



### **The classical research in vocabulary acquisition.**

**Paul Meara** *Swansea University*

A few years ago it was fashionable to describe vocabulary acquisition as a neglected aspect of language learning. Recently, however, interest in this area has unexpectedly grown at an enormous rate. There seems to be a general feeling among teachers, publishers and researchers that vocabulary acquisition has not been treated seriously enough in the past, and that our beliefs about how people acquire vocabulary in a foreign language are due for overhaul. At times like this, what generally happens is that people working in the field rediscover work that was carried out many years before. Papers which have remained in peaceful obscurity for many years suddenly start to be quoted in bibliographies, and become part of the collective folklore. Since 1980, or so, this tendency has become increasingly apparent in current research on vocabulary acquisition.

This paper is a summary of "major" empirical research on vocabulary acquisition carried out between 1900 and 1960. It is not a definitive survey, because such a survey would be beyond the scope of an article of this length. My working definition of "major" is any empirical study which has been cited at least twice in the current literature (i.e. since 1980) as part of a theoretical argument about vocabulary acquisition. This literature is comprehensively reviewed in Meara (1987) and in Meara (1992). This criterion gives us a set of 13 papers. The article provides a summary of all this work, and a critical analysis of some of the issues these studies point to. I hope that my reading of this research might highlight some areas in which these experiments need to be treated with caution, and how some of the more obvious misinterpretations might be avoided.

The 13 papers to be discussed are summarised in Tables 1 and 2. Table 1 identifies the author(s) and date of each paper, the type and number of subjects who took part in each experiment, the language of the stimuli, and the number of words the subjects were required to learn. Table 2 reports the type of experimental task used to measure acquisition and a brief summary of the findings.

**Table 1**

13 major studies in vocabulary acquisition: basic materials. details in parentheses are inferred from the text :b indicates that the subjects were absolute beginners.

<i>Author</i>	<i>date</i>	<i>Subjects</i>	<i>Target Lg</i>	<i>No of Target words</i>
<b>Thorndike</b>	1914	28 adults :b	German and Nonsense	4*20
<b>Grinstead</b>	1915	1 adult	German	(300+)
<b>Seibert</b>	1927	81 students	French	(3*30)
<b>Anderson and Jordan</b>	1928	31 children :b	Latin	(10*25)
<b>Stoddard</b>	1929	328 children :b	French	2 * 25
<b>Seibert</b>	1930	60 adults	French	12
<b>Chapman and Gilbert</b>	1937	121 children :b	Hindustani	48
<b>Forlano and Hoffman</b>	1937	65 children :b	Hebrew	2 * 20
<b>Morgan and Bailey</b>	1943	84 adults :b	Ru-ro	(120)
<b>Morgan and Bonham</b>	1944	148 children :b	Ru-ro	(6 * 20)
<b>Morgan and Folz</b>	1944	58 children :b	French	(100)
<b>Kopstein and Roshal</b>	1954	76 adults :b	Russian	8
<b>Kopstein and Roshal</b>	1955	96 groups :b	(Russian)	8

**Table two: methods and results****Thorndike 1914**

Thorndike compared two ways of learning words in listed pairs: a) the repetition method, where each list was simply read aloud until learning was achieved, and b) the recall method, where the list is read through once, and the subject then covers up one member of each pair and guesses the other. Thorndike recorded the time taken to learn the lists, and the number of correct responses on a test. Repetition is faster and produces slightly better test results, but these differences are not significant.

**Grinstead 1915**

Grinstead compared a) learning words from context with b) learning words from lists. In (a) the single subject read a text and listed any unknown words. In (b) a list of words was presented and the subject deleted from this list any words he already knew. The unknown words were then looked up in a dictionary and later retention of their meanings was tested. Method (a) produces a

Meara 1996

## **Table 2 (continued)**

small advantage over (b).

### **Seibert 1927**

Seibert compared 3 ways of learning word lists: a) silent learning, b) learning aloud and c) learning aloud accompanied by an immediate written recall of the words learned. In all methods, Seibert required subjects to learn the lists by rote, and the time necessary for this was roughly recorded. All subjects performed each task over an extended period; relearning trials and tests were administered after 2, 10, and 42 days. Seibert claims that methods (c) and (b) produce fastest relearning, and that accuracy was greatest when items were learned aloud.

### **Anderson and Jordan 1928**

Anderson and Jordan taught sets of different types of Latin words to 12yr old children, and looked at rates of learning and forgetting. Words were learned in lists of about 25 words, one list a day, with 15 seconds allowed for each word. Learning was measured by a recognition test. A&J report that cognate words are learned better than non-cognates or semi-cognate words, and that single words are learned better than phrases. About half the items learned are still retained after two months, but cognate and semi-cognate words hold up best; these stimuli produce a shallow U-shaped forgetting curve. Children who learn lots of words retain them better than those who learn fewer words.

### **Stoddard 1929**

Stoddard compared the learning of word lists presented as a) English-French pairs or as b) French-English pairs. 20 minutes was allowed for the entire task. Subsequent testing showed that French-English order produced better learning.

### **Seibert 1930**

Seibert compared acquisition of vocabulary in a) paired associate lists, b) in sentence contexts and c) in a combination of both of these. Recall was tested after 50 minutes, 2 days, 15 days and 40 days. Seibert reports that (a) is consistently superior to the other methods, and that this superiority is maintained over the 40 day test period. (cf. also Seibert 1945).

### **Chapman and Gilbert 1937**

Chapman and Gilbert tested the claim that it is easier to learn a foreign language word if you also know what the word means in your native language: i.e. if you do not know the meaning of **tompson**<sup>1</sup> in English, learning its Russian translation will be hard. C&G paired known and unknown words with random Hindustani words, and taught these pairs to the children by running through the entire list in a number of successive presentations with a test after each run. Results, not surprisingly, showed that known words always produced better learning than unknown words, and that the familiar words are less likely to be forgotten.

### **Forlano and Hoffman 1937**

Forlano and Hoffman compared two methods of teaching word lists. In the telling method, each word is read aloud by the teacher, and the children are then told what it means; in the guessing

**Table 2 (continued)**

method, each word is read aloud and the children have to guess its meaning. They are then told the true meaning, and correct their guess if it was wrong. Retention of meaning was tested immediately after learning, and again after two days. The telling method produces better learning.

**Morgan and Bailey 1943**

Morgan and Bailey used a set of artificial stimuli based on French to test the claim that words could be learned effectively in context. Subjects were provided with either a) a simple story, or b) a set of decontextualised materials. Both groups were provided with a dictionary and asked to produce a translation of their material. They were later tested on their ability to produce English translations given the Ru-ro word forms. The results failed to show any significant differences due to the type of material studied.

**Morgan and Bonham 1944**

Morgan and Bonham investigated the effects of part of speech on vocabulary acquisition, by requiring learners to learn lists of 20 word pairs in English and Ru-ro. The lists contained 3 nouns, 3 pronouns, 3 verbs, 3 adjectives, 3 adverbs, 3 prepositions and 2 interjections. Each word pair was exposed for 1.5 seconds, and after all 20 words had been seen, a test was administered. This procedure was repeated until all twenty words could be recognised three times in succession. Number of exposures required by each word was recorded. Significant differences between word types were found: Nouns were easiest, adverbs hardest.

**Morgan and Foltz 1944**

Morgan and Foltz's experiment is a replication of Morgan and Bailey, but using French words in place of Ru-ro. Results showed no significant differences between groups.

**Kopstein and Roshal 1954**

Kopstein and Roshal compared teaching words a) with pictures and b) in printed form. A recognition test was used to assess learning. Better learning was found with (a).

**Kopstein and Roshal 1955**

Kopstein and Roshal compared two methods of teaching foreign words. In a) foreign word and English word were displayed simultaneously for 3.5 secs; in b) the foreign word was displayed for 2 secs and then simultaneously with the English word for 1.5 secs. Method (a) was very much better than method (b) at first, but the difference disappeared as more substantial levels of learning are reached.

**DISCUSSION**

Although these experiments deal with a number of quite different aspects of vocabulary acquisition, there are a few common themes which deserve comment. These are, at the simplest level: subjects, languages, target language words to be learned, and the tasks used to

assess learning; at a more complex level, the important issue seems to be the role that models play in research.

### **subjects**

A glance at Table one reveals that there is an enormous discrepancy in the number of subjects taking part in these experiments. Such discrepancies are not peculiar to vocabulary studies, of course, but they are fairly important in this area, since we know that there are very large individual differences in the way people handle words (cf. for example Hunt:1978). In the studies I have listed, the number of subjects ranges from Grinstead's single subject, to Kopstein and Roshal's 738. None of the papers gives any explanation for the number of subjects used, and there is clearly a general feeling that the number is arbitrary as long as it is obviously sufficient. This is a rather surprising conclusion. Ideally, what you want is a number of subjects big enough to iron out the variation due to individual differences, but ironically in these studies, the experiments with the largest number of subjects make no attempt to control for extraneous variables that might affect vocabulary up-take. Stoddard, for example, mentions that his subjects produced huge amounts of variation, their scores ranging from 2 to 50 items on a 50 item test, but he does not attempt to explore the sources of this difference.

A second, but in some ways more important problem is that in most of the experiments the subjects are naive in the sense that their prior experience of the language they were supposed to be learning was nil. This is an important limitation, because it means that these experiments deal in effect only with the very earliest stages of learning a language. It could be argued that acquiring a few words in a language that you are wholly ignorant of, is quite different from acquiring the same number of words in a language you know moderately well. In these later stages, a learner will already have developed a good feel for the formal aspects of words in the target language. This should reduce the learning burden considerably and make it easier to acquire the target language words, the more proficient the learner is. At the same time, morphological information and comparisons with known words of similar meaning should also make it easier to fix the meaning and form of a target language word. Considerations of this sort obviously restrict the way that the results reported here can be applied to other types of learners. In effect, only three of these studies dealt with non-beginners, and of these the two studies by Seibert deal with "81 students who had the equivalent of one year of college work in French" (Seibert 1927: p296) and "sixty college students in second year French" (Seibert 1930: 299). It is not very clear what sort of level this description represents, but it is probably fair to guess that the level is not very advanced. This leaves Grinstead's study as the only one which reports on the behaviour of advanced subjects, and this study deals with only a single individual. The limitations of this data base are very clear, and it obviously going to be very difficult to make statements about "vocabulary acquisition" in general on the strength of these reports.

### **languages**

The same sort of limitation is apparent in the range of languages studied in these experiments too. A number of the studies use target languages which are totally artificial. Morgan and Bailey's Ru-ro consisted of 114 artificial words with a syntax that was an exact parallel of French. Chapman and Gilbert technically used Hindustani, but in fact they merely gave their children lists of English words paired randomly with Hindustani words; the word

pairs were NOT translations. Forlano and Hoffman used Hebrew words and their English translations. The remaining studies use TL words from four closely related Indo-European languages. In spite of these restrictions, all the authors make a point of extrapolating from their rather limited data to general, universal statements about learning foreign vocabularies. This is an important simplification, because it ignores the fact that different types of language present quite different learning problems to individual learners. Take for instance, the cases of a Dutch speaker, a Spanish speaker, an Arab and a Vietnamese learning English. By and large, the Dutch speaker will find basic English vocabulary easy, since most of it is cognate with items in his own language. He might have problems with less frequent vocabulary, but by the time he gets to that stage, he probably has reached a high level of independence and autonomy anyway. In contrast, a Spanish speaker will generally find basic English vocabulary difficult: it is structurally very different from basic vocabulary in Spanish, and there are few cognates. However, Spanish speakers have a huge latent vocabulary of low frequency English words which are cognate with Spanish items, and this should mean that their ability to acquire new words improves dramatically with their general level of competence in English. The Arab and the Vietnamese speakers have no such help from their L1, and the process of acquiring new words will never get any easier for them. At the same time, however, these two learners will find English vocabulary difficult in different ways because of the way their L1 lexicons are shaped and structured.

### **words to be learned**

Just as we saw that there was a discrepancy in the number of subjects used in the experiments, there is a similar discrepancy in the number of words to be learned. The figures here range from Kopstein and Roshal's amazingly low figure of eight words to Grinstead's more ambitious target of 300plus. Most of the studies cluster in the 20-40 word range, though sometimes, the subjects are asked to learn a number of lists of this length. Again, however, there is no justification for why these figures are chosen, or whether such a figure is an appropriate one. The basic problem is that all the authors are assuming it is possible to model the acquisition of an entire vocabulary by looking at how effectively a tiny subset of this vocabulary is acquired in tightly controlled conditions.

There are a number of obvious reasons why this position is untenable. Firstly, learning a set of 20-40 words may pose some difficulties for short-term memory, but seen from a long-term perspective, and in comparison with the number of words a fluent speaker needs to know, such numbers are basically trivial. Many people can handle a vocabulary of a few tens of words by using simple mnemonic techniques, for example. It is not obvious, however, that these techniques would enable a learner to handle, say, two thousand new words - the number of words you need to handle about 80% of English text. Secondly, and more importantly, a vocabulary of 30-40 words can be efficiently handled by treating it as an unconnected list of discrete items. Bigger vocabularies on the other hand will contain subsets of words which are linked together either on semantic or morphological grounds, and these linkages must make it inefficient to treat the vocabulary as a simple list. At the very least some sort of network structure must develop in a large vocabulary which reflects these relationships between the component items of the total vocabulary. Presumably, what makes it difficult to acquire a large vocabulary is that it takes time and effort for these connections to develop, and for a properly organised

lexicon to emerge. This problem does not arise when the target lexicon contains only a handful of words.

### **learning methods**

The studies we have been discussing make use of a highly restricted set of learning methods. Most of them use the time-honoured method of learning the target language words in lists, paired off with their L1 translations, and there are a number of variations on this theme. In most cases, however, the method is treated in a rather rigid fashion: what is measured is the number of trials, or the total time required to learn the entire list of words, so that the learners are allowed no flexibility in their learning. Anderson and Jordan's 15secs per word is a good example of this approach. The other method used is to present words in contexts, but apart from Grinstead's real life contexts, the studies reported here all used highly artificial contexts. In addition, since most of the learners were absolute beginners, it is difficult to see why the contexts might be expected to help very much in any case. Morgan and Bailey's contexts, for example, consisted of a story made up of 114 totally unknown words in a language whose syntactic structure was also unknown. This means that the new words may have been related to each other through the situations described, but it is hard to imagine that this context could provide any real support to the learner, and it is not really surprising, therefore, that M&B fail to find any significant differences in their study.

It is easy to explain this concentration on lists of isolated words in historical terms, but the use of a single presentation method limits this research in a very basic way. There are two reasons for this. Firstly, we know from studies of "good language learners" that effective study of vocabulary often involves the use of many different learning methods (cf. for example, Naiman, Frölich, Stern and Tedesco: 1975). Good learners rarely rely on list learning as a way of increasing their vocabulary; rather, they actively seek out new words and incorporate them into their personal word stock using a variety of learning techniques. The second reason why list learning is unsatisfactory is that though it may be effective for small numbers of words, it is much less obviously effective as a way of learning large vocabularies. In these experiments, for example, Stoddard allowed his subjects twenty minutes to learn twenty five words: i.e. roughly one word per minute. Even with this generous time allowance, and with a fairly loose set of criteria for accepting a word as acquired, Stoddard's subjects managed to acquire only fifteen words. If you allow for subsequent forgetting of a proportion of these words, say 50%, then the effective learning rate is something like 3 minutes per word. At that rate, it would take 50 hours of study time to acquire a vocabulary of 1000 words - not a great deal of time objectively, perhaps, but a very long time when judged against the amount of time typically available to language learners. Experiments that are restricted to the learning of very small numbers of words clearly mask this basic problem.

### **assessment**

Much the same criticisms can be made of the way in which the subjects' grasp of the vocabulary was tested in these experiments. All the studies reported simply ask for the Target Language word to be translated into English, and this means that even in the experiments where words were initially learned in contexts, only the ability to recognize decontextualised words was measured. It is not obvious to me that this measure is a good test of how well vocabulary items have been

learned. At best it tests passive recognition skills rather than active acquisition of items; at worst it tests passive recognition of one item out of a set which has only just been the studied - this means that recognition of any characteristic which distinguishes the word from the rest of the set would be enough to give a correct answer. Testing in this way gives no indication of whether a particular word can be put to active use, or whether some partial knowledge might have been acquired which could facilitate learning in future encounters. Furthermore, this kind of testing gives no indication of how resistant the might be to forgetting or to confusion with other words, both problems which increase as the number of words to be learned gets larger.

## CONCLUSION

Although I have been ostensibly discussing work on vocabulary acquisition carried out some thirty or forty years ago in this paper, readers who are familiar with the contemporary literature will recognise that many of the comments I have made apply equally well to work that has appeared more recently (cf. Meara 1983, Meara 1987 and Meara 1972).

In some ways, we seem to have made very little real progress since Grinstead. None of the questions asked by the researchers listed in Table One has been definitively resolved, and it is perhaps worthwhile asking why there is so little sign that any of the work carried out since 1960 has had any serious impact on course design or teaching practice.

My own view is that a lot of current research is making the same mistakes that can be found in the earlier work on vocabulary acquisition. The main problem areas that we have noted in the early work are still apparent today.

The first, and perhaps the most important of these problem areas is that none of us has any idea of the extent to which individual differences affect vocabulary acquisition, though we do know that individual differences in L1 vocabulary skills are very large indeed. This means that we still do not have a motivated reason for using experimental groups of a particular size, and we do not know how far our results are generalisable to wider populations. A recent collection of studies on "the use and acquisition of the second language lexicon" (Gass 1987), which I will take as representative of current experimental practice includes 7 papers, where the number of subjects ranges from 15 to 244. Again, there is no principled explanation for these numbers apart from availability, and one is left wondering whether it is really necessary to do experiments with 244 subjects in one condition, and just how much variation you would expect to find in a group as large as this. Conversely, one might ask whether a group of 26 subjects comprising 5 Arabic speakers, 1 chinese, 1 Farsi speaker, 1 Greek, 1 "Indonesian", 5 Japanese, 1 Portuguese, 9 Spanish speakers, 1 Thai and 1 Turk can really be considered representative of anything at all (Ard and Gass: 1987). The short answer to this question is that we don't know, of course. There is, however, plenty of evidence that vocabulary handling skills in L1 vary enormously, and it seems very likely that the same sort of variation will apply in the case of L2 learners. However, until we know the extent of this variation, and how it is made up, we are unlikely to come up with convincing models of how vocabularies are acquired. It seems to me that the question of how much individual variation there is in vocabulary skills really needs to be made a top priority in L2 vocabulary acquisition research.



The second problem area is that most of the current research still looks at a very restricted range of languages, and generalises from this narrow base to vocabulary acquisition in general. As in the case of the early research, almost all the current work is based on Indo-European languages, despite the fact that cognate vocabularies seem relatively easy to learn, and that non-Indo-European languages are known to cause special problems in the area of vocabulary acquisition. As an example, consider Gass's 1987 collection again. Here, the seven papers cover French and English bilinguals, Dutch learners of English, Swedish learners of English, and groups of mixed subjects learning English or learning Hebrew - if anything, a rather narrower spread of languages than we found in the classical research, with a heavy emphasis on the acquisition of English as an L2. This narrow spread in itself is not a problem, but the concentration on the acquisition of English must distort the field. English vocabulary is very peculiar: it seems to be inordinately large compared to the basic vocabularies of other languages, for instance, and compared to other languages, English seems to rely less on systematic combinations of items, and rather more on discrete items. In terms of Cruttenden's distinction between **items** and **system**, English seems to have more items and less system than many other languages do (cf. Ringbom 1983).

There are in fact, very few studies which make comparisons between learners from different backgrounds acquiring the same L2, and very few studies of learners from the same L1 background acquiring different L2s. To my knowledge, there is for example no work on the acquisition of Chinese or Arabic vocabularies by English speakers, and only a handful of papers on the acquisition of English vocabulary by native speakers of non-Indo-European languages. There are two notable exceptions to this claim. A series of studies comparing the English vocabulary of Finns and Swedes shows without exception that Finns have to work much harder at acquiring English vocabulary than Swedish speakers do (cf. Takala 1984, Ringbom 1987). There is also a major cross-linguistic study of migrant L2 learners, where systematic comparisons between learners from the same L1 background acquiring different L2s were carried out (Broeder et al. 1988). This work is difficult to evaluate because the number of subjects in each comparison group is tiny (usually 3), and this takes us back to the question of individual differences and variation at the lexical level.

What we really need, then, is a systematic program of replications, where studies carried out with small, homogeneous groups of learner are repeated with other groups with different characteristics. At the moment, there is no sign of a systematic program of this sort.

The third similarity between current work and the early research is a general assumption that acquiring words in an L2 is basically the same process, regardless of the stage the learner is at. A great deal of current research deals with volunteer subjects who learn a set of words in laboratory experiments, and are therefore effectively absolute beginners. Rather more work on advanced learners is being carried out, but this work still concentrates on how well a particular set of words are acquired in learners at a given level of proficiency. Few, if any attempts have been made to assess how the overall structure of the L2 lexicon develops, or how the size of a learner's lexicon affects the way he acquires new words. It is easy to see why this should be, of course. There are no obvious ways of describing complexity in the lexicon, or how a small lexicon differs from a large one (other than by its size). It is much easier to think of the

lexicon as an unstructured list of words, rather than as a complex, interlocking structure. This is clearly a problem that we will have to address if we are to understand what is really going on in a developing lexicon.

Finally, we still do not seem to have made any real advances in the measurement of vocabulary acquisition. A number of more recent papers (notably Richards 1976), have discussed what "knowing a word" means. Unfortunately these discussions have not yet led to any working models or tests which might be used in psycholinguistic experiments, and most of the current research still looks at vocabulary as discrete items which can be marked correct or incorrect on the basis of simple recognition or production tests. The simple binary model which underlies these tests is a very crude one; it doesn't accord well with most learners' personal experience and it seems inherently incapable of allowing us to develop a sensitive model of what learning words in an L2 really involves.

At the same time, the modern literature on vocabulary acquisition includes a huge number of different experimental tasks: lexical decision tasks, word association tasks sentence completion tasks, and so on. In theory this diversity should allow us to "triangulate" on the really important issues in vocabulary acquisition. In practise, it seems to me that it contributes to a serious fragmentation of the field. There are too few of us working in this area to explore fully what each of these different techniques does. The data each of us produces is often quite incompatible with data other people produce, and this makes it very difficult to put all our findings together into a single coherent picture.

A solution to both these problems might be found if we could develop a series of standardised tasks, or agreed "bench marks" which we could use to assess the acquisition of words. It ought to be relatively easy to agree on what these bench marks should be. They would need to be technologically simple; they would have to be usable with learners from different L1 backgrounds, and usable with learners at different levels of proficiency; but, at the same time they would need to be sufficiently rich and sensitive to reflect the richness of the real world. In short, a challenging combination of real world constraints and rich theory. This may sound like a tall order, but if we cannot develop standard, reliable tools of this sort, then in the long run we will continue to produce data whose relevance is strictly limited. Until we begin to tackle these problems systematically, we are likely to continue covering much the same ground as the classical researchers did, and with equally unconvincing results.

## NOTES

1: **tompion** - "a small pellet of mud and saliva which a bear inserts in its anus before hibernating for the winter to stop the ants getting in".

## ACKNOWLEDGEMENTS

Thanks to Angela Lilley, Sue Gass and Jan-Arjen Mondria who provided detailed comments on earlier drafts of this paper.

Meara 1996

## REFERENCES

**Anderson, J and A Jordan.** 1928.

Learning and retention of Latin words and phrases. *Journal of Educational Psychology* 19, 485-496.

**Chapman, I and L Gilbert.** 1937.

A study of the influence of familiarity with English words upon the learning of their foreign language equivalents. *Journal of Educational Psychology*, 28, 621-628.

**Forlano, G and M Hoffman.** 1937.

Guessing and telling methods in learning words in a foreign language. *Journal of Educational Psychology*, 28, 632-636.

**Grinstead, W.** 1915.

An experiment in the learning of foreign words. *American Psychologist*, 9, 407-408.

**Hunt, E.** 1978.

The mechanics of verbal ability. *Psychological Review*, 85, 109-130.

**Kopstein, F and S Roshal.** 1954.

Learning foreign vocabulary from pictures vs. words. *American Psychologist*, 9, 407-408.

**Kopstein, F and S Roshal.** 1955.

Method of presenting word pairs as a factor in foreign vocabulary learning. *American Psychologist*, 10, 354.

**Meara, PM.** 1983.

*Vocabulary in a second language*. London: Centre for Information on Language Teaching.

**Meara, PM.** 1987.

*Vocabulary in a second language: vol 2*. London: Centre for Information on Language Teaching.

**Meara, PM.** 1992. .

*Vocabulary in a second language: vol 3. Reading in a Foreign Language* 9,(1992), whole volume.

**Morgan, C and D Bailey.** 1943.

The effect of context on learning vocabulary. *Journal of Educational Psychology*, 34, 561-565.

**Morgan C and DN Bonham.** 1944 .

Vocabulary learning affected by parts of speech. *Journal of Educational Psychology* 35(1944), 369-377.

**Morgan, C and M Folz.** 1944.

The effect of context on learning a French vocabulary. *Journal of Educational Research*, 38, 213-216.

**Naiman, N, M Frölich, H Stern and A Tedesco.** 1975.

Meara 1996

*The Good Language Learner*. Toronto: Ontario Institute for Studies in Education.

**Nation, ISP.** 1990.

*Teaching and Learning Vocabulary*. New York: Newbury House.

**Richards, J.** 1976.

The role of vocabulary teaching. *TESOL Quarterly*, 10, 77-89.

**Ringbom, H.** 1983.

On the distinction of item learning vs. system learning and receptive vs. productive competence in relation to the role of L1 in foreign language learning. In: **H Ringbom** (ed.) *Psycholinguistics and Foreign Language Learning*. Åbo: Åbo Akademi. 1983.

**Ringbom, H.** 1987.

*The role of the first language in foreign language learning*. Clevedon, Avon: Multilingual Matters.

**Seibert, L.** 1927.

An experiment in learning French vocabulary. *Journal of Educational Psychology*, 18, 294-309.

**Seibert, L.** 1930.

An experiment on the relative efficiency of studying French in associated pairs vs. studying French words in context. *Journal of Educational Psychology*, 21, 297-314.

**Seibert, L.** 1945.

An experiment on the practice of guessing word meanings from context. *Modern Languages Journal*, 29, 305-314.

**Stoddard, G.** 1929.

An experiment in verbal learning. *Journal of Educational Psychology*, 20, 452-457.

**Takala, S.** 1984.

Evaluation of students' knowledge of English vocabulary in the Finnish comprehensive school. Jyväskylä: Jyväskylän Yliopisto.

**Thorndike, E.** 1914.

Repetition versus recall in memorizing vocabularies. *Journal of Educational Psychology*, 5, 596-597.

A version of this paper appeared in: **G Anderman and M Rogers** (Eds.) *Words, words, words*. Clevedon: Multilingual Matters. 1996. 27-40.