EXTENT OF PERSONAL VOCABULARIES AND CULTURAL CONTROL

By Professor J. M. GILLETTE
UNIVERSITY OF NORTH DAKOTA

In view of the fact that language is so fundamental in human affairs, it provokes much less attention than might be anticipated. We are surrounded by a medium of culture of which language is the vehicle as fish are surrounded by air or land animals by the atmosphere, and yet the vast majority of human beings are quite as unaware of this contiguous stratum as animals are of that which impinges upon them. After a type of animal developed which had sense, mentality enough to turn round upon its environment, inspect it, draw inferences about it, great interest was taken in its study and important results therefrom ensued.

Just how language sprang into existence no one knows exactly, but there is an abundance of theories concerning it. There is considerable agreement that close to the beginning of human life pantomime played a large rôle and was the chief mode of passing on information. Hunters, warriors and adventurers came back full of their exploits and had to give expression to them in some way. To-day a man turns loose a barrage of words and relieves himself of the tension. Then, in default of words, he had to gesticulate, imitate his former actions, represent by attitudes and postures, describe situations by means of graphic movements. It is supposed that in the excitement of this portrayal grunts and cries which may have accompanied the original activities were expressed. Probably a part of the mimiery of the situation one was trying to depict consisted of the imitation of sounds by vocal expressions. Ultimately the various kinds of sounds came to have a somewhat definite meaning and could be used apart from pantomime as signs of meaning. And after ideas could be communicated by sounds and descriptions given, these utterances would undergo a development. They would differentiate, grow in number, become more definite and so come to be an oral tongue or language. With the evolution of human society through savagery, barbarism and civilization words multiplied, syntax and grammar emerged, writing and printing came to crystallize forms of speech and embalm them in human memory. Travel and trade greatly augmented the number of terms and when science developed a still greater extension of words and terms took place.

Some one has given rise to a theory of language which is a reverberation of that early period of human life when physical effort was such a pronounced and vital part of carrying on communication. It centers on the idea that language is primarily oral with us, that it is vocal, made up of sounds produced by the manipulation of the vocal organs. We can not produce sounds without making an effort to work our organs of speech in just the right way. As a consequence of the effort we make to talk, we have a lively memory of the attitudes, adjustments, and movements of the vocal organs we experience in talking. So it comes about that our memory of words is largely a memory of the physical adjustments and efforts accompanying talking and that when we are listening to some one talking we are trying all the time to reproduce the muscular adjustments of vocal organs necessary to produce the words we hear. And beyond
this, we are intently scanning the facial movements of the speaker and interpreting the muscular activities rather than the sounds registered in our ears.

Human beings are enveloped by a medium on which they are dependent for getting a living and adjusting themselves to others. This is culture or the cultural surplus. This culture is not the academic stuff, that infinitesimal portion of all extant knowledge which college boys and girls get during their few years sojourn in academic halls and are told represents the sine qua non of existence. It is the totality of all the ideas, inventions, plans, ways of doing things, customs, manners, codes, institutions, sciences, arts and whatever men’s minds have brought into existence during the whole course of social evolution that have continued as a part of the social environment. This cultural surplus, beginning as a tiny trickle hundreds of thousands of years ago in the nascent society of eolithic man, becoming a sluggish brook among paleolithic men, a small river with neolithic men, has incremented more and more rapidly during later culture epochs and with torrential rapidity during the last century until to-day it envelops all civilized individuals in sea-like embrace.

It ought to go without saying that what we call civilization is the culture surplus in its highest and most recent reaches and that were the culture increments of the last few centuries excised, higher civilization would be removed and such civilized life would be impossible.

Since this cultural surplus is so tremendously important, its control or use must be, of course, quite as important. In fact, it is by the employment of this culture in multitudes of ways that the work of the world gets done and the satisfactions and joys of life get accomplished. It is significant to realize that the avenue of access to this culture and the central, fundamental agency for its manipulation are unobtrusive and immaterial symbols: words, vocabularies.

As culture grows, the number of words multiply accordingly. Mr. Karl Voght estimates we have added 250,000 to our language since 1900.¹ The extent of our knowledge of words is a measure of the degree of our participation in the civilization of our age. The proof of this is of a very evident sort. If we think and know in symbols and only by the use of symbols, and if all the symbols we employ are words, either heard or seen, then we can have a knowledge of civilization and its accomplishments and undertakings only to the extent that we command a knowledge of the words and symbols by means of which the culture of our age is represented. Knowledge is the avenue to an understanding of the universe at large and of the human, social world with which we are most intimately associated. The child or the man who commands slight knowledge can have only the vaguest adumbration of the significance of either. Individual intelligences graduate upward from that point through larger and larger ranges of knowledge until the most gifted or cultured minds are reached—the minds which have a fairly adequate comprehension of our social and material worlds. But so far as we know, the only depositories our minds have for knowledge are words and symbols in the form of words.

The extent of one’s vocabulary and the accuracy of the estimate made of it are both dependent upon the validity of the method employed in making the count. In consideration of this fact, it may be well to devote some attention to the question of methods.

Probably the oldest method is that which builds an opinion concerning the number of words a given person knows on counting the different words contained in his writings. Following this device, the students of Shakespeare’s works have assigned a vocabulary of twenty-four thousand words to that

famous writer and those of Milton’s works have found that the blind bard made use of seventeen thousand words. The numerous writers of the English Bible all together had to content themselves with handling only seven thousand two hundred words. But they lived in very primitive times for the most part and we could not use our standards of judgment on them. However, in the case of all publicists, it is to be remembered that the number of words they know is undoubtedly very much greater than they use in their writings; and that even the number of words they are capable of using is much more extensive than that appearing in their published works. It is interesting to note that some one has made a reckoning—how good a one it is impossible to say—that the late President Wilson knew over seventy-two thousand words. This estimate was made on the basis of the addresses and papers he produced during the war together with his earlier published works. And as we shall see later, the estimate of this large number is probably not only not exaggerated but is entirely too conservative. It is likely that Woodrow Wilson knew twice that number.

There have also been attempts, largely in the nature of pure guesses, to judge the extent of vocabularies of the rank and file of people. There are statements to the effect that Italian grand opera singers employ only about six hundred words, that the words used by many peasants number still fewer, being only three or four hundred, and that the vocabulary of persons above the average contains not more than three thousand or four thousand words. We must regard these as only rough and entirely inadequate guesses, for it is certain that primitive peoples who live in a much lower stage of culture than grand opera singers and peasants have considerably larger vocabularies.

A knowledge of the extent of vocabularies of primitive people throws a good deal of light on the subject under discussion. It furnishes a background on which to form a judgment concerning the capacities of peoples a great deal more advanced. If the daily spoken language of a savage folk is a matter of ten thousand words, what should be that of an enlightened people, and more especially that of those persons placed at the summit of the highest cultural attainment? We certainly should expect a developmental expansion of considerable moment.

There have been many dictionaries of primitive peoples compiled by missionaries and philologists. It is, therefore, quite within our power to make an estimate of the number of words in their lexicons. In these primitive vocabularies there are few or no dead and obsolete words, for, in nearly all cases, there was no written language and all the words the investigators collected were those which were in current use by the living men and women making up the groups. The chances, therefore, are that most of the adults of the group were in command of the bulk of the words constituting the vocabulary. According to Kroeber, the following are the number of words contained in the vocabularies of some primitive peoples: the Aztec Nahua, 27,000; the Central American Maya, 20,000; the Plains Dakota, 19,000; the African Zulu, 17,000; the Navaho of our Southwest, 11,000; the Klamath of our Northwest, 7,000. The range of culture of these groups is from upper barbarism to upper savagery. If the vocabulary of savages ranges from seven to nineteen thousand words and that of barbarians ranges still higher, what should be the extent of the vocabulary of the average civilized or enlightened Frenchman, German or Englishman? Undoubtedly it must be a multiple of the other.

There have been formulated in recent years certain standardized modes of estimating the number of words in any given person’s vocabulary. One of these is the Terman test, which was devised in the
following fashion. From a dictionary containing eighteen thousand words, which are supposed to be the most common ones, was made up a list of one hundred words selected by taking the last word of every sixth column of the dictionary. This method of selection would prevent the operation of bias or pre-judgment. Now the one who wants to test his vocabulary selects from the hundred words those which he can fairly well define. The resulting number is multiplied by 180. Did one know all the hundred words, his resulting vocabulary would be eighteen thousand. Did he only know one half of the hundred, his vocabulary would be nine thousand words. Upon applying the standard test to himself, the present writer found he has a vocabulary of 16,833 words. This is between one seventh and one eighth the number obtained by the method to be described later in this article.

According to the results obtained from the application of this test to a large number of persons, some average or general conclusions were obtained. A person eight years of age knows 3,600 words, one of ten knows 5,400, one of twelve knows 7,200, one of fourteen knows 9,000, the average adult knows 11,700, and the superior person knows 13,500. Using the test on college students, it was found that male students know the following number of words: Freshmen, 9,240; sophomores, 10,860; juniors, 13,040; seniors, 12,700. The vocabularies of female college students according to the same order of classes are: 8,860; 9,325; 10,130, and 10,700.

The general method of random or arbitrary sampling of the phenomena to be measured employed here in the selection of the basic list of one hundred words is scientifically correct. It is the method commonly used by scientific statisticians in their study of phenomena of every kind. A sample that is large enough and arbitrarily chosen is found to be an accurate representation of the whole realm of phenomena of which it is a part. Nevertheless, we are inclined to regard the Terman test as coming short of the results which such a test should attain. It probably underestimates the extent of the vocabularies of the persons to whom it is applied. It seems likely that the dictionary of eighteen thousand words is much too restricted and that consequently the multiple applied to the known words is not nearly large enough. Were the same method of sampling applied to an unabridged dictionary and a multiple applied which was adjusted to the size of the lexicon, the results would unquestionably be much more competent.

It should be clear by this time that one's estimated vocabulary is largely the outcome of the kind of method used in making the estimate and of the limitations imposed in the use of the method. Thus there are several bases for making an estimate which, if employed, would yield widely different results or sizes of vocabulary. There are: the ability to define words, the ability to use words irrespective of the ability to define them, knowing words for reading but not for speaking or writing purposes. The resulting vocabularies will vary in size in the order of the bases named.

The extent of the dictionary employed in formulating the standard or in making the test should be sufficiently extensive to provide for the inclusion of two very essential items: first, the technical vocabularies of professional persons, like doctors, lawyers, chemists, engineers; second, of business men, such as merchants, manufacturers and managers. A professional man or a technician has a knowledge of thousands of technical terms used in his calling in addition to the usual vocabulary of educated persons. As a sample of this I find that the manager of a Woolworth store in a city of twenty thousand inhabitants must know and have at his tongue's end the name of all the five thousand articles
PERSONAL VOCABULARIES

It has been apparent to the writer of this article that there is a great surprise in store for any one who will make an elaborate and painstaking investigation of the extent of his own vocabulary, using as the basis of such test one of the great unabridged dictionaries of the English language. If he has hitherto looked up to Shakespeare as the consummate master of the language in all the phases and regarded himself as a mere pigmy in his knowledge of words when compared with the score and a quarter thousands Shakespeare is said to have made use of, his pride in himself will swell enormously after he has completed the self-investigation. Not that it will be demonstrated that he can utilize a greater number of words in writing or speaking than that great master of drama, but only that he really has a knowledge of a very much greater number of terms than the former is known to have employed in his written productions. The short description of some results the present writer has obtained in this field will give a semblance of veracity to these statements.

The statements I have read concerning the extent of vocabularies known or used by humans served to make me quite curious concerning the exactitude of the results and methods used to obtain them. Many of the estimates seemed to be no better than mere guesses and the results secured by some very good tests appeared to fall short of those which might have been attained had other methods been employed. Consequently I undertook to make a test of the extent of my own vocabulary, that is, of the number of words I can fairly claim to have some considerable knowledge of.

The dictionary employed in making the test is the Standard unabridged of 1922. This lexicon contains 2,737 pages, each page containing three columns of words. The words of these columns appear in two different situations: those

contained in the store. Many of these articles are, of course, common articles which every one would know. But a well-equipped practitioner possesses a store of many thousands of technical scientific terms which are unknown to the laity. Second, it should include, also, the words built up on the foundation of certain basic words by a process of compounding through the use of prefixes and suffixes. Such words usually do not appear at the left margin in the columns of unabridged dictionaries, or at the foot of the column; consequently they are liable to escape inclusion in samples and so are not allowed for in multiples. Thus from the word abolish are built nine other words by the addition of suffixes. From the word accent are derived ten additional words. Accept gives rise to six and access to seventeen. Then by the use of prefixes twenty other words are added to the fifty-four so far derived from the four original words, altogether a total of seventy-four words. One gets the force of this method of multiplying words relatively to the size of one’s vocabulary by supposing that the person in question knows five thousand words of the basic sort from which it is possible to build up other words. The average number of words built from the four words cited above was eighteen. Multiplying 5,000 by 18 yields a total of 90,000, the enlargement to the number of words the individual pretty certainly knows because he knows the foundational ones. This is a purely suppositional case, and instead of five thousand we might need to substitute four thousand, three thousand, or some other number. It certainly would not do to multiply one’s entire vocabulary by such a multiple as 18 because much of the vocabulary is made of proper names, historical, geographical, and personal, which are not generally subject to taking on prefixes and suffixes.
heavy type words set even with one another at the left margin; and non-marginal words, such as synthetic and compound words, and many proper names, phrases, etc. One kind of synthetic words is built out of such more elemental words as accent, abolish and the like by annexing prefixes or suffixes. Another kind is made by compounding some marginal word with other words.

The first estimate concerned the marginal words. It was conceived that if enough columns to fill fifty-two pages, or an average of two pages for each of the twenty-six letters of the alphabet, were examined, a sample sufficiently large would be obtained to be representative of all the columns of the lexicon. However, it seemed more scientific to select columns under each letter in the proportion each such letter's number of columns bears to the total number of columns. This gave to such letters as x and z only a part of a column and to such as c, sixteen, and to s, seventeen columns; but no letter was assigned less than one column. The particular columns to be employed under each letter were chosen by the hit-and-miss plan in order to secure purely arbitrary and random sampling, thus obviating personal bias. Then the words about which I knew much or considerable were marked, the total number of words in the column were counted, and the results were tabulated column by column until the total 159 columns had been covered. The total number of words counted was 4,019, the number known was 1,150, 28.66 per cent. It was estimated from the average number of words per column counted that the dictionary contains about 209,000 marginal words. It was assumed that if one knows 28.66 per cent. of the large sample of over four thousand words, he will know the same proportion of the 209,000 words. In this case the number of marginal words known was 59,800.

Since a full explanation has been made of the method used to secure an estimate of the proportion of known marginal words contained in the dictionary columns, it will be sufficient to state that the same general method was followed in securing an estimate of the known words of all the non-marginal terms. The editors of the Standard indicate that the particular edition used to make the test contains 450,000 words. We noted that the total number of marginal words amounted to about 209,000. This would leave about 241,000 of the non-marginal kind. The percentage of words of this kind which were known was found to be 28.2, or almost the same as the proportion of the marginal sort. Applying this percentage to 241,000 yields the total number of this kind of words that are known, which was 68,000. And now totaling the two estimates, 59,800 and 68,000, we have a grand total of 127,800 words known.

In consideration of the other estimates that have been made in the past, this number seems enormous. Some may be inclined to discredit the estimate altogether. Before doing this, however, let the reader observe two cautions: First, take the time to give the method used here a fair test by applying it to himself. It is likely that then he will be satisfied that this estimate may not be far wrong. Second, then consider that a large number, probably far the larger number, of known words are forms or derivatives of more elemental words. Think of circum in combination with a lot of other elementary words whose root meanings are known, resulting in circumvent, circumscribe, circumambulate, circumlocution, etc., etc. Then recall such a word as accent and what derives from it: accent, accentuation, accenual, accentualist, accentually, accentuality, accentuate, accentuable, accentuation, accentus. It is pretty certain that if one knows accent about all the other forms
of that word are known. And what is true of these forms of words is also true of compound words and of phrases.

So far, I have been able to induce only two persons to apply this test to themselves. It requires much time and persistence, so that few have the requisite scientific curiosity to carry it out. But fortunately, two of my very mature students, graduate men in their thirties, have made estimates. One used a Practical Standard Dictionary containing 140,000 words. He selected two pages under each letter of the alphabet by the hit-and-miss plan for counting purposes. He found he knew 2,055 of the 4,368 words contained in the fifty-two pages employed, or 47 per cent. This yielded a total of 65,800 words out of the 140,000. Had he employed an unabridged dictionary, the number he knew might have been considerably swelled.

The other student made use of an unabridged Standard dictionary of 2,757 pages. He considered both marginal and non-marginal contents. His estimate of known words in the entire dictionary was 52,489.

It is noted that the vocabularies of the students are only about half as large as the vocabulary of the first estimate. Perhaps this difference may not appear as evidence of the incompetency of the method when it is known that the one who made the first estimate has been a student during a rather mature life-time and has a command of languages the others do not possess. But one must expect variations even in the estimates made by those of about the same degree of culture because of the fact that there is no absolute criterion of the degree of knowledge which is to be applied in selecting the words which are to be considered as known. Two persons might have the same knowledge concerning a given word and one would class it among his known words and the other among his unknown words. This is a difficulty which it would be hard to overcome. Nevertheless, it is probably not a very important difficulty.