

*Jarosław Wołoszyn*

#### MEASURES OF VOCABULARY SIZE AND VOCABULARY ACQUISITION

Programmed language instruction in the field of vocabulary acquisition can make proper use of the learner's vocabulary size. The vocabulary capacity needed for the active and passive use varies with age, education, knowledge and profession. The everyday use of the language covers not more than 2000 words. Professional use expands the vocabulary in thousands of words. Hence difficulties in assessing the active vocabulary in educated people who know a number of fields of science only superficially and some fields quite thoroughly.

The passive vocabulary of a person with average education is about 40 000 words, the extreme cases reach 100 000 words, while an average novel can be read with comprehension with the use of only 5-10 thousand words.

There has been a wide variation in the estimates of the size of children's vocabulary. Some of the studies date back to the beginning of the century. A normal child of 3 years should have a vocabulary of 2000 words. In 1913 Heilig<sup>1</sup> had counted only 350 words at the age of 23 months; another researcher Bush had counted 1900 words at 30 months. French workers cited by Pieron and Fries achieved estimates divergent from the above. They hold that children of 2 years and 9 months have a vocabulary size of 650 words and 1400 at 3 years and 10 months. Yet they find true the figure of 1600 at 3 years. And again Lorrimer<sup>2</sup> having tested twenty children found 896 words at the age of three. The variation here is too wide to regard the estimates

---

<sup>1</sup> According to C. Fries, *English Word Lists*, Michigan 1950, p. 44.

<sup>2</sup> *Ibid.*

as reliable. Since the method of obtaining these results was not given the studies are not useful in assessing children's vocabulary size.

Various tests have been devised to measure the size of the individual's vocabulary. These tests, however, can measure only passive vocabulary of people with certain education (tests on children could only record their active vocabulary as used in speech). The tests do not always guarantee reliable results. Henmon's test, for example, has only statistical significance and that of Burt's cannot be applied to group testing<sup>3</sup>.

Jankowski proposes representative sampling<sup>4</sup> which employs the method of counting known words in every twentieth page in a 300 page - or every 50th in a 700 page dictionary. For convenience, the first lexical item without its obvious derivatives should be counted as a lexical unit. The result from all representative pages is multiplied by the number of pages represented.

The method described is in contrast with what has always been stressed by many linguists that words should be viewed as parts of a context. Nevertheless the test can be useful for other purposes e.g. for the evaluation of progress in vocabulary learning of the same learner<sup>5</sup>.

The present author followed the method proposed by Jankowski and examined the English vocabulary size in 10 students of English Department. Three dictionaries had been chosen: a monolingual American dictionary with 60 000 entries (The American Everyday Dictionary, New York 1955), a bilingual English-Polish Dictionary (J. Stanislawski, The Great English-Polish Dictionary, Warszawa 1964), and, finally, a specialized bilingual dictionary (Maritime Dictionary, ed. by J. Milewski, Warszawa 1963).

In the first dictionary (AED) every 33rd page was tested to obtain 16 representative pages. In these 16 pages the students

---

<sup>3</sup> Ibid., p. 46.

<sup>4</sup> B. A. J a n k o w s k i, *Nauka języka obcego: Spojrzenie psychologa*, Warszawa 1973, p. 287-288.

<sup>5</sup> The shorter time of progress the bigger number of representative pages with the use of the same dictionaries.

recognized 556 words which multiplied by 33 give 'about 18 000 words which averages the knowledge of the AED to 33%.

15 pages were sampled giving the average of 32 000 words (32%) in the English-Polish dictionary with over 100 000 entries.

Lastly, the technical dictionary gave 6900 words in 21 000 which give approximately 30%<sup>6</sup>.

Statistically speaking, the students had achieved more than 30% passive command of English vocabulary which does not hold true for the knowledge of lexical items in a given context. There, the students appeared to have understood the literary text and medical text in more than 90% and almost 85% respectively<sup>7</sup>.

It should be noted that the passive vocabulary size exceeds the lexis in live speech in all cases. Therefore it is obvious that the results obtained with the method discussed above do not represent the real knowledge of the language or, rather, of the vocabulary. Some studies reveal high estimates, others extremely low estimates of vocabulary sizes of individuals at various age levels. The divergence of results may be due to an incorrect manipulation of statistical data. The researchers seem to have failed to notice the difference between the passive and active vocabulary for one can meet a popular statement of a linguist that "an ordinary man uses about 1000 words on any occasion, out of which 500 are used habitually, the other 500 to be used in emergencies only"<sup>8</sup>.

Another researcher found that any person of average intelligence could not fail to use and understand 9000 words and their 7000 derivatives out of 16000 words found in a 45 000 word abridged dictionary; even at the lowest grade of intelligence

---

<sup>6</sup> The AED entries contained proper names, geographical and historical names (typical of monolingual dictionaries); the technical dictionary had only abbreviations among normal entries and no proper names; the bilingual dictionary had a separate section of abbreviations, proper and geographical names which however were not taken into account.

<sup>7</sup> Using the same method; guessing at the word was not excluded. Two texts were used: *End of the Chapter* by Galsworthy, Moscow 1960; and "The Journal of Experimental Therapeutics". Dec. 1973, vol. 183, no. 3.

<sup>8</sup> Ogden acc. to F r i e s, op. cit., p. 42.

nearly 6000 main words and over 3000 derivatives would be familiar.

The American Everyday Dictionary gives probable vocabulary size of an individual with certain reading habits. The figures are based on sampling word counts in typical American publications:

Medium	Vocabulary Size
Sunday N. Y. Times	40 000
New Yorker	35 000
Astounding Science-Fiction	30 000
N. Y. Herald Tribune, Washington Post,	
Atlantic Monthly, Harper's	25 000
Fortune, Time	20 000
Newsweek	12 000
N. Y. Daily News, Daily Mirror	9 000

These estimates are only hypothetical since they constitute only a potential vocabulary size. According to the same dictionary G. B. Shaw had a reading vocabulary of 80 000 words and a professional American lexicographer Henry Cecil Wyld probably knew around 100 000 words. These vocabularies were based on multiplying their writing vocabularies by five - the figure ascertained by count.

Tests which give an actual size of vocabulary are those of Kirkpatrick's and of Terman and Childs'<sup>9</sup>. The score of Kirkpatrick's test is multiplied by 280 and the other test by 180 in order to achieve the estimates of total vocabulary.

These tests are similar to the test proposed by Jankowski thus it is not unlikely to assume that the scores were similarly striking. The authors do not state if they studied the size of active vocabulary which would, even in the mother tongue, give much lower estimates. The results seem to have little influence on the assessment of actual vocabulary size.

The test which would seemingly produce most reliable data is that of Michael West's<sup>10</sup>. He based his test on the frequency word lists as a means of building up the test in such a way

<sup>9</sup> *Ibid.*, p. 47.

<sup>10</sup> *Ibid.*

that it could have a wide range at each stage testing only a small number of words with the condition that in the higher stages there would be smaller samplings and thus the larger the vocabulary tested the higher the multiplication factor. The test carefully prepared and conducted could give a very accurate determination of the individual's vocabulary. The test conducted by West in Bengal, though abandoned at some stage, revealed that a Bengali student at the matriculation age had the English vocabulary corresponding to that of an English child of 9 1/2 years, i. e. 5000 words. This means that the Bengali students were confined to books appealing to 9-10 year old children in Britain.

It would be very prudent to arrange a similar test among Polish students at matriculation age and at graduation age to achieve and analyze the data of the size of their vocabularies in order to reinforce the current programme of teaching vocabulary. Such a test, however, should be conducted at the beginning and the end of the same course which may naturally vary with time and student groups. The test may be also introduced in the secondary school at the end of each year of the language course.

The most recent measurement of vocabulary size was carried out by Diller<sup>11</sup>. He found that a 7 year old child knows 50 000 words, secondary school pupil 100 000 words and an educated person about 225 000 words which is nearly half of Webster's Dictionary. During the first 20 years of life a native speaker acquires about 10 000 new words a year. Most probably the same can be said about Poland.

Since on one claims that the foreign language vocabulary size should be equal to the native one, textbook writers and teachers should make certain that foreign language learners and, in our case secondary school learners should acquire the basic corpus of foreign language vocabulary regardless of the method employed or the future needs of the learner.

The vocabulary acquisition is a step-by-step process and it should never be overexaggerated. Finocchiaro ascertains the

---

<sup>11</sup> Cit. by W. Marton, *Dydaktyka języka obcego w szkole średniej jako maksymalizacja uczenia się ze zrozumieniem*, Poznań 1974.

teenage learner's ability to acquire new words as 15-20 per lesson<sup>12</sup>. This is probably an overestimate considering such factors as the durability, constancy, readiness and selectivity of the learner's memory. The loss of memory which amounts to 20% of the acquired load in one month<sup>13</sup> is smaller when the learner is not overloaded with new vocabulary. His vocabulary size can be expanded if the new items are associated or linked with things already known.

Therefore it is easier to teach new vocabulary on the basis of a certain fundamental lexis to be extended and built up progressively.

This idea has been successfully followed by authors of simplified texts edited in England. One of the series of those books - Longmans' Pleasant Books in Easy English consists of four stages:

- 1) The first stage is limited to about 480 most common words,
- 2) The second stage is limited to about 750 of the commonest words in English,
- 3) The vocabulary of stage 3 is limited to about 1100 the most common words, and
- 4) The fourth stage is limited to 1500 commonest words.

This series seems to have the smallest possible load of vocabulary yet it includes 27 written pieces on all possible subjects. They are: adaptations of plays, narrative, anecdotes, essays and easy articles, fiction and real stories.

Longmans' The Bridge Series generally do not exceed The Interim Report on Vocabulary Selection. The words exceeding the Report are included in the first 3000 words of Thorndike's Teacher's Word Book. Words outside the first 3000 are usually explained in the glossary but they do not occur at higher frequency than 25 per 1000 running words. This means that the range of 2000 and then of 3000 enables the reader of the series to read books on any subject.

Unfortunately, textbook writers are not unanimous in further

---

<sup>12</sup> Pinocchiario cit. by T. P. Krzeszowski, *Teaching English to Polish Learners*, Warszawa 1970, p. 125.

<sup>13</sup> Jankowski (after Ebbinghaus), *op. cit.*, p. 142.

ranges of vocabulary selection. For the benefit of the Polish learner it would be worthwhile, it seems, to include in the texts as many international words (rejected by Ogden in his Basic English) as possible. Similarly, borrowings from English into Polish are grossly underused. An attempt should be made to study the possibilities of creating such a "familiar" vocabulary load as a starting point of a textbook.

Institute of English Philology  
Department of Applied Linguistics  
University of Łódź

*Jarosław Wołoszyn*

#### POMIARY ZASOBU SŁOWNICTWA A AKWIZYCJA

Artykuł niniejszy przedstawia niektóre wyniki przeprowadzonych na świecie badań w zakresie pomiaru zasobu słownictwa u dzieci i dorosłych. Wyniki te autor artykułu porównuje z wynikami badań przeprowadzonych wśród studentów filologii angielskiej UŁ.

Badania zasobu leksykalnego, przeprowadzone w trakcie studiów, mogą w zasadniczy sposób wpłynąć na nauczanie słownictwa na neofilologii, a ponadto mogą stanowić dodatkowy sprawdzian znajomości języka obcego wśród studentów.