning. Acta neurobiologiae experimentalis, 54: 59-62.

Pavlik, P., & Anderson, J. (2008). Using a model to compute the optimal schedule of practice. Journal of Experimental Psychology: Applied, 14(2), 101-117.

Bayetto, A. (2010). The words children write - beyond the early years. Summary report of the Oxford Wordlist, stage 2 research study. Oxford University Press Australia, Australia. http://www.oxfordwordlist.com/pages/search.asp

Leech, G., Rayson, P., & Wilson, A. (2001). Word frequencies in written and spoken English: based on the British National Corpus. Longman, London. http://ucrel.lancs.ac.uk/bncfreq/