STUDIES IN LANGUAGE BEHAVIOR

I. A Program of Research

WENDELL JOHNSON
University of Iowa

The studies by Fairbanks (5), Mann (11) and Chotlos (4), which are presented in this issue of The Psychological Monographs, constitute the beginnings of a program of research in language behavior. They have been completed in the order named. The present paper is intended as an introduction to them, and to the general program which they represent.

The importance of language and of symbolization generally, as a distinctly human form of behavior and as a basic factor in personal and social problems, is generally recognized (9, 10, 12, 20). The effective scientific investigation of such behavior, however, depends upon the development of highly reliable and differentiating measures, by means of which specified aspects of language behavior might be systematically observed in relation to one another and to other variables. With such measures, significant testable hypotheses can be formulated and checked, and a body of dependable information can be accumulated.

SPECIFIC OBJECTIVES

The proposed program of research is designed to:

1. Develop reliable and differentiating measures of specified aspects of language behavior.
2. Determine the degree to which the resulting measures are intercorrelated.
3. Determine the degree of correlation between these measures and such other pertinent variables as those involved in environmental influences, physiological conditions, intelligence, and personality adjustment.
4. Apply the measures to a comprehensive investigation of language development.
5. Determine the degree to which language behavior, as measured, is modifiable under specified conditions.
6. Determine the degree to which modification in language behavior is associated with modifications in other aspects of behavior or adjustment.
7. Indicate the normal characteristics of language development and language behavior, and the varieties of disorder or abnormality in such behavior, in terms of the measures used.

Types of Language Measures to Be Investigated

No attempt will be made here to present a review of the theoretical and experimental literature dealing with the problems with which this program is concerned. It is sufficient to say that previous work in the field has suggested many of the procedures to be employed, and that others have been suggested by preliminary research carried out by the writer, or under his direction. A comprehensive review of language behavior studies has been published by Sanford (12). The following types of language measures are to be investigated:

Type-token ratio (TTR). This is a measure of vocabulary "flexibility" or variability, designed to indicate certain aspects of language adequacy. It expresses the ratio of different words (types) to total words (tokens) in a given language sample. If in speaking 100 words (tokens) an individual uses 64 different words (types), his TTR would be .64. In order to develop the most highly reliable and differentiating form of the TTR, it is to be computed for
given language samples in the following various ways:

a. For all words spoken or written by a given individual, or in a given language sample, and separately for words representing the various grammatical categories; for words in different frequency categories—for example, the 500 most frequently used words, the next 500 most frequently used words, etc., as determined by the published word-counts of Thorndike (17), Horn (6), and others, or by the word-counts to be derived from the present investigations; etc.

b. With varying statistical or mathematical procedures, thus:

The over-all TTR, as computed for an entire language sample. TTR’s for samples of different magnitudes are not directly comparable because of the tendency for the TTR to vary inversely with size of sample. A knowledge of the precise character of this inverse relationship might make it possible to compare directly TTR’s for samples differing in length, by means of a correction table. The feasibility of constructing such a table is to be investigated. The study by Chotlos (4) throws considerable light on this problem.

The mean segmental TTR. TTR’s for samples of different magnitudes can be made comparable by dividing each sample into like-sized segments of, say, 100 words each, computing the TTR for each segment and then averaging the segmental TTR’s for each sample. It can be safely assumed that such segmental TTR’s are directly comparable, so long as they represent segments of equal size, and that means of such segmental TTR’s are also directly comparable. Results obtained by using segments of different magnitudes—as 100-word segments, 500-word segments, etc.—are to be compared, in order to ascertain the size of segments that will allow for the most reliable and differentiating mean segmental TTR. The above mentioned study by Chotlos (4) is concerned with this problem also.

The cumulative TTR curve. A curve of the cumulative TTR for a given language sample can be plotted by computing successive TTR’s as increments are added to the sample. For instance, the cumulative TTR for a 1000-word sample would be plotted as follows: TTR values are to be represented along the ordinate and number of words along the abscissa. The abscissa values may be in units of one word, or ten words, or 100 words, etc., as desired. If the unit is one word, 1000 TTR’s would be computed in plotting the cumulative curve for the 1000-word sample; if the unit is ten words, 100 TTR’s would be computed; if the unit is 100 words, ten TTR’s would be computed, etc. Thus, if the unit is ten words, the first value will represent the TTR for the first ten words of the sample, the second will represent the TTR for the first 20 words, the third will represent the TTR for the first 30 words, etc. The problem of fitting an equation to the resulting curve is dealt with in some detail by Chotlos (4). Basically, the problem concerns the relation between D (number of different words, or types) and N (number of words, tokens) in the given sample. This problem has been given considerable attention by Zipf (20), Carroll (3), and Skinner (16). The relevant data presented by Chotlos (4) indicate the degree to which the relation of D to N promises a means of predicting vocabulary, in the sense that the value of D for a given N provides a basis for predicting D for a specified N of larger magnitude.

The decremental TTR curve. Sup-
pose a 1000-word sample to be divided into ten 100-word segments. The TTR is computed for the first segment. Then, the number of different words in the second segment that did not occur in the first segment—i.e., the number of new types introduced in the second segment—is found. The TTR for the second segment is then computed by dividing this number—not the number of types, but the number of new types—by 100, which is the number of tokens in the second segment. In the same way, the TTR’s for the third, fourth, and each of the other segments may be computed, by dividing the number of tokens, 100 in each case, into the number of new types introduced into the sample for the first time in the segments under consideration. The resulting curve of these successive segmental TTR’s may be expected to show a relatively steeper slope than the cumulative TTR curve, and the measure representing the slope of this curve may be found to be of special interest. It represents, of course, the rate of decrement in the use of new types, the rate at which the individual “uses up” his vocabulary in producing a language sample. Decremental TTR’s should represent in a peculiarly direct quantitative manner one aspect of language development, when applied to language samples secured successively from the same children. The decre­mental TTR curve is, of course, the first derivative of the cumulative TTR curve, and thus it is not actually necessary to fit a curve to the decremental TTR data if the cumulative TTR curve has been computed.

Type-frequencies. A simple objective language measure is that which expresses the frequency of occurrence of each different word, or type. Such frequencies, as reported for large samples of written language by Thorndike (17), Horn (6), and others, have been used chiefly in the preparation of school readers, spelling books, etc. Certain other uses of such data are obvious. When type-frequencies are based on the kinds of language samples to be used in the present program they may be regarded as representing language behavior norms. In previous studies of word-frequencies it would seem that the primary objective has been simply to determine the relative frequency of occurrence of each word, and with some exceptions special interest has attached to those words which have been found to occur with especially high frequencies. The main objective of the present program in this connection is somewhat different. Chief interest lies in ascertaining individual and group differences in the relative frequency with which particular kinds of words are used. One may determine (a) type-frequency changes that characterize language development; (b) type-frequency characteristics of the language of special groups, especially those that may be found to differentiate one group from another, as schizophrenics from normal subjects, scientists from novelists, etc.; (c) the particular type-frequencies that correlate significantly with such other variables as intelligence, emotional stability, educational level, etc. Attention may be given to the following types of words (and to any others that may be found to be useful):

a. Self-reference words.
b. Quantifying terms (precise numerical words).
c. Pseudo-quantifying terms (words loosely indicative of amount, size, etc., such as much, many, lots; or very, highly, etc., used as qualifiers of other pseudo-quantifying terms, as in such expressions as “very much”).
d. “Allness” terms (superlative or extreme
words, such as never, always, all, nobody, everyone, etc.).

e. Words expressive of negative evaluation, such as no, don't, etc., and horrid, unsatisfactory, dislike, etc.

f. Words expressive of positive evaluation.

g. Qualification terms (words that serve to qualify or limit statements, such as except, but, however, if, etc.).

h. Terms indicative of consciousness of abstracting (such words as apparently, seems, appears, as if, to me, etc.; as indicated by the last two examples, for purposes of this type of analysis it will be necessary to treat certain phrases as single words. What we call the dogmatic or “closed mind” attitude might be expected to be characterized by language in which these terms are relatively lacking.)

Ratios of any one of the above types of words to any one of the other types might be computed for given language samples, and their significance evaluated. The ratio of the terms indicative of consciousness of abstracting to “allness” words, for example, might be expected to differentiate individuals and groups in ways that should be of theoretical and practical importance in the study of personality.1

The relative frequency of use of the various grammatical types of words—nouns, adjectives, verbs, adverbs, etc.—might also be determined, as well as ratios of nouns to adjectives, adjectives to verbs, verbs to adverbs, nouns to verbs, adjectives to adverbs, etc., and the ratio of these four to all other words. With language development, the relative frequency of nouns particularly and also of verbs may be expected to decrease with reference to the relative frequency of adjectives and adverbs. The degree to which these and other possible relationships can be utilized as measures of language development and of individual and group differences should be ascertained. Busemann (2) and Boder (1) have employed the adjective-verb quotient to indicate certain kinds of personality differences, and to differentiate samples of written language. Sanford (13) has reported a personality study involving this and other related measures. The present series of studies involves analyses in this general connection. Mann (11) applies the adjective-verb quotient and also adjective-noun and adverb-verb quotients in her comparative study of the written language of schizophrenic patients and university freshmen. Fairbanks (5) investigates the relative frequencies of occurrence of various parts of speech in comparing the spoken language of schizophrenic patients and university freshmen. Chotlos (4) presents similar data in terms of types and tokens, respectively, and he also presents TTR values for nouns, verbs, adjectives and adverbs, respectively, for written language samples obtained from Iowa school children.

Proportionate vocabulary. How many different words or types make up 25, or 50, or 75 per cent of a given language sample? In the study by Fairbanks (5), 30,000-word samples of spoken language were obtained from schizophrenic patients and “superior” university freshmen, respectively. For the freshmen just 46 different words or types comprised 50 per cent of the 30,000-word sample, and for the schizophrenic patients this figure was 33 types. This is the more striking, perhaps, when expressed by saying that for the schizophrenic patients approximately one-tenth of one per cent of the words made up 50 per cent of the total

1 Reference made here to “allness” terms and to the notion of consciousness of abstracting implies the writer’s debt to Alfred Korzybski. See especially Korzybski, A. Science and Sanity, An Introduction to Non-Aristotelian Systems and General Semantics, Lancaster, Pa.: Science Press, second edition, 1941.
sample. In fact, one word, the one most frequently used by the schizophrenics, which was the word / made up slightly over 8.3 per cent of their entire 30,000 words.

A sample of, say, 1000 words might be analyzed in such a way as to yield a curve as follows: Along the ordinate percentages would be represented; these percentages would correspond to numbers of tokens. For example, suppose that 100 tokens make up 10 per cent of the 1000-word sample; it is this 10 per cent and other percentage values so computed that would be represented along the ordinate. Other percentages would lie along the abscissa; these percentages would correspond to numbers of types. Thus, suppose that 10 types comprise 1 per cent of the total of 1000 tokens; this 1 per cent and other percentage values so computed would be represented along the abscissa. The curve showing the relation between these two sets of percentages would be made up of points expressing such values as the one cited above: for the schizophrenic patients 0.1 per cent of the words (this percentage representing types) made up 50.0 per cent of the sample (this percentage representing tokens). The relation symbolized by this curve can be expressed mathematically, of course, and it is proposed to examine its usefulness as a basis for comparing different language samples or any given sample with a norm or standard sample. The relationship discussed here can be expressed, of course, in terms of rank and frequency. That is, a curve that is fitted to word-frequencies as a function of rank, the most frequent word having the lowest rank number, 1, represents in an alternative way the same phenomenon that is discussed here in terms of proportionate vocabulary. (See Zipf [20].)

Standard frequency vocabulary. The word counts that have been published by previous workers, and the one to be done in the present program, can be used separately or pooled in arriving at a standard frequency-of-use rank number for each different word included in them. Such rank numbers would represent the relative frequency with which each word had been used in the total language sample—presumably drawn from a more or less representative population of individuals—not in terms of the actual number of times each word was used, but in terms of its rank. Thus, the most frequently used word would have a rank number of 1, the next most frequently used word would have a rank number 2, etc.

With the resulting table of rank numbers, it would be possible to score any given language sample by noting the rank number of each word (token) contained in it, and computing the mean (or median) of these rank numbers. The lower the mean of the sample the more heavily loaded it is with words that are used relatively frequently by people generally. We may say, then, that this mean rank number of a language sample represents the "standard frequency vocabulary" employed in it. It is to be reasonably expected that language development would be characterized by increase in this measure, and that the measure would serve to differentiate individuals and groups.

A less refined, and perhaps nearly as adequate, form of this measure could be worked out in terms of standard frequency rank numbers on a categorical basis. That is, the first 100 most frequently used words, for example, could all be given the same rank number, the number 1, the second 100 words could all be assigned the rank number 2, etc.
Statistical analysis may indicate advantages in classes or categories of unequal magnitudes, putting the more frequently used words in smaller groups, for instance, and the less frequently used words in larger groups, or vice versa, perhaps varying the number of words in a group in some relation to the frequency with which they are used. Comparison of results obtained from use of various forms of the measure will determine the relative merits of each.

**Verbal output.** A very simple language behavior measure is that which expresses the verbal output of an individual. Individual differences and intra-individual variations with respect to verbal output are, of course, obvious. Their significance in relation to the various aspects of personal and social adjustment have not been thoroughly or systematically investigated. It is planned to include an attempt in this direction in the present research program.

Verbal output is not meant to be synonymous with speaking or reading rate, as that term is used to refer to verbal output under relatively optimal conditions. An individual's verbal output under various conditions may, and usually does, fall considerably under what it is when he speaks at or near his optimal steady rate. Verbal output may be expressed, of course, in terms of rate.

The measure may express number of words spoken or written per unit of time, or in response to a specified stimulus under standard conditions. It may also express the proportion of a time unit during which an individual produces spoken or written language. For example, two individuals could be compared by placing them together for one hour and recording (a) the speaking time of each, (b) the total number of words spoken by each, and (c) the verbal output of each in terms of words spoken per minute. It is to be noted that these measures are different from a measure of the rate of verbal output while speaking. It would be of interest, of course, to correlate such a measure of rate with the other verbal output measures.

**Word length.** Since the studies of Zipf (20) have shown word length to be highly correlated negatively with frequency of use—the shorter the word the more frequently it occurs—it is not planned at this time that measures of word length will be included to any important degree in the present program. It is mentioned here, however, because the data to be utilized will be so tabulated that word length could be studied if findings indicate that this would be advisable. It is a rigorously objective and highly reliable measure (15).

**Sentence length.** Sentence length is a measure that presents serious operational difficulties in the study of spoken language, although it may be generally satisfactory in the analysis of written language. It is planned to include it in the analysis of at least a selected set of the written language samples.

**SPECIAL TYPES OF LANGUAGE BEHAVIOR TESTS**

The Extensional Agreement Index (EAI) expresses the degree of agreement among n persons in defining a given term extensionally—i.e., by pointing to or exhibiting somehow the actual objects, phenomena, etc. to which the term refers.² Thus, the kind of behavior which the EAI is designed to measure is not observation so much as word-fact relat-

---

²This measure was introduced and briefly discussed in Johnson, W., Language and Speech Hygiene, referred to in footnote No. 1.
The EAI may range in numerical value from 0.0 to 1.0, 0.0 representing no agreement and 1.0 representing maximum possible agreement among \( n \) persons in relating or applying a given word as a label to actualities. Its theoretical and practical significance lies in the fact that it makes possible not only an index of a person's conformity or idiosyncrasy in his extensional use of words, but also a measure of the degree to which any given term may be regarded as testable, or extensional, or operational— or vague. If in the statement “Stutterers are psychoneurotic” the term “psychoneurotic” has an EAI of, say, .18, the statement is not to be regarded as highly testable or factually meaningful, since \( n \) persons would disagree considerably as to just what is to be observed in order that the validity of the statement might be tested. The EAI offers, therefore, a means of quantifying to some degree such notions as are represented by the terms “verifiable,” “operational,” etc.

The EAI may be computed in several different ways. Tuthill (18) in a study made as part of the present program demonstrated a variety of ways of computing such a measure of extensional agreement. The basic formula is

\[
EAI = \frac{x}{y}
\]

in which \( x \) represents the number of obtained agreements and \( y \) the maximum possible number of agreements. The EAI, then, represents the per cent of the maximum possible number of agreements that are obtained in a given case.

For example, imagine four different pictures and ten different persons who are each asked to apply the label “most artistic” to one of them. Suppose the label is applied to picture A by 3 persons, to picture B by none, to picture C by 5, and to picture D by 2. If there had been perfect agreement, all 10 persons would have applied the label to the same picture. Thus, the number of agreements among the 10 persons that would have occurred under these conditions is to be regarded as the maximum possible number of agreements. This number may be determined by the formula \((n - 1) \cdot 5n\) and since \( n = 10 \), the maximum possible number of agreements is \( 9 \times 5 = 45 \). The number of agreements actually obtained is to be computed as follows: The three persons who applied the label to picture A agreed 3 times, since when \( n = 3 \), \((n - 1) \cdot 5n = 3 \). There were no agreements with regard to picture B, in terms of the technique for computing the EAI that is here being used. Using the formula \((n - 1) \cdot 5n\), there were 10 agreements in the labeling of picture C, and one in the labeling of picture D. In all, then, 14 agreements were obtained. Therefore, \( EAI = \frac{14}{45} = .31 \), which may be interpreted as indicating that the number of agreements obtained was 31 per cent of the maximum possible number.

This is an example of an extremely simple case, used to illustrate the application of the basic formula. Another example will serve to indicate an important modification of the basic formula. On July 9, 1939, the American Institute of Public Opinion released to newspapers the results of a survey in which each of several thousand persons had been asked to apply one of the labels, “Conservative,” “Liberal,” and “Radical,” to each of ten prominent Americans. The results were presented
in percentages as follows:

<table>
<thead>
<tr>
<th></th>
<th>Conservative</th>
<th>Liberal</th>
<th>Radical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopkins</td>
<td>4</td>
<td>55</td>
<td>41</td>
</tr>
<tr>
<td>Roosevelt</td>
<td>1</td>
<td>62</td>
<td>37</td>
</tr>
<tr>
<td>La Guardia</td>
<td>8</td>
<td>64</td>
<td>28</td>
</tr>
<tr>
<td>Farley</td>
<td>13</td>
<td>68</td>
<td>24</td>
</tr>
<tr>
<td>Dewey</td>
<td>45</td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td>Hull</td>
<td>51</td>
<td>49</td>
<td>3</td>
</tr>
<tr>
<td>Garner</td>
<td>64</td>
<td>32</td>
<td>4</td>
</tr>
<tr>
<td>Vandenberg</td>
<td>67</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td>Taft</td>
<td>86</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Hoover</td>
<td>92</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

These figures represent only the labeling reactions of persons "who knew or had some idea of the terms when later in the survey they were asked point-blank what the words . . . meant." From these data it is possible to compute an EAI for each of the three terms involved.

The procedure to be used will differ in three important respects from that used in the above example of the four pictures. In the first place, in the first example there was only one label to be applied by each of ten persons to only one of four possible referents. In the present case, there were three labels, any one of which was to be applied to each of ten referents. In the second place, there were ten labelers in the first example; in this one there were many thousands, and the numbers have been converted into percentages. These percentages will be used instead of the raw numbers in computing the EAI's. In $(n - 1) .5n$, $n$ will represent 100 in computing the maximum possible number of agreements. Lastly, instead of assuming, as was done in the first example, that agreements occur only when labels are applied, and not when they are not applied, we shall assume that both the application of a label and the refusal to apply it may involve agreement. When this assumption is made, the net number of agreements involved in the application and non-application of a given label to a given referent can be computed as follows. Let $x =$ the number who apply the label, and $n - x =$ the number who do not apply it. Then, the number of agreements among those who do apply the label is found by the formula, $(x - 1) .5x$. Similarly, the number of agreements among those who do not apply the label equals $(n - x - 1) .5 (n - x)$. The net number of agreements is found simply by subtracting the smaller of these values from the larger. And the EAI is found by dividing this net number of agreements by the maximum possible number of agreements. Thus,

$$EAI = \frac{2x - n}{(n - 1) n} = \frac{2x - n}{n}$$

In this way, the EAI of each given term is computed for each referent, and the EAI's of the term for the various referents (in the present case, 10) are averaged. For the term "Liberal" the following results were obtained:

<table>
<thead>
<tr>
<th></th>
<th>% Labeling</th>
<th>2x - n</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hopkins</td>
<td>55</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Roosevelt</td>
<td>62</td>
<td>.24</td>
<td></td>
</tr>
<tr>
<td>La Guardia</td>
<td>64</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>Farley</td>
<td>63</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>Dewey</td>
<td>47</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Hull</td>
<td>46</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>Garner</td>
<td>32</td>
<td>.36</td>
<td></td>
</tr>
<tr>
<td>Vandenberg</td>
<td>29</td>
<td>.42</td>
<td></td>
</tr>
<tr>
<td>Taft</td>
<td>13</td>
<td>.74</td>
<td></td>
</tr>
<tr>
<td>Hoover</td>
<td>9</td>
<td>.90</td>
<td></td>
</tr>
<tr>
<td>Ave. EAI</td>
<td></td>
<td>.34</td>
<td></td>
</tr>
</tbody>
</table>

The obtained number of agreements was, on the average, only 34 per cent of the number representing complete agreement as to the extensional meaning of the word "Liberal," as applied or not, to the ten men listed, by the presumably
random sample of persons surveyed by the Gallup organization in the summer of 1939. The variability is of interest. As applied to Hopkins, Dewey and Hull, the term “Liberal” proved to be almost entirely meaningless; there was virtually no agreement as to whether these men were or were not suitable referents of the term. There was relatively high agreement, on the other hand, that Taft and Hoover were not to be labeled “Liberal.” The mean EAI was .58 for “Conservative” and .69 for “Radical.”

Dr. Gallup, under whose name the survey report appeared in the press, did not, of course, report his findings in terms of these EAI’s. Moreover, he seems to have missed a basic point, in stating that the survey results indicated “the way American voters—rightly or wrongly—are classifying the figures in United States political life.” (The italics are the present writer’s.) The words “rightly or wrongly” seem to imply the assumption that there is a “right” way and a “wrong” way to apply such a label as “Liberal,” that such a term has somehow an intrinsic “meaning,” presumably known by some means to someone somewhere, quite aside from and more valid than the extensional meanings ascribed to it by those persons who actively relate it to various referents. There would appear to be, from an extensional point of view at least, no “right” or “wrong” about it, except in the sense that in matters of this kind one might (or might not) prefer to assume that the majority is “right.” Be that as it may, however, Dr. Gallup carried out, in this particular survey, what amounted to a very ambitious effort to determine by vote the extensional meanings of a group of words. And by using his results to compute the EAI’s of these words, it becomes possible to measure fairly precisely the vagueness or factual meaningfulness of some of our important political terms.

The resulting EAI’s afford a degree of insight into the processes of political controversy, and point to one of the fundamental problems in connection with social organization. The EAI of .34 for “Liberal” strongly suggests that such a statement as, “America should (or should not) have a liberal in the White House,” is to be regarded as essentially “lyrical.” Like our remarks about the weather, which are not to be mistaken for meteorological reports, the remark that “So-and-so is a liberal” is not to be regarded as a statement chiefly descriptive of So-and-so. For the most part, it merely serves to announce one of the ways in which the speaker proposes to apply the word “Liberal,” and thus it is mainly indicative of an aspect of the speaker’s language behavior. To know the EAI of a word, as computed from data as adequate as those provided by Dr. Gallup, is to know something quite precise and significant about the language behavior of a speaker or writer who uses it, particularly if he gives no indication of awareness of the word’s descriptive limitations, as these are implied by its EAI. The descriptive limitations of a word with an EAI of .34 are probably so great as to render it practically meaningless referentially in many contexts. It is to be regarded as being in many instances little more than noise or ink marks, meaningful chiefly in being symptomatic of the speaker’s or writer’s neurosemantic state. That is to say, it is more revealing as behavior than as language; it symbolizes the speaker more than it symbolizes anything he may appear to be speaking about.

This rather long discussion of the EAI has been given in order to make more or less clear, not only the basic operations
involved in its computation, but also certain of its implications. The EAI of a term, computed from data obtained under adequate conditions, is indicative of one of the most important characteristics of word usage, the relatively precise degree to which words may be regarded as factually meaningful—or vague.

Use of the measure requires that it be computed from data obtained under known and specified conditions; moreover, the particular form of the basic formula to be used in computing it will vary somewhat with the nature and purpose of the investigation. It is proposed that preliminary work to be done in the present program will involve construction of a test by means of which EAI's for a number of different terms can be determined under a variety of conditions. Work already done indicates that the reliability of such a test can be expected to be quite high, that its administration and scoring offer no insurmountable problems, and that data obtained by means of it will reveal differences between words and between individuals and groups.

In the administration of this test it is planned that the subject will be given a word in a standard context, as in the statement: “Point to the pictures that show people doing good things.” The subject then points to such pictures, among a standard set of pictures, as to him represent referents of good as so used. Each picture in the set is numbered and the number of each picture to which the subject points is recorded. From data so obtained from each of a group of subjects, the EAI of each word in the test is to be computed, as was done for the Gallup poll data presented in the preceding pages.

As part of the present program, a study has been made by J. Wilson and the writer (7) in which graduate students and instructors in psychology defined extensionally, by reference to a list of statements taken from psychology texts, the terms “law,” “theory,” and “hypothesis.” The mean EAI’s obtained were .62 for “law,” .40 for “theory,” and .28 for “hypothesis.”

**Extensional Synonymity Index (ESI).** Such EAI’s represent the relative degree of vagueness of words as used. By treating the test data in other ways, they can be made to yield two other types of information, represented by an extensional synonymity index (ESI) and an extensional conformity index (ECI), respectively. By recording the percentage of all the subjects who point to each picture, or other types of referent, in defining each word, it is possible to measure the degree of synonymy between any two words. The formula \( \text{ESI} = \frac{c}{\sqrt{xy}} \), in which \( c \) represents the percentage of subjects pointing to a given picture in defining both of two given words, and \( x \) and \( y \) represent the percentages of subjects pointing to the picture in defining each of the two words, respectively. This value is to be computed for each picture, and the values thus obtained for all the pictures are to be averaged in deriving an expression of the mean degree of synonymity between any given pair of words.

**Extensional Conformity Index (ECI).** The percentages of subjects pointing to each picture in defining each word can also be used as word-fact relating behavior norms. Thus, the pictures may be “weighted” according to these values, and on the basis of them the pointing or labeling of a given individual can be evaluated. For example, if a given individual in defining the word “good” were to point to certain pictures, he
would be showing less conformity to the group that he would be in pointing to certain other pictures. The mean of the percentage values of the pictures to which an individual points in defining a given word would represent his degree of conformity to the group in his extensional use of that word. We may call this his extensional conformity index (ECI), and individual differences expressed in terms of the ECI might be found to be a factor in personality adjustment.

The Intensional Agreement Index (IAI) expresses the degree of agreement among \( n \) persons in defining a given term intensionally—i.e., by giving its verbal equivalents. A dictionary definition is to be regarded as an intensional definition, as the term is here used. Like the EAI, the IAI may range in value from 0.0, representing no agreement, to 1.0, representing maximum possible agreement.

In a preliminary study carried out by N. Whitman and the writer (8), an attempt was made to determine IAI's for each of certain terms used in the field of psychology (learning, perception, emotion, and personality) and certain terms used in the field of biochemistry (fats, lipids, enzymes, oxidation, and basal metabolism). Textbooks in each field were examined until for each term six definitions (from six different authors) had been found. These definitions were then edited so as to exclude all words except nouns, verbs, adjectives, and adverbs (the adverbs when and where, the adjectives that, these, those, and which, and articles used as adjectives were also excluded). Then for each term the number of types (different words) used in all six definitions was recorded, and the number of definitions in which each type was used was determined. The number of obtained agreements, in the use of any given type by the six textbook authors, was found by means of the formula \( (n - 1) .5n \); in which \( n \) represents the number of definitions in which the type occurred. The values thus obtained for the various types were summed in determining the total number of obtained agreements shown by the six textbook writers in verbally defining the term in question. The maximum possible number of agreements was computed by using the formula \( x(n - 1) .5n \), in which \( n \) represents the total number of definitions, six in each case, and \( x \) represents the total number of types used in all the definitions. The maximum possible number of agreements was then divided into the obtained number of agreements in determining the IAI of a given term. The IAI's as thus determined, were:

<table>
<thead>
<tr>
<th>Psychological terms</th>
<th>No. of types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>.024</td>
</tr>
<tr>
<td>Perception</td>
<td>.006</td>
</tr>
<tr>
<td>Emotion</td>
<td>.010</td>
</tr>
<tr>
<td>Personality</td>
<td>.007</td>
</tr>
<tr>
<td>Ave.</td>
<td>.012</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Biochemical terms</th>
<th>No. of types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fats</td>
<td>.080</td>
</tr>
<tr>
<td>Lipids</td>
<td>.150</td>
</tr>
<tr>
<td>Enzymes</td>
<td>.127</td>
</tr>
<tr>
<td>Oxidation</td>
<td>.067</td>
</tr>
<tr>
<td>Basal Metabolism</td>
<td>.035</td>
</tr>
<tr>
<td>Ave.</td>
<td>.092</td>
</tr>
</tbody>
</table>

The difference between the mean IAI's may be regarded as indicating a measurable difference between the fields of psychology and biochemistry with regard to the degree of terminological agreement that has been achieved in them to date. One important aspect of scientific development is to be observed in the fact that among biochemists at the present time there is a tendency to abandon the term fats in favor of the term lipids—a tendency to replace one term with another that has a higher IAI. Increasing agreement as to definitions, both intensional and extensional,
is a basic characteristic of the development of a science; and a means of measuring the degree of agreement that has been achieved within the various fields makes possible a peculiarly objective comparison of them in this important respect. Degree of similarity among verbal formulations generally can be measured in terms of the IAI.

The procedure followed in the above study of psychological and biochemical terms can be modified in at least three ways. First, the definitions can be obtained directly from the subjects rather than from textbooks or other published material. Second, the subjects can be instructed to define each word by listing synonyms of it, and the number of synonyms to be listed can be limited. Third, the words to be defined need not be presented only in isolation, but they may be presented also in context, other words to be substituted by the subject for the word in question, or a definition to be written for the word as used in the particular context. The influence of differences in context on the meaning, and on agreement as to the meaning, of specific words can thus be investigated.

**Intensional Synonymity Index (ISI).** From data of the type just discussed it is possible to obtain measures of intensional synonymity. Degree of synonymity of given pairs of words defined extensionally can be measured by means of procedures already described. Similar procedures can be used in the present connection. For example, suppose the words good and worthwhile to have been defined by each of 100 subjects, each of whom defined each word by listing three synonyms for it. The degree of intensional, or verbal, synonymity between these two words can then be computed by means of the formula

$$\frac{c}{\sqrt{xy}}$$

in which c represents the number of terms (types) given by the 100 subjects as synonyms for both words, and x and y represent the number of terms (types) listed as synonyms for each of the two words, respectively. The correlation between extensional and intensional synonymity indexes would be of interest.

**Semantic vocabulary test.** As has been indicated previously in this outline, vocabulary measures can be obtained from a language sample obtained from any given individual in terms of type-token ratios, type frequencies, proportionate vocabulary, and standard frequency vocabulary. Another type of vocabulary test might be attempted. A common criticism of ordinary vocabulary tests is that while they are indicative of the number of words an individual "knows" or "recognizes," they are not necessarily indicative of the range of "depth" of the individual's knowledge of or skill in using each word that he "knows." The problem raised by this criticism involves technical difficulties, but certain approaches to its solution appear to be possible.

Investigation could be made of the possibility of constructing a vocabulary test of such a nature that the individual's ability to use each word would be sampled in detail. It is possible to distinguish types of meaning, such as meaning in terms of use, variety, differentiating characteristics, sources, etc. For example, the word orange can be defined in terms of (a) the various uses of oranges, (b) the kinds of oranges, (c) the characteristics that differentiate oranges from other things, (d) the geographical areas where oranges are grown, the methods by which they are grown, the history of these methods, etc., and (e) in terms of the scientific research that has been done on oranges, the methods used in
picking, packing, processing, marketing, transporting, etc. This does not exhaust the problem of defining orange, but it illustrates the possibility of devising a vocabulary test of a type that should make possible a measure of vocabulary “depth” as well as “range.”

Measures of “allness.” Previous mention has been made of “allness” terms, such as all, everyone, nobody, every, never, absolutely, etc. Language spoken during moments of anger or despair, or other relatively profound affective states, appears to be particularly characterized by such terms. They give to language a character which reflects what is usually referred to as dogmatism, or stubbornness, inflexibility, etc. Orientation on the basis of dichotomies, or of the excluded middle—a two-valued, either-or orientation—appears to be basic to and to be fostered by, this sort of language. The degree to which one is prone to two-valued orientation is probably an important aspect of one’s general adjustment, personality development, intelligence, etc. Insofar as it might prove possible to set up rigorous criteria of allness terms, the frequency of their use in language samples could be studied. Another approach to the study of allness, however, is also to be proposed. From one point of view allness may be regarded as manifested in extreme responses in situations where they are not mandatory. An attempt could be made to construct a reliable test involving, say, 100 items, to each of which a response can be made along a graduated scale expressive of extreme and intermediate degrees of preference, attitude, behavioral tendency, etc. At least five and possibly seven or more alternative responses to each item should be provided, one expressive of neutrality or average tendency and the others distributed on either side and graduated toward the two extremes. The test would be scored, not in terms of the preferences, etc. expressed, but in terms of the proportion of extreme (allness) responses. It is anticipated that two main types of evaluative tendencies might be indicated by such a test, the tendency to give extreme responses, or allness, and an extreme tendency to give indecisive, indefinite, neutral responses. The latter might characterize certain schizoid conditions, for example. It is to be noted that this type of test should get away from one common weakness of pencil-and-paper tests, in that the effect of falsified responses on the score will be minimized, since the “intensity” rather than the “content” of the responses will determine the score.4

Tests of verbal differentiation. It would appear reasonable to assume that the adequacy of generalization or “abstract thinking” depends largely upon the adequacy of the analysis or differentiation upon which the generalizing is based. This is indicated by an examination of practically any generalization process; it is especially obvious, perhaps, in medical diagnosis. The ability to observe, respond to, and relate differences would appear to limit the ability to abstract similarities effectively. In fact, abstracting (roughly, generalizing) can be defined as a process of leaving out details or differences; similarities are recognized and formulated in accordance with the way differences are disregarded, not observed, or related. Consciousness of abstracting (9, 10), therefore, in any given instance, is seen to depend on an awareness of the differences that are being disregarded or related in the abstracting of similarities.

4Previous work suggestive of this approach has been reported by Watson (19).
It is proposed to construct a test specifically designed to measure an individual's ability to express differences, or to perform verbal differentiation. It is the intention to begin with the simple procedural plan of presenting the subject with pairs of objects, designs, etc. and requesting him to tell the differences between them. A time limit, to be determined, is to be set for each response. An attempt is to be made to score the responses in each of three ways. First, the mere length of response is to be measured; it is hardly to be expected that this will suffice, except possibly as a very gross measure. Second, the number of differences enumerated is to be noted; it will be necessary to formulate rigorous criteria of a “difference.” Third, various forms of the type-token ratio are to be tried as possible expressions of the subject's level of performance.

Assuming the construction of a reliable test, scores on the test are to be related to other variables. The relation of differentiating ability to intelligence, as measured by current standard tests, and to other criteria of competence, is of particular interest.

**SUPPLEMENTARY MEASURES**

The entire research program here proposed involves not only the language behavior measures discussed above, but also certain other measures which are to be used in order to obtain data concerning the relation of the language measures to other aspects of behavior. Among these supplementary measures are tests of intelligence, measures of mental and chronological age, achievement and aptitude tests, measures of silent and oral reading, of speech and writing, and various indices of personality.

**STUDIES COMPLETED**

To date six studies have been completed, and a considerable amount of preliminary or exploratory investigation has been done. The six completed studies have been done by Fairbanks (5), Mann (11), Chotlos (4), Tuthill (18), Johnson and Whitman (8), and Johnson and Wilson (7). The investigations so far completed have been concerned mainly with problems of method, although they have been designed to contribute, also, to a fuller understanding of language behavior in its various relationships.

Individual 3,000-word spoken language samples were obtained by Fairbanks (5) from each of 10 schizophrenic patients and 10 university freshmen. Mann (11) obtained 2,800-word written language samples from each of 24 schizophrenic patients and 24 university freshmen. The writer has obtained 3,000-word written samples from each of approximately 1,000 Iowa public school children, selected on the basis of age, sex, I.Q., type of school (rural, town, city) and socio-economic status. A selected set of 108 of these written language samples have been analyzed in considerable detail by Chotlos (4).

The studies which follow the present article in this monograph will serve to illustrate some of the above types of approach to the investigation of language behavior. The present discussion is offered as a general introduction to these and to the further studies that will, it is hoped, be included in the program of research which has here been outlined.

---

*Acknowledgement is hereby made of grants from the Work Projects Administration for Iowa Projects 4,892 and 5,960, by means of which these data were obtained and tabulated, and to Professor George D. Stoddard, then Director of the Iowa Child Welfare Research Station, who sponsored the projects.
BIBLIOGRAPHY


5. Fairbanks, H. Studies in language behavior: II. The quantitative differentiation of samples of spoken language *Psychol. Monogr.,* 1944, 56.

6. Horn, E. *A Basic Writing Vocabulary: 10,000 Words Most Commonly Used in Writing.* Univ. of Iowa Monogr. in Educ., 1926, No. 4, 225 pp.

7. Johnson, W., & Wilson, J. The degree of extensional agreement among twenty psychologists in their use of the labels hypothesis, theory and law. Unpublished research.


