

ESTIMATING STUDENTS' VOCABULARY SIZES IN FOREIGN LANGUAGE TEACHING

Sauli Takala

1 Introduction

In this paper I will discuss some issues related to the estimation of people's vocabulary sizes and present some results from one large-scale assessment study. I will first outline different approaches to vocabulary research and then focus on the methodological problems related to quantitative estimation of acquired vocabularies. I will conclude by citing empirical results obtained from one study where some new ideas in test theory were applied to vocabulary learning.

2 Different approaches to vocabulary research

2.1 Why study vocabulary?

At the outset we should address the basic question: Why should anyone be interested in vocabulary research? Why should vocabulary knowledge be an interesting and important area for research? In sum, why bother about vocabulary? There are some indications that linguistics (e.g., Bolinger, 1963; 1970; 1976; Fillmore, 1979; Halle, Bresnan & Miller, 1978; Halliday 1966; Melchuk & Zolkovsky, 1974; Raskin, 1983) is showing a growing interest in the role of the lexicon and in lexical processes as an important part of linguistic theory. Psychologists and psycholinguists have demonstrated clearly for quite some time ago that vocabulary knowledge is the best predictor of reading comprehension (e.g., Anderson & Freebody, 1981). According to some estimates (e.g., Freebody & Anderson, 1981; Frumkina, 1967; Johnson, 1972; Klychnikova, 1973), about 70 % of the words in a text should be known for a global understanding of its meaning, about 90 % for understanding all main ideas, and about 95 % for understanding also details. Thus, we can conclude that vocabulary knowledge is definitely an important prerequisite for discourse comprehension, and seeing how central learning from text is in school and out-of-school, we have ample reason to maintain that vocabulary research is an important area for research and deserves, if anything, to be strengthened and intensified.

2.2 Approaches to vocabulary research

Vocabulary research can have a number of different approaches. In this paper I will discuss three such approaches. I will call them psychological, pedagogical, and quantitative, respectively.

If vocabulary research has a psychological bias, several questions arise as possible research problems. How is vocabulary processed in comparison to e.g., perception, syntax or whole discourse? What is meant by knowing a word? How does memory work in learning vocabulary (encoding, storage and retrieval) and how can different techniques (e.g., keyword method, hook method) possibly facilitate vocabulary learning? What causes difficulty and what facilitates vocabulary learning?

If vocabulary research has a pedagogical bias, several other questions merit attention. What words should be learned (issue of selection)? What should be the nature of learning outcomes at different stages of a course: beginning, intermediate, final stage (issue of objectives/goals concerning desired vocabulary knowledge and skills)? How should words be semanticized, i.e., how should their meanings be taught? How should word meanings be consolidated? What should be the role of conscious vs. incidental vocabulary learning?

If vocabulary research has a quantitative bias, as it may have due to its nature - consisting as it does of a large amount of different words - we may ask somewhat different questions. What is the total size of vocabulary in a language? How many different words do people know? How many words do ordinary people use, and how many words do writers use? How does vocabulary grow in childhood and in the later stages of life? How common are different words?

In order to get answers to such questions, several methodological problems have to be solved. What kind of test types can be used to test different kinds of vocabulary knowledge (validity issue)? How can we get good estimates of total vocabulary sizes on the basis of a sample of words (issue of research design, and problems related to reliability/dependability and generalizability)?

3 Estimation of students' vocabulary sizes

3.1 Problem

The main purpose of the study was to estimate the size of students' active and passive vocabulary in English after they had studied English for seven years (about 600 lessons, about 450 clock hours). For a more detailed description of the research problem, see the author's doctoral dissertation (Takala 1984).

3.2 Design

In this paper we are interested in estimating the overall size of English vocabulary learned by students in the Finnish comprehensive school. Thus we are dealing with program evaluation and domain-referenced (or criterion-referenced) measurement. We wish to generalize into the whole universe of content (i.e., taught vocabulary) and into the whole population of students. This means that it is necessary to specify the content domain and draw a random sample from it. Only this kind of design makes such two-way generalization possible. In such a design, it is useful or even almost necessary to apply multi-matrix sampling, which means that different students answer partly or totally different items. Thus several test forms are randomly rotated in class.

Population. The final target population of the study was defined as "all Finnish-speaking students in the final grade of 'normal' comprehensive school classes".

Student Sampling. Preliminary studies (Takala 1984) had shown that it is important to sample a sufficient number of schools, while it would not be necessary to sample many students from each school. The sampling method was a two-stage stratified cluster sample. The primary sampling unit was the school and the secondary sampling unit was the class. Four strata were used with the size of school and the degree of urbanization of the school community as the two bases of stratification.

The designed sample of school consisted of 42 schools and the executed sample of 39 schools. Altogether, 2,415 students took part in the study.

Item Sampling. Vocabulary size estimation promised to be a good starting point for generalizability studies. It is laborious but possible, due to Finland's fairly centralized school system, to define the domain and even list and count the items in the domain.

Two textbooks, which were practically the only ones used in schools, were reviewed and words taught in them were listed separately. Textbook 1 taught about 2,500 words for the two higher sets (Sets A and B) and about

1,500 words for the lowest set (Set C). Textbook 2 taught about 2,850 words and 2,340 words, respectively. From the two separate lists, a total of about 950 words was randomly drawn and distributed among 40 different test forms. Thus each student had to respond only to 40-50 items.

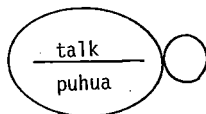
Certain design issues were tested in the study so that items were distributed to either "a robust student sample" and a "less robust student sample". They are not reported here (see Takala 1984).

3.3 Choice of test type

Several test types were considered. The constructed answer technique, in which students wrote the English equivalents of decontextualized Finnish words ("active vocabulary") and vice versa ("passive vocabulary"), was chosen on both theoretical and practical grounds. For a more detailed description of the rationale for the choice of the test type, see Takala (1984).

Sample items

Instructions: "In this test you can show how well you know the English vocabulary included in your course work. Below are presented a number of Finnish words. Your task is to write the English equivalent on the line above the Finnish word. Write the word even if you may not be quite sure about the correct spelling, since spelling mistakes are a minor consideration in scoring."



"Write the Finnish equivalents of the following English words."



4 Data collection and data analysis

Data on student vocabulary knowledge, and on the context of teaching and learning, were collected in the spring of 1979. Data file building took more than a year.

Student answers were scored 0-1 with meaning equivalence as the ultimate criterion (e.g., disregarding spelling). Interrater agreement was of the order of 95 %.

Data were analyzed using a logistic item analysis program and vocabulary size estimates were obtained through a new variance components analysis, which uses the generalized symmetrical sums (gss) method. It was shown that the results obtained with a new program are identical with those computed with Cronbach's formulas from the SPSS Reliability Program mean squares indices.

5 Some main results

The main results of the study can be briefly summarized as follows.

There was no reliable difference in the students' passive and active vocabulary knowledge, as they were measured in the study. Also, students' knowledge of simple word-formation rules and their contextual inference ability were poorly developed, in comparison to typical L1 skills. The following reasons were assumed: (1) Finnish and English are not related languages, which may not encourage such skills. (2) The emphasis at this stage is on syntactical patterns, while morphology is largely neglected. (3) The treatment of texts is "intensive", giving students little exposure to English. The estimated average size of vocabulary (see table 1, original estimates) was about 1,000 words, with great variability in performance. Fast learners knew about 1,500 words, average students about 900 and slow learners about 450 words. Due to the limited word-formation skills, the estimates ought to be adjusted by up to 45 %, by 17 %, and by 7 % for the three sets, respectively (see table 1, corrected estimates). The relationship between taught and learned vocabulary was 55 %, 32 %, and 20 % for the three sets, respectively.

Table 1. Original and Corrected Estimates for the Total Passive and Active Vocabulary Sizes, by Set

Set	Original estimates		Corrected estimates	
	Passive	Active	Active	Passive/content-aided
Set A	1,550	1,450	2,000	2,200
Set B	950	850	1,025	1,050
Set C	450	350	450	-

Variance components analysis showed that words made a greater difference in scores than students and that error of measurement can be lowered more efficiently by increasing the number of word items than by taking a larger student sample. There may also be an optimal size of input in vocabulary learning. Students who used a textbook with a lower input learned less than those whose textbook taught more words.

6 Implications and conclusions

Now that a new approach to a large-scale assessment of vocabulary size has been developed, tested empirically and found to be a promising line of study, several research questions suggest themselves. These can be divided into two major groups. One has to do with the test types and the other with student populations.

As was mentioned in the above, it was possible to test only limited aspects of vocabulary knowledge, namely relatively solid and easily accessible passive and active knowledge of words. Several experiments ought to be conducted with other test types that tap more partial knowledge of word meanings and see how vocabulary size estimates are affected.

Similarly, students' knowledge of vocabulary in the context of discourse comprehension and production ought to be estimated. Such experiments would provide data to complement the baseline data collected in the present study. It would then be possible to estimate, with a certain degree of confidence, that if students' decontextualized and firm knowledge of L2 words is X , their more partial knowledge of vocabulary is $X + Y$ words, etc. It can be conjectured that partial knowledge of a fair amount of basic words combined with some knowledge of basic morphological rules and the availability of an adequate context can lead to an adequate comprehension of test passages and to provide a good opportunity for more word learning.

The study ought to be extended to other populations. With regard to the present study, it would be important to test students' knowledge of lower stage vocabulary at the end of that school stage. This would make it possible to explain with greater confidence the finding that lower stage vocabulary was known better than upper stage vocabulary. Is this so already at that stage or is lower stage vocabulary repeated during the upper stage, and thus the difference in learning is attributable to an increase in the opportunity to learn lower stage vocabulary? This question could be studied in even greater detail by looking at each successive grade and comparing the results.

Vocabulary size assessment should also be extended to older populations. How many words do students know at the end of the senior secondary school? How many words do L2 majors at the university know?

Other studies ought to address the question of how students' ability to use word analysis skills develops over time as the study of L2 progresses. Teaching experiments ought to be carried out in which students of different age levels are taught word analysis and context utilization skills in order to see what effect such direct teaching would have on students' vocabulary efficiency.

Further, since it was found that exposure to more words had a favorable influence on vocabulary learning, it should be studied what exposure leads to optimal word learning for students of varying ability. It seems likely that the relationship is not linear but more likely an inverted U-shaped curve.

In terms of curricular implications and educational equality concerns, it would be important to study when the observed large differences in vocabulary size in L2 emerge, and whether setting/streaming (and using different textbooks with different input) tends to increase or decrease such differences. Is limited input (i.e., smaller vocabulary size taught) better for slow learners or is that a misguided notion?

In addition to such empirical research, it would be useful to devote some attention to more theoretical questions on the nature of vocabulary learning, teaching, and research. Is it, for instance, in the very nature of a domain like vocabulary that the input should be large, and that the number of words known solidly would be low or conversely the number of words almost forgotten would be high? What would that mean for teaching, testing and grading? Is, for instance, the observed large item variance component an indication of the failure of teaching, or is it a natural characteristic of L2, and for that matter L1, learning and performance?

It is obvious that a whole research program is needed to increase our knowledge about vocabulary teaching and learning both in L1 and L2. Close links between L1 and L2 vocabulary research are of great importance for optimal progress. It may be more laborious to keep track of what is being done in both L1 and L2 research, but that is necessary to avoid duplication of effort and to utilize the state of art knowledge. This is one of the main lessons that work on this investigation has provided. It is time to put that belief into practice, now that the data invite further elaboration. This will be a rewarding experience, since vocabulary research tends to have a

special fascination of its own. Its range of interest is as wide as life itself. As Vygotsky so aptly put it, a word is a microcosm of human consciousness.

References

Anderson, R. C., and P. Freebody. 1981. Vocabulary Knowledge. In J. T. Cuthrie (ed.) 1981. Comprehension and Teaching: Research Review. Newark, Del. International Reading Association. 77-117.

Bolinger, D. 1963. "The uniqueness of the word". Lingua 12: 113-136.

Bolinger, D. L. 1970. "Getting the words in". American Speech 45: 78-84.

Bolinger, D. 1976. "Meaning and memory". Forum Linguisticum 1: 1-14.

Fillmore, C. J. 1979. On fluency. In C. J. Fillmore, D. Kempler & W. S-Y. Wang (eds.). Individual differences in language ability and language behavior. New York: Academic Press. 85-101.

Freebody, P., and R. C. Anderson. 1981. Effects of differing proportions and locations of difficult vocabulary on text comprehension. (Tech. Rep. No. 202). Urbana: University of Illinois at Urbana-Champaign, Center for the Study of Reading.

Frumkina, R. M. 1967. "Slovar-minimum i ponimanie teksta" (Minimum vocabulary and text comprehension). Russkij jazyk za rubezom 2: 15-21.

Halle, M., Bresnan, J., and G. A. Miller. 1978. Linguistic Theory and Psychological Reality. Cambridge, Mass.: The MIT Press.

Halliday, M. A. K. 1966. Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, and R. H. Robbins (eds.) 1966. In memory of J. R. Firth. London: Longmans. 148-162.

Johnson, D. B. 1972. "Computer frequency control of vocabulary in language learning reading materials". Instruction and Science 1: 121-131.

Klychnikova, Z. I. 1973. "Psychologiceskie osobennosti obucenija cteniju na inostrannom jazike" (Psychological peculiarities of teaching reading in a foreign language). Moscow: Nauka.

Melcuk, I. A., and Zolkovsky, A. K. 1974. Towards a functioning 'meaning-text' model of language. In V. Ju. Rozencvejk (ed.) 1974. Essays on lexical semantics. Vol. 2. Stockholm: Språkförlaget Skriptor. 1-53.

Raskin, V. 1983. A concise history of linguistic semantics: I. West Lafayette: Purdue University.

Takala, S. 1974. Evaluation of students' knowledge of English Vocabulary in the Finnish Comprehensive school. (Tech. Rep. No. 350). University of Jyväskylä. Institute for Educational Research.