A Personal Learning Tool for Second Language Engineering Vocabulary Acquisition

Hua-Li JIAN
Dept. of Foreign Languages and Literature, National Cheng Kung University, No. 1, Ta Hsueh Road, Tainan City 70101, Taiwan, hljian@mail.ncku.edu.tw

Frode Eika SANDNES
Dept. Computer Science, Faculty of Engineering, Oslo University College, P.O. Box 4, St. Olavs plass, N-0130 Oslo, Norway, frodes@iu.hio.no, URL: http://www.iu.hio.no/~frodes/

KEYWORDS: *vocabulary-learning, web-based learning, English for special purpose, student evaluation*

ABSTRACT: *In this paper we suggest how to use Internet material as a source for reading and computer aided assessment of the students' reading performance. The intention is to give students' more freedom and choice when reading, allowing the students to read texts that stimulate their interests. The rationale is that the "joy of reading" is enhanced by pursuing the students' own interests and direction, and the students are more likely to allot more of their time to reading. The students' vocabulary acquisition is assessed through frequent tests. The Internet-based tests are automatically generated such that the workload of the teacher is reduced. The teachers can therefore focus their valuable time on guiding the students rather than painstakingly designing individual tests for each student. Students can also work more independently and obtain immediate feedback from the tool while teachers can closely monitor the students' reading progress.*

1 INTRODUCTION

Reading activities are believed to stimulate vocabulary acquisition when learning a second language. Students either learn vocabulary from the context, or by consulting a dictionary. Vocabulary for special purposes is particularly important in engineering education as English has become the de facto international engineering language, and most non-English speaking countries have at least one English language module. Traditional classroom language teaching typically focuses on a set reading text followed by a set test. The test is carefully designed by the teacher to assess the students' comprehension of the text and understanding of the vocabulary. However, one text is unlikely to catch all the students' enthusiasm. Further, the texts are often non-technical, or technically irrelevant, as the teacher is often a language specialist and not an engineer. In this paper we suggest how to use Internet material as a source for reading and computer-aided assessment of the students' reading performance. Either the students freely select their own texts or they select a text from a diverse predefined reading list. The intention is to give students' more freedom and choices when reading, allowing the students to read texts that stimulate their interests. The rationale is that the "joy of reading" is enhanced by pursuing the students' own interests and directions, and the students are more likely to allot more of their time to reading. The students' vocabulary acquisition is assessed through frequent tests. The Internet-based tests are automatically generated such that the workload of the teacher is reduced. The teachers can therefore focus their valuable time on guiding the students rather than painstakingly designing individual tests for each student. Students can also work more independently and obtain immediate feedback from the tool while teachers can closely monitor the students' reading progress. This paper describes the implementation of our web-based testing tool. First, the tool takes an Internet address or a "cut-and-paste" text as input and extracts the relevant vocabulary in the text. The tool applies data-mining and text-processing techniques to extract the relevant vocabulary from a text and consult current special-purpose online dictionaries. Second, strategies for automatically generating vocabulary and comprehension assessment tests based on this vocabulary are described. Several strategies for language test design are described and also how these strategies can be implemented automatically in the web-based teaching tool.

2 BACKGROUND

Vocabulary is believed to be a key factor in language learning and much has been written about vocabulary (for a theoretic overview see [Coady, J & Huckin, T. 1997] and for a vocabulary research overview see [Coady, J. 1993]. Many strategies for vocabulary learning exist [Ming, T. 1997; Nation, P. 1995] such as glosses [Jacobs, G M. 1994], text-based activities [Joe, A. 1995a], re-telling [Joe, A. 1995b]. A substantial amount of research has especially been devoted to vocabulary learning in the South East Asian countries, especially Chinese learning environments [Sengupta, S. 1996; Wang, Z. 1996] where English teaching pose particular challenges.

Studies have shown that students give up a reading task if the text contains too difficult vocabulary [Grabe, W. & Stoller, F. L. 1997]. Further, several studies have demonstrated the link between vocabulary learning and success in reading [Bossers, B. 1992; Coady, J. 1993]. If students are comfortable with the vocabulary they are more likely to continue the reading activity. Studies by [Hirsch, D. & Nation, P. 1992] shows that students need to know 95% of the words in a text in order to comprehend it comfortably. Further, [Laufer, B. & Sim, D. D. 1992] states that a minimum vocabulary size of academic level English comprises 3000 words, yet [Zechmeister, E. B., D'Anna, C., Hall, J. W., Paus, C. H. & Smith, J. A. 1993] claim that native English-speaking undergraduate students typically master 15,000 words.

Several measures of vocabulary knowledge have been proposed beyond simply counting the vocabulary size. [Paribakht, T.S. & Wesche, M. 1997] proposed a scale based on the ability to apply the vocabulary into a sentence context, [Leung, C. B. & Pikulski, J. J. 1990] apply a simple knowledge factor describing how accurately the student knows the meaning of a word, while [Grabe, W. & Stoller, F. L. 1997] classify vocabulary on familiarity to the user.

Vocabulary learning can be classified as intentional or incidental [Hulstijn, J. H. 1993; Hulstijn, J. H., Hollander, M. & Greidanu, S. T. 1996]. Intentional vocabulary learning is typically the vocabulary acquired explicitly through a language course, while incidental vocabulary learning occurs during independent reading, watching televisions etc.

Obviously, all initiatives leading to the students' increase in vocabulary are commendable. Yet, the incidental vocabulary learning process has been criticised by [Watanabe, Y. 1997] for being inefficient. Some of the reasons for this have been articulated by [Hulstijn, J. H., Hollander, M. & Greidanu, S. T. 1996] which claim that incidental vocabulary learning may give the reader a false understanding of the word, the student may ignore the word and yet understand the overall meaning, the reader may fail to infer the word from the context, new words may not reoccur (repetition essential in the learning process, say 8 to 10 times on average), and learners may fail to connect a word to the new meaning of the word intended in the specific context.

If words are not inferred from the context, the reader will have to consult a dictionary. [Hulstijn, J. H. 1993] has shown that students are more likely to consult a dictionary if the vocabulary is relevant to the goals of the reading tasks. Further, students who consult a dictionary are more likely to remember the word [Knight, S. 1994].

It is natural to assume that [Watanabe, Y. 1997]'s warnings become irrelevant if the students' incidental learning process is continuously monitored, the student is tested and misunderstandings are corrected in a timely manner – especially when the student is deducing the meaning of vocabulary from the context. The problem is that incidental learning is driven by the students' individual interest where the students select their own reading material. The teacher will then have to specifically read the individual texts of all the students, single-handedly evaluate all the vocabulary and prepare individual tests for the students. Obviously, this is not feasible in practice, given a medium or large class.

Testing of vocabulary knowledge and acquisition is another integral area of research [Friel, S. & Johnstone, A. H. 1978; Handy, J. & Johnstone, A. H. 1973; Mobarg, M. 1997]. The most common test form is multiple-choice [Meara, P. & Buxton, B. 1987] and variations on the multiple-choice test [Arnold, J. C. & Arnold, P. L. 1970; Dressel, P. L. & Schmid, J. 1953; Willey, C. F. 1960]. Researchers have also experimented with hypermedia language teaching and online self-assessment testing [Koren, S. 1999], although the materials and tests in this study were fixed and created by the teacher. Web technology provides on many accounts a more flexible and economic learning and testing environment than traditional paper-based methods.

The objective of this work is therefore to develop and test a web-based system that automatically extracts the relevant and new vocabulary from a text that the user selects in electronic form on the Internet, and automatically generates vocabulary tests for the students to take, hence overcoming the difficulties identified by [Watanabe, Y. 1997]. The tests are predominantly multiple-choice tests and are instantly automatically graded giving the student an immediate score and indications of where mistakes were made. The vocabulary, tests and test results are also stored in a database so that the teacher remotely can monitor the reading and vocabulary testing activities of the students. The student can then also be re-tested on previous vocabulary at certain times. The project can be summarised as a computer-aided tool for efficient incidental vocabulary learning.

An automatic vocabulary learning assessment tool for reading activities will have the following benefits: a) Allow students to read self-selected material on the Internet that matches their personal interest, stimulating the students' "joy of reading" and hopefully encourage them to read more. b) Relieve teachers of the tedious burden of designing tests. The teachers' valuable time can be better spent on other activities such as guiding students with particular difficulties. c) Allow teachers to easily and closely follow each individual student's reading activities and progress. d) Allow larger class sizes. e) Allow students to receive immediate feedback on their performance.

2.1 Engineering technology and language learning

A study by [SPRETNAK, M. S. 1982] reveals that engineers on average spend 25% of their effort on writing tasks, 23% reading technical and business material, 11% checking and supervising the writing of colleagues and 7% giving oral presentations. In other words, more than 50% of an engineer's time is spent on communication-related times. However, language and communication training usually occupy a small slot in the engineering curriculum.

Task-based language learning is believed to be more effective than other forms of language learning (see for example [CANDLIN, C. 1987] and [JOHNSON, E. 1997]. Further, reading activities are believed to be more important than writing. As [SPRETNAK, M. S. 1982] puts it, "readers are good writers". In one study by [HEYS, F. 1962], a class which was given only regular reading assignments produced better writing than a class that was given only writing assignments.

2.2 Reading activities - examples from Computer Science

Computer science students can include both their personal reading activities and relevant work-related reading activities into the curriculum of a vocabulary-training course using the proposed tool. Work-related reading activities include reading documentation on the web, which is mostly in English for the most current topic. For example, students working on java projects might have to read either the documentation provided by Sun Microsystems or by one of the many related sites providing technical information such as JGuru, Apache Jarkarta etc. These sites are usually organised around short articles on various related topics. The students can then select relevant articles and add these to the official reading list. Another useful resource frequently exploited by students is discussion lists and FAQs (frequently asked questions). A student may want to add an entire FAQ or parts of a FAQ to the reading manager – capturing the essential vocabulary of the specific problem domain.

Personal reading activities are similar to the work-related reading activities in that text is usually organised into articles. From our encounter with students in the computing laboratories, we find that students spend a long time in reading computer-oriented news. Students of technology often have a personal interest in technology. Undeniably, such reading activities stimulate learning, as reading can help connect the theory from a course to a real world context. Further, reading also accelerates language learning. One of the popular technical websites for technical news is Slashdot. Other students may be interested in specific things such as gaming, computer graphics, audio and computer music etc – which are all typical hobbyist topics with a high degree of relevant technical content.

3 THE VIRTUAL VOCABULARY TUTOR

The virtual vocabulary tutor has several views, namely a reading manager, a test manager and a progress manager. Each view is described in the following sections and the pedagogical benefits are highlighted.

3.1 Authentication and privacy (students and teachers)

Profound to a safe learning environment is data-protection and security. The reading environment of the students is thus protected, i.e. all the information about the students is only visible to the students and the teacher. In order to access the virtual vocabulary tutor the students must enter through a secure point-of-entry that will ask the students to authenticate themselves with a username and a password. Once the students have successfully logged in they will have access to the available facilities of the system, namely the reading selection, test-section and progress selection. The login procedure is identical for both teachers and students.

3.2 Selecting readings (students)

The first action required by the students when first using the system is to add texts to the reading list, and one unique reading list is maintained for each student. First the students must identify suitable texts on the Internet that they want to study. This can for instance be an article about some new computing technology. Alternatively, the teacher can suggest a list of acceptable readings, i.e. a list of URLs (Internet addresses to the texts), which the students select from. The advantage is that the student is making the decision, whether it is a completely free decision or a guided decision. The rationale is that students will put more effort and devotion into studying a text covering material in which the students have an interest.

The students add text to the system by copying and pasting the text directly into a dedicated text edit field. The physical copying of the text has several advantages. First, a text can be collected from a wide range of sources such as web pages, CD-ROMs, word processing documents, PDF-documents or postscript documents. However, one unresolved challenge is that some PDF-documents are copy protected by a mechanism that prevents the students from copying and pasting the text in the document (read only). Second, the students can select parts of the page of interest. This is particularly useful when reading web pages that have a limited portion of textual context. Occasionally, vast amounts of web page space are used for navigational or advertisement purposes, for instance, online newspapers. Third, text spanning multiple web pages can be (manually) merged into one unified text. Automatic traversal of web sites by following hyperlinks is not always straightforward, as it is hard to automatically determine the boundary of the text, i.e. where does the relevant text end? (HTML does not provide any standard mark-up mechanisms for "relevant" text). Finally, some material has a limited life span and source locations change. Such changes are often sudden and unpredictable. One crucial requirement for online-readings is that the texts at least are available for the duration of the course – allowing the material to be revisited. The texts entered by the students are therefore stored in a database to satisfy this requirement.

Reading parameters are immediately calculated, allowing both the teacher and the students to assess the quality and difficulty of a particular text. In particular these parameters include the total number of words in the text, the average number of words per sentence, the size of the vocabulary and a vocabulary difficulty index.

The students will at any time have access to the reading list via hyperlinks to the local copy of the texts. The students will also see if they have accessed (read) the text, how many times and when it was last accessed.

The teacher can also inspect the students' reading lists and evaluate if the material is suitable, both in content and in difficulty level, as well as monitoring that the students have progress in their reading schedule (i.e., the reading statistics for each text is also available to the teacher).

3.3 Sitting a test (students)

The students can choose to be tested on a specific reading. The students will then select the text from the reading list. The tool automatically extracts the vocabulary of interest from the text (described in a subsequent section) and generates a test for the students to take. The performance of the students is recorded and the score is immediately presented as well as the recommended answers, once the test is completed. This is similar to standard online vocabulary tests, but the difference is the dynamic manner in which the test is generated. Each time a student decides to be tested on a particular reading, the student is given an individual and unique test. The test will comprise different vocabulary and different multiple-

choice alternatives, and questions will be presented in pseudo-random order. The tests are presented sequentially, i.e., one question per page.

The students can also sit a "final exam", where the students are tested on vocabulary from their entire reading collection.

3.4 Reviewing vocabulary and annotation (students and teachers)

The assimilated vocabulary can be reviewed at any time. The students may want to inspect the vocabulary that the system has identified as important and their explanations. The students can make amendments to the automatically generated alternatives in situations where a term may have a different meaning than the one described (terminology for special purposes). The students can give a direct translation in their native language and also write their own definition in English. Note that these entries are not checked by the system, but are provided as an aid to the students and also for the teacher to assess the quality of the students' vocabulary understanding. The teacher can choose to make these annotation facilities optional or compulsory as part of their pedagogical strategy.

The teacher may also want to review the students' vocabulary lists to evaluate the suitability of the reading materials and verify that the students are reading texts with an appropriate level of difficulty. A vocabulary list is relatively easy and quick to inspect and can easily be used by an experienced language teacher for providing constructive feedback to the students. If necessary, the teacher can guide the individual student and suggest reading list alterations.

3.5 Reviewing progress (students and teachers)

The students can at any time review their own progress, and the teacher can monitor the same information for every student taking the course. The progress manager produces various statistics regarding the students' activity. These include test scores over time, reading activity over time and overall summary. In addition, the teacher has access to a view that ranks the students according to activity and progress, enabling the teacher to quickly identify passive students that require special attention.

3.6 Student manager

The student manager is used to register new students and remove existing students from the system.

4 TECHNIQUES FOR AUTOMATIC VOCABULARY EXTRACTIONS

The techniques used for extracting vocabulary are inspired from techniques used in web-mining [CHAKRABARTI, S. 2002] and terminology mining based on text corpora (see for example [JUSTESON, J. & KATZ, S. 1995; BOWKER, L.. 1996]. The text is first organised into sentences. Then the words of each sentence are identified. The first step is to filter high frequency words, also known as stop words [WILBUR, J. W. & SIROTKIN, K. 1992; HO, T. K.]. This is achieved using Kilgarriff's English word frequency list [KILGARRIFF, A. 1996; KILGARRIFF, A. 1997]. Then a text dictionary is constructed. The text dictionary lists all the unique words in a text. Thus all duplicates are removed and the material thus shrinks significantly. Word stemming can also be employed to obtain the general form of words [PORTER, M. F. 1980]. For example, the stem of "computers" is "computer".

We assume that the learner is familiar with and master the most frequently used words in the language, such as "is", "are", "the", "in", "on" etc [KILGARRIFF, A. 1997]. The next step is therefore to remove all high frequency words, or stop words, from the text dictionary. Thus, a large number of the entries are then removed. We have no guarantee that words found on the Internet are correct, and the next step therefore comprises spell checking. A dictionary-based spell checking technique is used, that is, a reference dictionary consisting of all the possible valid words in the language in all grammatical forms is used. If an entry in the text dictionary cannot be found in the reference wordlist, then the entry is tagged as a potential incorrect spelling. All entries marked as potential incorrectly spelt words are crosschecked against the reference wordlist using Metaphone [PHILLIPS, L. 1990]. Metaphone is a technique, inspired by SOUNDEX for matching words based on their phonetic sound and is particularly suitable for spell checking applications [KUKICH, K. 1992]. Entries with no Metaphone match in the dictionary are tagged as "potential correct special terms". Words with no metaphone match are very different from all the entries in the dictionary and are likely to be vocabulary for special purposes not present in the dictionary. Entries with a metaphone match are most likely incorrectly spelt words; however, this may not

necessarily always be the case. Note that this tool uses several dictionaries and wordlists for different purposes.

4.1 An illustration of automatic terminology extraction

Imagine a student identifies the following paragraph that he or she wishes to add to the reading list:

> *"The more things change, the more they stay the same. Andrew D. Kirch, security administrator for AHBL, infiltrated several script kiddie groups and shared some of his findings with us via IRC. From the (edited) interview transcript, you'll learn that one of the "new waves" in DDoS coordination is hijacking corporate conference call facilities, which is really an update of good old '60s-style phone phreaking, plus some insight into why some DDoSers do what they do -- and some tips on how they might be stopped"*
>
> *(Taken from "A peek at script kiddie culture" by Robin Miller, Slashdot.org, March 5th, 2004.)*

Further, suppose we have a list including the following common stop words:

> *an, and, be, call, change, do, for, from, good, groups, his, how, in, into, is, learn, might, more, new, of, old, on, one, phone, really, same, several, shared, some, stay, stopped, that, the, they, things, tips, us, what, which, why, with, you'll*

When filtering the stop words from the text we are left with the following terminology:

> *60s-style, AHBL, DdoS, DdoSers, IRC, administrator, conference, coordination, corporate, edited, facilities, findings, hijacking, infiltrated, insight, interview, kiddie, phreaking, plus, script, security, transcript, update, via, waves*

This list of terminology can be reduced further by removing the terms occurring in previously registered readings.

5 AUTOMATICALLY GENERATED TESTS

5.2 Fill in the blanks tests

A commonly used vocabulary test strategy is to present the student with a short text, perhaps one sentence where term to be tested is removed. The student task is therefore to fill in the blanks. One strategy to generate such tests is to use an Internet search engine to search for text with the given vocabulary, either each term in isolation or combinations of terms. The tool then extracts the relevant sentences from the search engine results and replaces the word with a text entry field. The result is presented to the student as a test. This strategy allows the tool with a high probability of success to produce meaningful sentences with the given vocabulary, which the student has not encountered before. Alternatively, to gain more control a specific corpus of text can be used, such as the British National Corpus (http://sara.natcorp.ox.ac.uk/lookup.html) or one of the many books available through the Gutenberg Project (http://promo.net/pg/). One danger of this strategy is the situation where the context of the term in the student's reading does not match the context of the term found by employing the search engine. However, the result can still be educational as it is important to understand multiple meanings of terms.

For example, imagine we wish to test the system on the term "infiltrated" extracted from the text in the previous example. A simple search on Google brings up a large number of alternatives. One of the first ones are from the website of ABC news. We then extract the sentence with "infiltrated" and remove this word and present the result to the user:

> *"Iraqis loyal to Saddam Hussein have _____ US operations and are feeding information to insurgents."*

In this example the term "infiltrated" is used in a war context and not in a computer science context, although computer security can be viewed as an electronic war on hackers.

5.1 Multiple-choice tests

Automatic generation of multiple-choice tests is not trivial. We follow two design principles: a) A test should consist of between 10 to 20 questions in total. If the test is longer the student may be overwhelmed and start avoiding taking the tests. b) For each question there should be between two and five alternatives. Obviously, there must be more than one alternative, and if there are more than five alternatives the tests may be overwhelming.

There are two main categories of multiple-choice tests. First, the word may be given and the student must choose the correct definition. Second, the word definition is given and the student must select the correct word. For a given word, multiple definitions may be generated based on: the actual definition (compulsory), antonyms, synonyms, completely unrelated words, the definition of another word in the set and homonyms. We envisage that the freely available WordNet database for English [MILLER, G. A. 1995] will serve as a valuable resource for automatically generating tests. The WordNet database comprises a large set of words with their interrelationship including hypernyms (generic words). For example, the word "machine" is a hypernym of the word "computer". Similarly, the same strategy can also be applied in the opposite way around for the tests where the student selects the correct word given a definition.

Obviously, a different dictionary is used during the creation of the multiple-choice tests. A simple wordlist will not suffice; therefore an extended dictionary with definition entries for each term is used.

Suppose we wish to create a multiple-choice question for the word "insight" extracted in the previous section:

What is the definition of the term "insight"?

1: Clear or deep perception of a situation
2: A feeling of understanding
3: Present and easily available
4: Try to solve a problem by thinking intensely about it
5: Anything that is seen

Clearly, alternative 1 and 2 are correct definitions of "insight" (source: WordNet). Alternative 3 and 4 are similar definitions based on synonyms (visible and brainstorm), and alternative 5 is a totally unrelated based on a similar sounding word ((in) sight).

6 SUMMARY

This paper describes a novel virtual vocabulary tutor that is currently being developed as a collaboration project between National Cheng Kung University in Taiwan and Oslo University College in Norway. The tutor allows students to select their own reading material. The tool automatically extracts the "difficult" vocabulary and automatically generates multiple-choice tests. Test scores, reading patterns and progress are automatically recorded by the tool. The teacher has full access to these records and can quickly identify the students requiring special assistance and help. The purpose of the project is to stimulate the joy of reading and independent study.

REFERENCES

ARNOLD, J. C. & ARNOLD, P. L., 'On Scoring Multiple Choice Exams Allowing for Partial Knowledge'. *Journal of Experimental Education* 39:1, 1970, 8-13.

BOSSERS, B. 'Reading in Two Languages: A Study of Reading Comprehension in Dutch as a Second Language and in Turkish as a First Language'. Rotterdam: Drukkerij Van Driel. 1992.

BOWKER, L.. 'Towards a corpus-based approach to terminography' Terminology 3:1, 1996, 27-52.

CANDLIN, C. 'Towards Task-Based Language Learning' In Candlin C., & D. Murphy (eds.), Language Learning Tasks. Lancaster Practical Papers in English Education 7. Englewood Cliffs, N.J.: Prentice Hall, 1987.

CHAKRABARTI, S. 'Mining the Web: Analysis of Hypertext and Semi Structured Data' Morgan Kaufmann, 2002

COADY, J. 'Research on ESL/EFL vocabulary acquisition: Putting it in context'. In: T. Huckin, M. Haynes, & J. Coady (Eds.), *Second Language Reading and Vocabulary Learning*. Norwood, NJ; Ablex, 1993, 3-23.

COADY, J & HUCKIN, T. '*Second language vocabulary acquisition: A rationale for pedagogy'*. New York: Cambridge University Press, 1997.

DRESSEL, P. L. & SCHMID, J. 'Some Modification of Multiple Choice Items' *Educational and Psychological Measurement*, 13, 1953, 574-595.

FRIEL, S. & JOHNSTONE, A. H. 'A Review of the Theory of Objective Testing.' *School Science Review* 59:209, 1978, 733-738.

GRABE, W. & STOLLER, F. L. 'Reading and vocabulary development in a second language: a case study'. In: Coady, J. & Huckin, T. (Eds.), *Second Language Vocabulary Acquisition*. Cambridge: Cambridge University Press, 1997, 98-122.

HANDY, J. & JOHNSTONE, A. H. 'How Students Reason in Objective Tests' *Education in chemistry* 10:3, 1973, 99-100.

HEYS, F. 'The Theme-a-Week Assumption: A Report of an Experiment' The English Journal LI:5, 1962, 320-322.

HIRSCH, D. & NATION, P. 'What vocabulary size is needed to read unsimplified texts for pleasure?' *Reading in a Foreign Language* 8:2, 1992, 689-696.

HO, T. K. 'Fast Identification of Stop Words for Font Learning and Keyword Spotting', Proceedings of the 5th Int'l Conference on Document Analysis and Recognition, 1999, 333—336.

HULSTIJN, J. H. 'When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables'. *The Modern Language Journal* 77:2, 1993, 139-147.

HULSTIJN, J. H., HOLLANDER, M. & GREIDANU, S. T. 'Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words'. *The Modern Language Journal* 80:3, 1996, 327-339.

JACOBS, G. M. 'What lurks in the margin: Use of vocabulary glosses as a strategy in second language reading.' *Issues in Applied Linguistics* 5:1, 1994, 115-137

JOE, A. 'Text-based tasks and incidental vocabulary learning.*' Second Language Research* 11:2, 1995, 149-158

JOE, A. 'The value of retelling activities for vocabulary learning.' *Guidelines* 17:1, 1995, 1-8

JOHNSON, E. 'How to Develop Reading-To-Do Skills in Engineering Education' Literacy Across Cultures 1:1, June, 1997.

JUSTESON, J. & KATZ, S. 'Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text', in Natural Language Engineering 1:1, 1995, 9-27.

KILGARRIFF, A. 'Which words are particularly characteristic of a text? A survey of statistical approaches', Proceedings of AISB Workshop on Language Engineering for Document Analysis and Recognition, Sussex University, 1996, 33–40.

KILGARRIFF, A. 'Putting frequencies in the dictionary', International Journal of Lexicography 10:2, 1997, 135-155.

KNIGHT, S. 'Dictionary: the tool of last resort in foreign language reading? A new perspective'. *Modern Language Journal* 78, 1994, 285-299.

KOREN, S. 'Vocabulary instruction through Hypertext: Are there advantages over conventional methods of teaching?' *TESL-EJ* 4:1, 1999, 18 pp.

KUKICH, K. 'Techniques for Automatically Correcting Words in Text', ACM Computing Surveys 24:4, 1992, 377-439.

LAUFER, B. 'Reading in a foreign language? How does L2 lexical knowledge interact with the reader's general academic ability?' *Journal of Research in Reading* 15:2, 1992, 95-103.

LAUFER, B. & SIM, D. D. 'Measuring and explaining the threshold needed for English for Academic Purposes texts. *Foreign Language Annals* 1:8, 1985, 405-413.

LEUNG, C. B. & PIKULSKI, J. J. 'Incidental learning of word meanings by kindergarten and first-grade children through repeated read aloud events'. In: J. Zutell & S. McCormick (Eds.), *Literacy Theory*

and Research: Analyses from Multiple Paradigms. Chicago, Il: National Reading Conference. 231-239. From: Grabe, W. & Stoller, F.L., 1990, 105.
MEARA, P. & BUXTON, B. 'An alternative to multiple choice vocabulary tests' *Language Testing* 4, 1987, 142-154.
MILLER, G. A. 'WordNet: A Lexical Database for English' Communications of the ACM 38:11, 1995, 39-41.
MING, T. 'Innovative ways to teach vocabulary – some practical ideas for the classroom teacher' *Guidelines* 19:1, 1997, 11-21
MOBARG, M. 'Acquiring, teaching and testing vocabulary' *International Journal of Applied Linguistics* 7:2, 1997, 201-222
NATION, P. 'Best practice in vocabulary teaching and learning' *EA Journal* 13:2, 1995, 7-15.
PARIBAKHT, T.S. & WESCHE, M. 'Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition'. In Coady J. & Huckin, T. (Eds.), 1997, 174-200.
PHILLIPS, L. 'Hanging on the Metaphone' *Computer Language* 7:12, 1990, 39-43.
PORTER, M. F. 'An algorithm for suffix stripping' Program 14:3, 1980, 130--137.
SENGUPTA, S. 'Creating a hypertext database to help Hong Kong English language teachers teach writing'. *System* 24, 1996, 187-198.
SPRETNAK, M. S. 'A Survey of the Frequency and Importance of Technical Communi-cation in an Engineering Career' The Technical Writing Teacher 9:3, 1982.
SPRETNAK, M. S. 'Reading and Writing for Engineering Students' JAC 4, 1983.
WANG, Z. 'A contrastive analysis of vocabulary teaching: Australian and Chinese university settings.' *Australian Review of Applied Linguistics* Supplement 12, 1995, 268-280
WATANABE, Y. 'Input, intake and retention: Effects of increased learning of foreign language vocabulary'. *Studies in Second Language Acquisition* 19, 1997, 287-301.
WILBUR, J. W. & SIROTKIN, K. 'The automatic identification of stop words' Journal of the American Society for Information Science 18, 1992, 45-55.
WILLEY, C. F. 'The Three-Decision Multiple-Choice Test: A Method of Increasing the Sensitivity of the Multiple-Choice Item'. *Psychology Review* 7, 1960, 475-477.
ZECHMEISTER, E. B., D'ANNA, C., HALL, J. W., PAUS, C. H. & SMITH, J. A. 'Metacognitive and other knowledge about the mental lexicon: Do we know how many words we know?' *Applied Linguistics* 14:2, 1993, 188-206.