COMPUTER ASSISTED SECOND LANGUAGE VOCABULARY ACQUISITION

Peter J. M. Groot
Utrecht University

ABSTRACT

During the initial stages of instructed L2 acquisition students learn a couple thousand, mainly high frequency words. Functional language proficiency, however, requires mastery of a considerably larger number of words. It is therefore necessary at the intermediate and advanced stages of language acquisition to learn a large vocabulary in a short period of time. There is not enough time to copy the natural (largely incidental) L1 word acquisition process. Incidental acquisition of the words is only possible up to a point, because, on account of their low frequency, they do not occur often enough in the L2 learning material. Acquisition of new words from authentic L2 reading texts by means of strategies such as contextual deduction is also not a solution for a number of reasons. There appears to be no alternative to intentional learning of a great many new words in a relatively short period of time. The words to be learned may be presented in isolation or in context. Presentation in bilingual word lists seems an attractive shortcut because it takes less time than contextual presentation and yields excellent short term results. Long term retention, however, is often disappointing so contextual presentation seems advisable. Any suggestions how to implement this in pedagogic contexts should be based on a systematic analysis of the two most important aspects of the L2 word learning problem, that is to say, selecting the relevant vocabulary (which and how many words) and creating optimal conditions for the acquisition process. This article sets out to describe a computer assisted word acquisition programme (CAVOCA) which tries to do precisely this: the programme operationalises current theoretical thinking about word acquisition, and its contents are based on a systematic inventory of the vocabulary relevant for the target group. To establish its efficiency, the programme was contrasted in a number of experimental settings with a paired associates method of learning new words. The experimental results suggest that an approach combining the two methods is most advisable.

INTRODUCTION

The naive view that the vocabulary of a language should be seen as a "set of basic irregularities" impervious to systematic study, and its acquisition as a haphazard process of learning largely unrelated elements is long outdated. Furthermore, the language teaching profession has come to realise that in foreign language teaching, a grammar-oriented approach is not, to understate the case, the most efficient way to achieve communicative competence. An integrated approach combining systematic attention to the acquisition of both grammar and vocabulary is considered much more effective. This fuller appreciation of the importance of vocabulary teaching gives rise to a number of questions concerning the way in which it should be selected and presented for learning. These questions will be addressed below.

In the early stages of instructed foreign language acquisition[1] students learn a few thousand mainly high frequency words. Such words occur so frequently in the teaching materials to which they are exposed that many are easily acquired. However, a vocabulary of that size, say 2,000 words, is not sufficient for functional language proficiency. To take reading as an example, estimates of the number of words required for understanding non-specialised texts vary (dependent, among others, on what is meant by

"words" and "adequate comprehension") but there is general consensus that 5,000 base words is a minimal requirement (Laufer, 1997; Nation 1990) while for non-specialised, academic reading a considerable larger vocabulary is needed (Groot, 1994; Hazenberg & Hulstijn, 1996). It is therefore necessary that a large number of words be learned in a short period of time at the intermediate and advanced stages of language acquisition. Incidental acquisition of these words is only possible to a point, because they do not occur often enough in the foreign language learning material. Learning new words from authentic L2 reading texts by means of strategies such as contextual deduction is not the answer either, for reasons to be given later. Although there is evidence that retention is better with L1 glosses than without (Hulstijn, Hollander, & Greidanus, 1996; Watanabe, 1997), isolated presentation of the numerous words to be learned in bilingual word lists results in long-term retention that is widely felt to be disappointing. Since the time available for the learning of the large number of new words is limited, it is essential to tackle this problem systematically, both in selecting the relevant vocabulary and in creating optimal conditions for the acquisition process. This article sets out to describe a computer assisted word acquisition programme which intends to do precisely this: the programme tries to systematically operationalise current theoretical thinking about word acquisition, and its contents are based on a systematic inventory of the vocabulary relevant for the target group. The programme (called CAVOCA, an acronym for Computer Assisted VOCabulary Acquisition) was developed over a trial period of several years. Its present database was constructed with the help of a government grant and contains some 500 words specially selected for their difficulty and relevance to the academic reading needs of Dutch university students. In the following paragraphs the theoretical and practical considerations involved in the construction of the programme will be dealt with.

HOW MANY WORDS?

Obviously, a detailed answer to this general question is impossible without a detailed description of the language activity and level intended. Therefore I shall confine myself to a specific example, namely, the vocabulary required for an adequate comprehension of academic reading texts of the type used in the foreign language reading comprehension tests annually constructed by the CITO (the Dutch central educational testing body) for the final exams of Dutch "vwo," an upper level secondary type of school preparing for university studies. These tests comprise a selection of authentic, argumentative and/or popular-scientific L2 texts on a variety of non-specialist topics. They specifically measure L2 reading skills, and comprehension does not depend primarily on textual features such as conceptual or structural complexity, or on reader characteristics such as familiarity with the topic.

To the extent that reading comprehension is dependent on word knowledge, there is empirical evidence (Groot, 1994) that for an adequate understanding of academic texts of this kind, a vocabulary of at least 7,000 words is required (Hazenberg & Hulstijn, 1996 mention an even higher number--10,000). Nation (1993) and Laufer (1997) suggest a target vocabulary of 5,000 as the minimum lexical requirement for understanding general, non-specialised texts. The rationale for these numbers is that only a vocabulary this size will result in a sufficiently dense lexical coverage of texts of this kind. Various studies (Groot, 1994; Hazenberg & Hulstijn, 1996; Hirsh & Nation, 1992; Laufer, 1989) have demonstrated that for adequate comprehension of texts at this level, readers must be familiar with more than 90% of the words used. With such a dense lexical coverage of a text, the percentage of unknown words is so low that, generally speaking, they will either not be essential for an understanding of the text or their meaning may be deduced from the context.

WHICH WORDS?

Apart from the most frequently used 2,000 words there are, a further 3,000 words that should be learned. It is not possible to indicate accurately which words, partly because beyond the first 1,200 words, the frequency of words rapidly decreases and depends greatly on the corpus. Additional selection criteria such as usefulness and valency do not solve the problem either. Every selection will therefore contain a certain degree of arbitrariness as far as inclusion or omission of certain words is concerned. A partial solution to this problem may be compiling a much longer list of words of which only a portion must be mastered (Groot, 1994). The advantage of a list of this length is that difficult choices as to whether or not to include a particular word can be largely avoided. The feasibility of this idea has been studied in relation to English (de Jong, 1998). For this purpose, the subdivision into six frequency levels of the head words listed in the Collins Cobuild Dictionary (1995) was used. Allocation of a particular frequency level to a word is based on an analysis of the 200 million-word "The Bank of England Corpus." Level 1 includes the 700 most frequent words, level 2 the next 1,200 words, level 3 the next 1,500 words, level 4 the next 3,200, level 5 the next 8,100 and level 6 all remaining words. It turned out that the application of a number of qualitative criteria such as relevance and difficulty, and quantitative criteria such as frequency resulted in a list of approximately 8,000 words drawn from levels 3, 4 and 5. Familiarity with any 3,000 words from this list added to the first 2,000 would result in a lexical repertoire of 5,000 words considered sufficient for general reading while command of any 5,000, again in addition to the first 2,000, would suffice for academic reading. Complementary to this approach, various other word lists relevant to reading at this level may also be used in the compilation of such a list (Nation, 1990).

HOW TO TEACH/LEARN THE WORDS?

In connection with word learning, a distinction is commonly drawn between incidental and intentional learning. Unless one narrowly defines incidental learning as excluding any conscious attention to the words being learned (cf. Singleton 1999, p. 274), the two learning modes are not always easy to differentiate and show a considerable overlap, not unlike the acquisition/learning dichotomy suggested by Krashen. In this paper, intentional learning will be used to refer to any learning activity the learner undertakes with the intention of gaining new knowledge. As such it differs from incidental learning where there is no such intention (Anderson, 1990). From a pedagogic perspective, however, the distinction is still useful in a discussion on the optimal way of presenting new L2 words in instructional contexts.

Most words in first language acquisition are learned incidentally in an incremental way because the language learner comes across them frequently in a wide range of contexts (De Bot, Paribakht, & Wesche, 1997; Nagy & Herman, 1987). In a short space of time, a large number of words are thus learned and this lexical repertoire then forms the basis for learning other new words. In the case of foreign language acquisition in instructional contexts, this process is virtually impossible to simulate. The exposure to new words is considerably less intensive and varied.[2] Undoubtedly, a limited number of high frequency words can be learned incidentally but that will certainly not be possible for the much larger number of less frequent words that must subsequently be learned if one wishes to speak of functional proficiency.

To solve this problem it has been suggested that learners be exposed to authentic L2 material and trained in communicative strategies such as contextual deduction of the meaning of new words so that incidental acquisition can take place, thus partially copying the L1 acquisition process (Krashen, 1989). Attractive though this idea may seem, it is not very realistic. Authentic language material is generally not produced with the intention of illustrating to learners the meaning or usage of certain words but rather to convey information to other native speakers who are already familiar with these words. More often than not, it is therefore largely unsuitable for the learning of new words for a number of reasons.

First, because of their relatively low frequency, the words to be learned will occur rarely in the inevitably small authentic L2 input. This means there is not enough repetition for an incremental learning process in which the various features of the words are picked up from the contexts, resulting in a solid embedding in the mental lexicon, as in L1 acquisition.

Second, in authentic use of language, it is frequently not the immediate context of an unknown word that contains the clues to its meaning but wider contexts that cumulatively illustrate its semantic properties. In most instructed L2 learning situations, however, the learner is only exposed to selected passages, which in themselves may not aptly illustrate meaning and use of the particular word at all.

But probably the most important reason why authentic L2 language is inadequate for incidental acquisition (except at highly advanced levels) is that it contains too many other unknown words. Of course, some of these may not be essential for understanding the context. Function words are generally less relevant for comprehension than content words and the same goes for adjectives compared to nouns. But others will be essential and not knowing them will make contextual deduction of the word to be learned problematic. Contextual deduction and, in its wake, incidental acquisition of an unknown word is only possible if the context is well understood and clearly illustrates its meaning. One might say that in such cases, for a proficient reader, the new word is redundant; in other words, it might as well have been left out (as, indeed, it is in cloze tests to measure comprehension of the context). But to the extent that the context contains other unknown words for the learner, there arises what one might call a cumulative reduction of the redundancy of the word in question. The number of possible meanings of the unknown word increases proportionally to the number of other unknown words in the context; the new word may mean "x" if another unknown word means "y," but if this is not the case, "x" must have a different meaning and this puzzle of semantic permutations gets more and more complex with each additional unknown word. The learner must form ever more hypotheses as to the possible meaning and systematically utilise previous and subsequent information to corroborate or refute these. This process will take so much attention and working memory capacity that higher reading processes, which are essential for understanding the context (such as recognition of suprasentential links and discourse markers), are seriously impeded.

The above line of reasoning may be summarised as follows. A thorough understanding of the context is essential for deducing the meaning of an unknown word. For any context to be well understood a dense coverage is required. This means the reader must have "foreknowledge" of most other words in the particular context, which in turn presupposes a large vocabulary. There is a certain irony to this phenomenon (sometimes referred to as the Matthew effect) in the sense that a learner can only pick up new words from authentic contexts if s/he already has a large vocabulary (Horst, Cobb, & Meara, 1998). The above arguments may serve to illustrate the principle that in the limited time available in an L2 teaching context such a large vocabulary cannot be incidentally acquired by dint of sheer exposure to authentic L2 material.

If in instructional L2 situations incidental acquisition of a large vocabulary of lower frequency words through exposure to authentic L2 texts is hardly possible, it follows that efficient acquisition of new vocabulary requires a conscious effort from the learner (Prince, 1996; Sternberg, 1987). There seems to be no viable alternative to intentional learning of a large number of words with the help of authentic L2 material that has been selected (or edited) specifically for this purpose. The limited time available for this huge learning effort makes it imperative that the acquisition process be, as it were, accelerated. This requires a careful analysis of what should be learned and how it should be learned or, in other words, which words should be selected for learning (cf. "Contents of the Programme") and how they should be presented (cf. "Theoretical Background"). A computer assisted word learning programme which intends to do this is described below.

CAVOCA

Theoretical Background

CAVOCA (Computer Assisted VOCabulary Acquisition) is a computer programme for vocabulary acquisition in a foreign language. It has been designed on the basis of generally accepted theories about the way the mental lexicon is structured and operates. Allowing for certain differences between the various theories on how words are learned, stored in, and retrieved from the internal lexicon (cf. Aitchison, 1995), there is general agreement that in a natural (L1) word acquisition process several stages may be recognised. They cannot always be clearly distinguished because learning a word is an incremental process that gradually develops with repeated exposure and because there is constant interaction between the various stages. However, for clarity's sake, they will be briefly described as if they were separate stages independent of one another.

1. Notice of the various properties of the new word: morphological and phonological, syntactic , semantic, stylistic, collocational, and so forth.

2. Storage in the internal lexicon in networks of relationships that correspond to the properties described in (1).

3. Consolidation of the storage described in (2) by means of further exposure to the word in a variety of contexts which illustrate its various properties. This results in a firmer embedding in the memory needed for long term retention.

Adequate implementation of the stages described above will result in a solid embedding of the word in the mental lexicon, which is necessary for efficient receptive and productive use. If one of the stages is neglected, the word will not properly fix itself in the internal lexicon and will be stored only superficially without the many associations and links with other words needed for efficient lexical retrieval. The learner will not or barely recognise the word in a reading or listening text and will certainly be unable to use it in speaking or writing. These ideas about the importance of an intensive processing of the new word were first presented in a systematic fashion in Craik and Lockhart's (1972) "levels of processing" theory. It postulated that "rates of forgetting are a function of the type and depth of encoding" information and distinguished between various levels of processing. Thus, in their view, processing semantic properties of a word represented a deeper level than just processing its phonological features. Certain aspects of their theory have been criticised (especially its inability to clearly define the differences between levels in operational terms) but it has since led to a general consensus among researchers that there is a stringent relationship between retention and intensity or elaborateness (Anderson, 1990) of processing lexical information about a new word (i.e., paying close attention to its various features such as spelling, pronunciation, semantic and syntactic attributes, relationships with other words, etc.). Important elements in this intensive processing are the variability (Anderson,1990) and specificity (Tulving & Thomson, 1973) of the encoding activity. This theoretical position appears to have several important pedagogic implications for the teaching/learning of new words.

The first is that exposure to words in context is preferable to exposure to words in isolation. Only contexts will fully demonstrate the semantic, syntactic, and collocational features of a word the learner has to process in order to establish the numerous links and associations with other words necessary for easy accessibility and retrieval (see also Nation, 1990, and Singleton, 1999, for a summary of the arguments and evidence supporting this position).

Another implication, although more controversial than the first, appears to be that having learners infer the meaning of new words from the context is a better way to safeguard elaborate, intensive processing than giving the meaning because of the greater cognitive effort required.

Mondria (1996) presents evidence that seems to refute this theoretical stance. He interprets his finding that vocabulary test scores for the two conditions (given vs. inferred meaning) indicated that there is no difference in long term retention effects between the two presentation methods and that, in teaching new

words, giving the meaning is a more efficient method than having learners contextually infer it, because it takes less time. His conclusions, however, are based on scores of tests of receptive knowledge only (a multiple choice and an open ended test) in which subjects were asked to recognise the target words. Whether tests of productive use (in which subjects have to recall the word themselves) would have yielded the same results leading to the same conclusions is doubtful (cf. the first remark in "Discussion").

The natural word acquisition process (as this occurs in first language acquisition) consists of gradual acquisition of the various properties of a word through repeated exposures in a wide range of authentic contexts illustrative of its various features. Bearing this in mind, we are faced with a dilemma in an instructed L-2 learning situation. On the one hand, there is not enough time for exposure to new words of the same intensity as in L1 acquisition. On the other hand, superficial exposure leads to shallow processing which fails to establish enough associations and links with other words for solid storage and efficient retrieval. Obviously, there is no easy solution to this dilemma. The most realistic approach seems to be to create an environment that is maximally conducive to learning new words by striking a balance between the two contradictory demands. The CAVOCA programme intends to do just that by speeding up the acquisition process; it takes the learners systematically through the various stages by exposing them to carefully selected L2 material which illustrates the salient features of the new L2 word and/or the differences between the L2 word and its nearest L1 equivalent or counterpart.

The Programme

The stages of the vocabulary acquisition process described above are operationalised in the various sections of the CAVOCA programme. The programme takes the learner systematically through the sequence of mental operations which make up the acquisition process. The word to be acquired is presented in contexts selected in such a way as to ensure an efficient and, as it were, condensed acquisition process. To secure learner involvement, the programme is interactive: at certain points the learner has to make choices ("What do you think the word means?" "Is the word correct/appropriate in this context?" "What is the word that is missing in this context?") and is given feedback by the computer. The current CAVOCA programme presents the words in modules, each consisting of 25 words and taking about 50 minutes to complete. The programme covers each word in four sections which embody the various stages of the word acquisition process.

The first two stages of the vocabulary acquisition process, learning the word's various properties (among which, most importantly in a L2 acquisition context, is its semantic properties, see Singleton, 1999, p. 189) and storing the word in the memory are operationalised in the first section of the programme, called "Deduction." The word to be learned appears on the screen for a few seconds. Next, it is used in three sentences, presented in order of contextual richness. The first sentence contains only a few clues as to the meaning of the word and mainly serves to draw the learner's attention to its morphological composition, spelling, syntactic function, and so forth. The second sentence contains more clues as to the meaning, and the third is so contextually rich that the meaning becomes entirely clear. Every sentence is followed by a multiple choice question to be answered by the learner with four options as to the possible meaning, the correct alternative being a (near) synonym. After each sentence the learner is given immediate feedback (whether the meaning s/he inferred was right or wrong) to avoid the wrong meaning from being retained. After the third presentation of the word, the key to the multiple choice item is given as final feedback for the learner. To a certain extent, this way of presenting new words may seem unnatural since in natural word acquisition first contexts need not but may very well contain clues to the meaning of an unknown word. It was nevertheless opted for to make learners process the word intensively by forcing them to form and test hypotheses as to its meaning. The word is presented three times in sentences containing ever more semantic clues and the learner has to deduce the meaning in stages. This method of presenting the new word is meant to trigger off a cognitive process of what might be called "graded contextual disambiguation"; step by step the learner reduces the uncertainty about the meaning of the word by

making use of the contextual clues increasingly present in the three consecutive sentences. It should yield better long term retention results than simply giving the meaning because it enforces a deeper level of processing (Mondria & Wit-de Boer, 1991). Here is an example. The word to be learned is "abrasive."

Figure 1. "Deduction"

The second and third context sentences in this example are:

> 2. He was offended by her abrasive tone of voice.

> 3. His abrasive criticism undermined her confidence and made her doubt herself.

The second section of the programme ("Usage") is geared to the second stage of the word acquisition process, consolidation. To further secure the word's position in the mental lexicon and to further illustrate its exact meaning, two sentences are presented in which the word is either used correctly/appropriately or not. The learner chooses and the computer gives feedback, explaining why the use of the word in question in that particular context was correct or incorrect. Also, whenever relevant, additional information about the word is given: other meanings (or, as in the example given here, the original, literal meaning of the word), derivatives, similar or misleadingly similar words, idiomatic usage, and so forth. In this section the learner is also requested to type the word in order to reinforce storage of the word's morphological properties. The computer points out and corrects any mistakes. An alternative version of the programme on CD-ROM gives the pronunciation of the word. The learner is then asked to repeat it and his/her

pronunciation is recorded so that it can be listened to and compared to the correct pronunciation. The theoretical rationale for this multi-modal presentation is its supposed positive effect on the retrievability of a word. A diversity of operations to be performed vis a vis a word is likely to lead to better storage of a word and ,as a result, more (efficient) retrieval routes (Chun & Plass, 1996; Gathercole & Conway,

Usage

abrasive
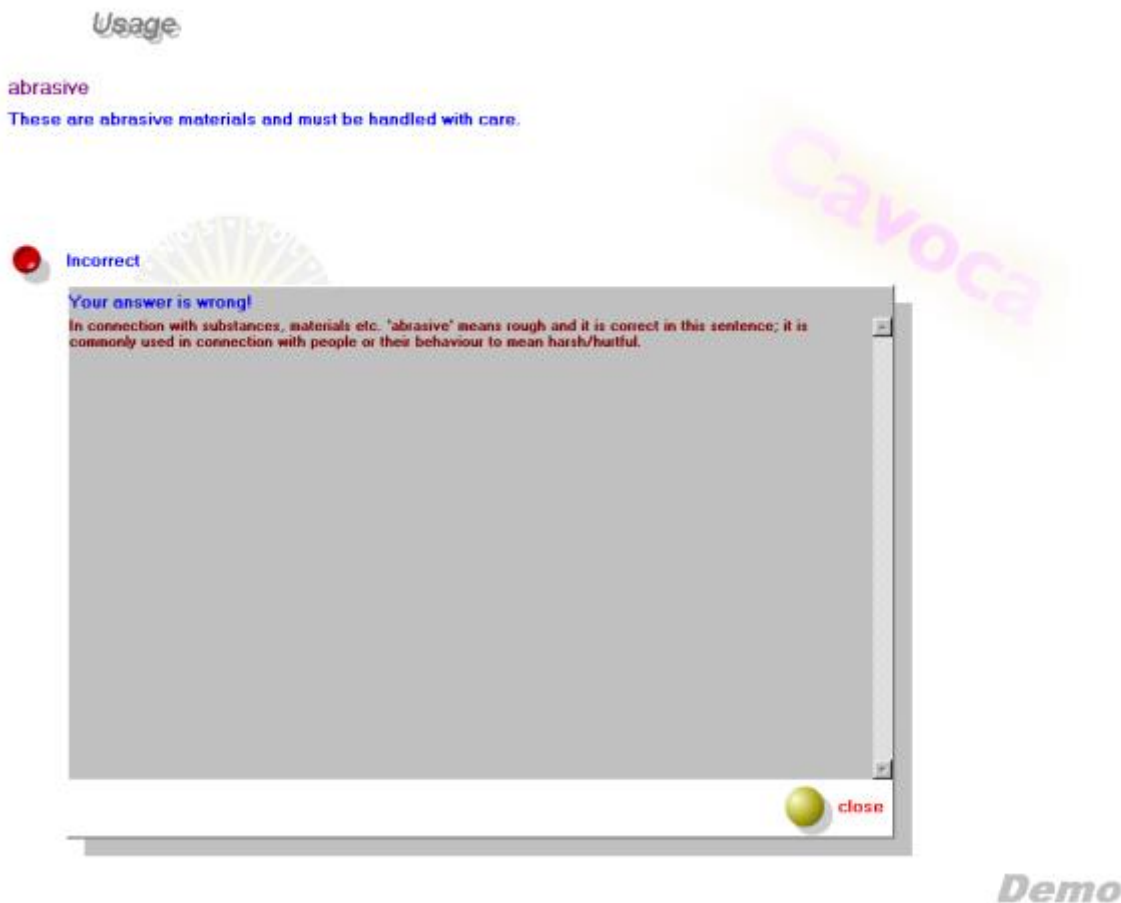These are abrasive materials and must be handled with care.

● Correct

● Incorrect

? Help

▮ Exit

Sentence 1 of 2 - Please, select an alternative.   *Demo*

1988).

Figures 2 and 3. "Usage"

The third section of the programme, "Examples," is likewise designed to reinforce consolidation and thus ensure long term retention. The learner is presented with a number of authentic L2 passages selected from large databases containing the word just learned. These passages have been specially selected to clearly illustrate both meaning and use of the word in question. An additional objective of this section of the programme is to increase the learner's motivation for learning words (or, to put it more realistically, to motivate them at all). The learner recognises that he/she (better) understands the authentic L2 passage thanks to the recently acquired knowledge of the word learned. Hopefully, in the learner's mind, this experience will serve as a specific illustration of the general principle of the importance of vocabulary for understanding authentic L2 reading texts.

Figure 4. "Examples"

Once the learner has dealt with each of the 25 words in the module in the manner described above, he/she comes to the fourth and final section of the CAVOCA programme, called "Lexical Retrieval." In this section, which also serves as a self-assessment test, the learner's active knowledge of the word is elicited. The learner is presented with 25 sentences, each with one word missing. These sentences have been selected specially so that the blank can be filled by one word only (i.e., one of the words covered in the module). To help the learner and to elicit specifically the word recently acquired, the first letter of the word belonging in the sentence is given. Once the 25 sentences have been completed, the learner's score appears on the screen and any mistakes are pointed out. Print-outs enable the teacher to check the student's performance in each module.
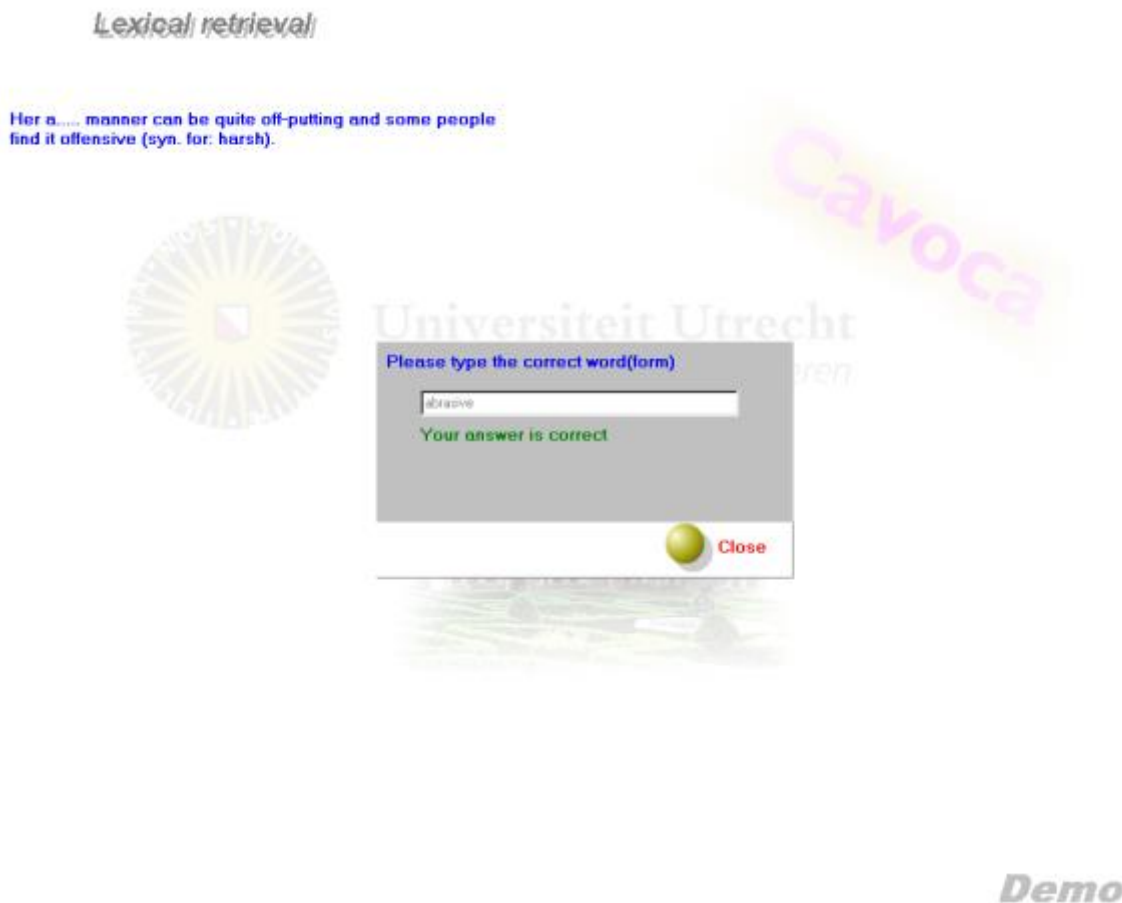
Lexical retrieval

Her a..... manner can be quite off-putting and some people
find it offensive (syn. for: harsh).

Please type the correct word(form)

abrasive

Your answer is correct

Close

Demo

Figure 5. "Lexical Retrieval"

Contents of the Programme

The words in the current CAVOCA database have been selected to fulfil two criteria: relevance and difficulty for the target group (Wijbenga, 1997).

To select a body of words relevant to first year university language proficiency courses, a preliminary list of several thousand words judged relevant to "academic reading" was put together based on frequency and a number of other considerations of a contrastive/linguistic and didactic nature. Seven experienced teachers of English from Dutch universities were presented with this list and were requested to put each word into one of five categories according to relevance. Subsequently, a list of 1,500 words with the highest mean score and the lowest standard deviation was made, in other words, all the words which were judged relevant and suitable by all or most of the teachers. From these 1,500 words a selection of about 500 difficult words was made, based on a contrastive analysis. This selection encompassed words which lack a Dutch equivalent or a one-to-one relationship with their Dutch counterparts in terms of usage or meaning. Examples are *acknowledge, encroach, fumble, enhance, oblivious,* and *anxious*, words for which there is no direct (context-independent) translation in Dutch. However, words like *abduct* (which denotes the exact same concept as the Dutch *ontvoeren* and which is used in the same way syntactically, stylistically, etc.) were not selected for the CAVOCA treatment. Such words receive a less intensive treatment (what we named the EDIT treatment: Extended DIctionary Treatment) on the assumption that words of this kind are easier to learn. In the EDIT treatment, the word is presented in one or two contextually rich sentences followed by a definition, derivatives, words related in form but not in

meaning, and so forth. A database of words from the higher frequency band width described in "Which Words?" relevant for intermediate stages of L2 acquisition, is under construction.

THE EFFICIENCY OF THE PROGRAMME

The CAVOCA programme is based on a theoretical analysis of the L1 word acquisition process and, in a sense, tries to replicate the various stages of this process, albeit in a condensed form to save time. It sets out to speed up the word acquisition process by means of intensified exposure to carefully selected L2 material. Thus, it fulfils the theoretical and practical conditions for efficient word learning. There are, however, a number of differences between L1 and L2 word acquisition. In L1 acquisition, the new word and its meaning are learned simultaneously, while in L2 learning the concept covered by the L2 label is either familiar to the learner (when the two labels cover semantic equivalents) or can be integrated into his/her conceptual framework. This difference alone justifies the question as to whether an intensive method of word learning like CAVOCA is efficient compared to less time-consuming methods such as "paired associates" learning (e.g., via bilingual word lists), efficiency being defined here as the ratio between the number of words learned and the time needed to learn them. One might argue that because the learner is already familiar with the concepts covered by the new L2 labels, the conceptual learning load in L2 word acquisition is lighter than in L1 acquisition so that intensive processing of the new L2 words is not essential for retention, even in the case of L2 words lacking one-to-one relationships with L1 counterparts such as those selected for the CAVOCA database. Another ground for comparing the two methods is that the CAVOCA method represents a way of learning new words which is very unlike what most students are used to. It takes more time per word than a bilingual list, students are not given a translation but have to work out the meaning for themselves, and all of the context material and the feedback is in the L2. In short, it is a much more difficult method than the familiar paired associates learning methods that they are used to.

In order to collect evidence relevant to this question, the CAVOCA approach was compared with the more orthodox approach of L2 word learning by means of bilingual word lists in a number of experimental settings. A detailed report of the experimental procedure and data would exceed the scope of this article and has appeared elsewhere (Groot, 1999) but the results most relevant to an evaluation of the efficiency of the CAVOCA approach will be discussed.

Experimental Procedure

The experimental (CAVOCA) and the control condition (bilingual lists) were compared in four experiments (Bonte, 1997; Dufour, 1997; Janssen, 1996; Nep, 1998). These experiments had a quasi-experimental, pre-test/post-test, differential treatment design, with the learning method as the independent variable and the scores on the post tests as the dependent variable. Subjects, ranging from upper level secondary school (vwo) pupils (aged 16 to 18) to first year university students (aged 19-20), were presented with two equivalent sets of words, one in the experimental and the other in the control condition, in two separate learning sessions of the same length. The words were selected according to the contrastive linguistic criteria described in "Contents of the Programme" and assignment of the words to either of the two conditions was random. In all experiments the effect of the two methods was measured twice: immediately after the learning session and two to three weeks later to determine the long-term retention effect. Subjects had not been told about the delayed test to prevent them from paying more than usual attention to the words after the learning session, which might invalidate the results. Due to their relatively low frequency, the chances that they would come across the test words in the period between the immediate and the delayed test were slim. Prior to the learning session, a pre-test was administered containing more words than the final set used in the experiments to check whether subjects were familiar with any of the words. This turned out to be the case in a few instances and these words were not used in the experiment.

In experiments 1 and 2, carried out by Janssen and Dufour (the second being a replica of the first in order to establish the generalisibility of the results), the effect of the two methods was measured by means of a test of receptive knowledge. The subjects were shown the two sets of words learned and asked to give a translation or definition. Obviously, this testing method favours the control condition since the method used in the testing session is the same as the one followed in the learning session (see Schneider, Healy, & Bourne, 1999, p. 89, and the observation in "Theoretical Background" about encoding specificity). Subjects need only remember the translation of the control condition words to achieve a high score. For the words used in the experimental condition, this direct association was not possible since no translation was provided.

Two follow-up experiments (3 and 4), carried out by Bonte and Nep (again, the second replicating the first to determine the generalisibility of the results), were set up in the same way in all other respects as the first two, except for the testing technique used to measure the effect of the two methods of word learning. To establish to what extent the scores obtained were the result of the particular testing method used in these experiments (or, in other words, to determine the constraints on their validity), a different testing technique was used, namely a cloze test. This testing format obviously measures more than just receptive knowledge of words since the word itself is not given but has to be provided by the testees themselves. This form of lexical retrieval clearly requires a deeper knowledge of a word than receptive knowledge. The context sentences used in the cloze tests were not the same as those used in the "Lexical Retrieval" part of the CAVOCA programme so that subjects could not come up with the target words because they recognised the sentences. Since pre-tests showed that subjects found this way of testing much more difficult than the receptive tests and to preclude them from filling in a semantically acceptable alternative word, the first letter of the word was given.

C	control method
X	experimental method
test 1	immediate test
test 2	delayed test
1-2	decrease in scores on immediate and delayed tests
mean	mean score
SD	standard deviation
ss	number of subjects
R	reliability (Cronbach alpha)
nax	maximum score

Table 1. Experiment 1: First year university students

	C test 1	C test 2	C 1-2	X test 1	X test 2	X 1-2
mean	74.67	49.96	24.85	55.46	36.42	19.03
SD	0.86	15.56		18.17	15.24	
ss.	14	14		14	14	
R	0.198	0.999		0.982	0.974	
max.	75	75		75	75	

Table 2. Experiment 2: vwo students

	C test 1	C test 2	C 1-2	X test 1	X test 2	X 1-2
mean	39.04	23.16	15.87	26.76	18.35	8.41
SD	1.94	8.36		7.19	7.33	
ss.	24	24		24	24	
R	0.219	0.974		0.913	0.916	
max.	40	40		40	40	

Table 3. Experiment 3: vwo students

	C test 1	C test 2	C 1-2	X test 1	X test2	X 1- 2
mean	12.65	5.00	7.65	9.60	7.00	2.60
SD	2.3	2.5		2.4	2.6	
ss.	15	15		15	15	
R	0.93	0.84		0.90	0.88	
max.	24	24		24	24	

Table 4. Experiment 4: vwo students

	C test 1	C test 2	C 1-2	X test 1	X test 2	X 1-2
mean	15.3	11.3	4	17.5	15.3	2.3
SD	1.8	2.1		2.1	2.4	
ss.	29	29		29	29	
R	0.87	0.85		0.83	0.86	
max.	25	25		25	25	

The reliability coefficients of nearly all the tests used were satisfactory (>0.80), which is not surprising for discrete point tests. The immediate receptive knowledge tests in the first two experiments were an exception: these showed a low reliability due to a clear ceiling effect in the scores resulting in a low standard deviation.

Significance levels were calculated for the following two most relevant results of the four experiments: a) the difference between the mean scores for both conditions on the delayed tests and b) the difference between the decrease in the mean scores on the immediate and delayed tests for both conditions.

The two-tailed t-tests for independent means resulted in levels never higher than $p < .05$. There was one exception for a) in experiment 3 ($n=15$), which was only significant at a level of $p < 0.10$

The experimental results consistently show certain patterns, independent of the subjects or the words used in the experiments.

1. In the first two experiments the scores on the immediate tests of receptive knowledge were considerably higher for the control condition. Recall of the fresh association between the words and their translation, as established by the bilingual word list, was sufficient for a high score. In fact, in both experiments the mean scores on the immediate tests of receptive knowledge of the words learned by means of bilingual lists were extremely high (> 95%). As observed above, this strategy of pairing associates resulting in high scores could not be applied in the case of the words learned through the CAVOCA programme, since in this condition no translation was provided and the meaning had to be worked out by the subjects themselves. Of course, subjects may have tagged their own L1 labels onto these (L2) concepts but, since no feedback in the form of the

correct translation was given, it is unlikely that these individual L2-L1 associations were always wholly correct or, if they were, as firmly established as those in the control condition.

2. As was expected, the scores on the delayed tests in these two first experiments were considerably lower for both conditions. Retention loss as manifested in the decrease in scores on the delayed test was larger for the bilingual word list method than for the CAVOCA condition: 24.85 versus 19.03 (= 33% vs. 25%) and 15.87 versus. 8.41 (= 39% vs.21%) in experiments 1 and 2, respectively, but the mean scores on the delayed tests of receptive knowledge were still higher for the word list condition than for the CAVOCA method.

3. In the last two experiments, where the effect of both methods was measured with cloze tests, the mean scores for both conditions on the immediate tests did not show the large differences observed in the first two experiments. However, as in the first two experiments, the decrease in the scores on the delayed tests was larger for the bilingual word list condition than for the CAVOCA condition: 7.65 vs. 2.60 (=31% vs.10%) in experiment 3 and 4 vs. 2.2 (=16% vs. 9%) in experiment 4, resulting in higher scores for the experimental condition.

4. If the figures found in the four experiments for the decrease in scores on the immediate and the delayed tests mentioned above in (2) and (3) are converted into forgetting rate percentages (i.e., the percentage of the words learned that was forgotten during the period of time between the immediate and the delayed tests), we get the following results for the control and the experimental condition in the four experiments: 33% versus 34% in the first experiment , 40% versus. 31% in the second , 60% versus 27% in the third, and 26% versus 13% in the fourth. With the exception of the first experiment, these figures confirm the retention loss pattern observed above in (2) and (3).

Discussion

For a correct interpretation of the above data three preliminary remarks are called for.

1. The experiments were carried out to determine which of the two methods of learning new words is more efficient in the sense of yielding the best long term retention results. The crucial question is then "When has a word been learned?" or, in other words, "What does it mean to know a word?" Clearly, as observed before, there are various levels of or dimensions to word knowledge (Nagy & Herman, 1987). Knowing a word may be seen in operational terms as a continuum ranging from vague recognition of its spelling to (semantically, syntactically, stylistically) correct and contextually appropriate productive use. Retrieval of a word from the mental lexicon for productive use requires a higher degree of accessibility or, in other words, a more solid integration in various networks than is needed for receptive use. For measuring this higher level of mastery, a test which asks testees to simply recognise a word and give its meaning is unsuitable; a test using the cloze technique, which measures testees' ability to produce the word themselves, is much more valid for that purpose. The experimental results reported in "Results" clearly demonstrate that for a meaningful interpretation of the data, it is essential to give an accurate description of what one understands by the trait "knowing a word" and of what trait is intended to be measured by what testing method.

2. The scores on the tests administered after the experimental learning session do not pretend to show the learningeffect of each separate part of the CAVOCA programme but rather the overall effect of the CAVOCA induced learning process as a whole. As observed in "Theoretical Background" in the description of the theoretical background of the programme, it is difficult, from a psycholinguistic

perspective, to discriminate between the various stages in the word acquisition process. All one can logically say is that there must be a temporal order: noticing a word's properties must of necessity precede any storage and consolidation can only follow when something has been stored. But even if it were desirable from a theoretical point of view, it is practically hardly possible to determine the relative contribution to the learning effect of the various stages in the learning process.

3.   It is not unlikely that a lack of familiarity with the CAVOCA method of word learning negatively influenced the scores in the experimental condition. This way of learning was completely new to the subjects. It was intuitively felt to be useful but also much more difficult than the more orthodox approach with its facile association between the L2 word and its L1 translation. Experimenters observed again and again that subjects are, as it were, conditioned for superficial learning and find it difficult to switch to a different style.

If long-term retention is the ultimate goal of learning new words, little significance should be attached to the extremely high scores for the control condition on the immediate receptive tests in the first two experiments. Considering the sizable fall in the scores on the delayed tests, these high scores possess no predictive value whatsoever with regard to long term retention of the words. The immediate tests measured superficial recognition of the words that had been presented in the bilingual list, automatically triggering fresh associations between the L2 and the L1 words. It is common knowledge that high ability learners in these age groups possess an admirable memorising capacity (Hulstijn, 1997; Knight, 1994). This enabled the subjects to achieve extremely high scores on the immediate tests. The associations, however, are not firmly established and two weeks later most of them are beyond recall.

It is not unlikely that the higher scores on the delayed tests in the first two experiments for the control condition should be attributed for a substantial part to the fact that for the control condition the method of learning and testing were identical (cf. the observation in "Experimental Procedure").

The higher scores on the delayed cloze tests and the smaller loss of retention for the experimental condition in all four experiments may be regarded as corroborative evidence for the theory that there is a strong relationship between retention rates and depth of processing. They appear to indicate that intensive processing of new words leads to a more solid embedding and better long term retention which is needed for active use of the words, than does superficial processing of the words out of context with the translation given, as in bilingual lists.

On the other hand, the higher scores on the delayed receptive tests for the control condition in the first two experiments point to the conclusion that, even where L2 word learning cannot be equated to just relabeling familiar L1 concepts (as was the case in the experiments described above where the L2 target words did not have a direct L1 equivalent), high ability learners at high L2 proficiency levels achieve receptive command more efficiently with the help of bilingual lists than with the CAVOCA method. Whether this also holds for L2 word learning at lower levels in lower age groups is a moot point. One might argue that high level learners have meta-cognitive strategies at their disposal which make their acquisition of new vocabulary much less dependent on externally imposed learning conditions (such as the intensive CAVOCA presentation that tries to copy the L1 word acquisition process in a condensed form) than is the case for younger, low level learners whose less developed cognitive maturity makes their L2 acquisition process more similar to L1 acquisition. The data reported here do not warrant any conclusions regarding this issue.

As to the significance of the above results from a pedagogic L2 teaching perspective, they strongly suggest that a combined approach, making use of the two methods simultaneously, is probably the most efficient. On the one hand, using bilingual word lists would fully profit from the L1 conceptual

framework, especially where the L1 and L2 labels are near equivalents in meaning and use, the effect being enhanced if these lists take a form such that they stimulate the learner to establish more than the superficial, minimalist associations between the L2 and L1 labels often attributed to this way of presenting new words. On the other hand, such an approach would yield better chances for long term retention due to the intensive processing of the words in the form of the various mental actions to be performed on them such as those offered by the CAVOCA programme.

In what order, proportion and form the two methods should be incorporated in the dual approach is an open question. One possibility would be to present the new words first in the CAVOCA programme immediately followed by a bilingual word list presentation. Further experimentation will have to provide data as to the efficiency of this particular way of combining the two methods. Of course, whatever the outcome of our endeavours to find the optimal mode of presenting new words, repeated exposure at certain intervals is essential for long-term retention. It is highly improbable that one learning session, however intensive, is sufficient. Modern technology offers unique possibilities for rehearsal practice that will ensure further consolidation. Using a concordance programme for finding the word in question, learners may be instructed to search in large electronic databases of authentic L2 texts for examples of the words just learned which best illustrate their meaning and usage. An exercise such as this refreshes the learner's awareness of the word, its meaning and how it is used. It is a useful exercise which also enables the teacher to assess whether the learner has retained the words in question. Also, repeated exposure to the recently learned words in short texts, in combination with words that frequently co-occur with them in authentic L2 material (either because they belong to the same semantic field or because they are linked up in standard phrases, collocations or idiomatic expressions), will stimulate further consolidation.

CONCLUSION

The CAVOCA computer programme is an attempt to operationalise theoretical ideas about word acquisition. As such, it is an instrument which enables us to empirically verify the theory on word acquisition in general and its validity for L2 word acquisition in particular. If it yields data incompatible with what the theory predicts, either the theory is partially incorrect (e.g., where it claims a basic similarity between L1 and L2 word acquisition), or there is something wrong with the way it has been operationalised in the CAVOCA programme. If the data collected with CAVOCA are in accordance with the theoretical predictions, they may be regarded as a confirmation of the theory. The evidence reported above may be regarded as a first indication that theories about word learning are correct in the importance they attribute to intensive processing for long term retention. But the data also indicate that there are marked differences between the L1 and the L2 word learning process. In particular, the fact that the L2 learner already has a system of conceptual categories at his disposal to accommodate the new L2 labels may imply that L2 word learning represents a simpler cognitive task than L1 word acquisition where new concepts and labels have to be learned simultaneously. To the extent that this is indeed the case the question arises whether attempts such as the CAVOCA programme to make L2 word learning a condensed copy of the L1 word acquisition process are cost effective, especially in the case of L2 words that have equivalent L1 counterparts. In such cases a simple bilingual presentation followed by some rehearsal practice may be more efficient. The overall conclusion must be that there is no simple answer to the key question what form the most efficient method of L2 word learning should take. It depends very much on variables like degree of L1-L2 equivalence of the words to be learned, the intensity (both qualitative and quantitative) of processing, the age and cognitive level of the learner, the quantity and quality of rehearsal practice etc. More experimentation systematically controlling these variables is needed to gather data that will provide more insight into their relative importance. Instruments like CAVOCA may help provide such data.

NOTES

1. In this paper, the acronym L2 will be used to include second and foreign language learnings as opposed to L1 learning.

2. Singleton (1999, 236) estimates one year of natural exposure to be the equivalent of 18 years of classroom exposure.

ABOUT THE AUTHOR

Dr. Groot is senior lecturer at the University of Utrecht (The Netherlands) and currently teaches and supervises research in second language acquisition. His research interests are: L2 testing, word acquisition and reading.

For a demo version of the programme, please e-mail the author at Peter.Groot@let.uu.nl

REFERENCES

Aitchison, J. (1995). *Words in the Mind.* Oxford: Basil Blackwell.

Anderson, J. R. (1990). Cognitive psychology and its implications. New York: Freeman.

Bonte, W.F.W. (1997). *Vocabulaire Verwerving. Een vergelijking van twee methoden* [Vocabulary learning. A comparison of two methods]. Unpublished master's thesis, Universiteit Leiden, The Netherlands.

Chun, D. M., & Plass, J.L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80,* 183-197.

Collins Cobuild English Dictionary. (1995). London: Harper Collins.

Craik, F., & Lockhart, R.S., (1972). Levels of Processing. A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11* 671-684.

De Bot, K., Paribakht, T.S., & Wesche, M.B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition, 19,* 309-329.

Dufour, M.J. (1997). *Foreign Language Vocabulary Acquisition: Two Methods Compared.* Unpublished master's thesis, Universiteit Utrecht, Faculty of Letters, The Netherlands.

Groot, P.J.M. (1994). Tekstdekking, tekstbegrip en woordselectie voor het vreemde-taalonderwijs (with a summary in English) [Lexical coverage, reading comprehension and wordselection in foreign language teaching]. *Toegepaste Taalwetenschap in artikelen, 3,* 111-121.

Groot, P.J.M. (1999). Computer ondersteunde vreemde-taalverwerving op de hogere niveaus (with a summary in English) [Computer assisted foreign language learning at higher levels]. *Toegepaste Taalwetenschap in Artikelen, 1,* 111-126.

Hazenberg, S.,& Hulstijn, J.H. (1996). Defining a minimal receptive second language vocabulary for non-native university students: an empirical investigation. *Applied Linguistics, 7,* 145-163.

Hirsch, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8,* 689-696.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11,* 207-223.

Hulstijn, J. (1997). Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady & T. Huckin, *Second Language Vocabulary Acquisition* (pp. 203-224). Cambridge: Cambridge University Press.

Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80,*(3), 327-340.

Janssen, A.E.P. (1996). CAVOCA. Computer-assisted Vocabulary Acquisition. Unpublished master's thesis, University of Utrecht, Faculty of Arts, The Netherlands.

Jong, A.P.H. de (1998). Between Abandon and Zoom. A Quantitative Study of Required VWO Vocabulary. Unpublished master's thesis, University of Utrecht, Faculty of Arts, The Netherlands.

Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78,* 285-299.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73,* 440-464.

Laufer, B. (1989). What percentage of lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special Language: from humans thinking to thinking machine* (pp. 69-75). Clevedon, England: Multilingual Matters.

Laufer, B. (1997). The lexical plight in second language. In J. Coady, & T. Huckin, *Second Language Vocabulary Acquisition* (pp. 20-34). Cambridge: Cambridge University Press.
Mondria, J.A. (1996). *Vocabulaireverwerving in het vreemde-talenonderwijs. De effecten van context en raden op de retentie* [Vocabulary acquisition in foreign language teaching. The effects of context and guessing on retention]. Doctoral dissertation. The Netherlands: Groningen University Press.

Mondria, J.A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied Linguistics 12,* 249-267.

Nagy, W.E., & Herman, P.A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisitio* (pp. 19-35). Hillsdale, NJ: Erlbaum.

Nation, I.S.P. (1990). *Teaching & Learning Vocabulary.* Boston: Heinle & Heinle.

Nation, P. (1993). Vocabulary size, growth and use. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon (*pp. 115-134). Amsterdam: Benjamins.

Nep, W. (1998). Computer ondersteunde woordverwerving in de tweede fase [Computer assisted vocabulary acquisition at higher levels]. Unpublished master's thesis, University of Utrecht, IVLOS [Teacher training department], The Netherlands.

Prince, P. (1996). Second language vocabulary learning: the role of context versus translations as a function of proficiency. *The Modern Language Journal, 80,* 478-493.

Schneider, V.I., Healy, A.F., & Bourne, L.E. (1999). Contextual Interference Effects in Foreign Language Vocabulary Acquisition and Retention. In A.F. Healy & L.E. Bourne (Eds.), *Foreign Language Learning* (p. 89). Hillsdale, NJ: Lawrence Erlbaum Associates.

Singleton, D. (1999). *Exploring the Second Language Mental Lexicon.* Cambridge: Cambridge University Press.

Sternberg, R.J. (1987). Most vocabulary is learned from context. In M.G. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89-105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Watanabe, Y. (1997). Input, intake and retention: effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition, 19,* 287-307.

Wijbenga, A. (1997). Compiling a Dutch University Word List. Unpublished master's thesis. University of Utrecht, Faculty of Arts, The Netherlands.