ED 017 714 VT 004 635

THE VOCABULARY OF INSTRUCTIONAL MATERIALS FOR TYPING AND STENOGRAPHIC TRAINING, RESEARCH FINDINGS AND IMPLICATIONS.

BY- WEST, LEONARD J.
CITY UNIV. OF NEW YORK, DIV. OF TEACHER EDUCATION

REPORT NUMBER NYCU-RR-67-10 PUB DATE OCT 67
EDRS PRICE MF-$0.25 HC-$0.88 20P.

DESCRIPTORS- *TYPEWRITING, *STENOGRAPHY, *INSTRUCTIONAL MATERIALS, *VOCABULARY, *STRUCTURAL ANALYSIS, SYLLABLES, WORD FREQUENCY, SILVER)HORN,

THE SILVERTHORN BASIC VOCABULARY OF WRITTEN BUSINESS COMMUNICATION WAS REANALYZED IN ORDER TO FURNISH ACCURATE DIFFICULTY INDEXES FOR INSTRUCTIONAL AND TEST MATERIALS FOR STENOGRAPHERS AND TYPISTS. AMONG THE 11,055 DIFFERENT WORDS IN THE REANALYZED LIST, 109 OCCUR AT LEAST ONCE IN EVERY 1,000 WORDS. MEAN SYLLABIC INTENSITY (NUMBER OF SPEECH SYLLABLES PER DICTIONARY WORD) WAS FOUND TO BE 1.54. MEAN STROKE INTENSITY (NUMBER OF TYPEWRITER STROKES PER DICTIONARY WORD) WAS FOUND TO BE 6.0 (4.67 LETTERS PLUS 1.0 SPACES PLUS .3 FOR THE INCIDENCE OF PUNCTUATION). THESE VALUES SUBSTANTIALLY EXCEED AND SHOULD REPLACE THE CONVENTIONAL 1.40 AND 5.0 ESTIMATES. OTHERWISE, STENOGRAPHERS AND TYPISTS WILL BE UNDERPREPARED FOR THEIR JOBS, AND THEIR PROFICIENCY WILL BE OVERESTIMATED. IT WAS ALSO FOUND THAT THE VALIDITY OF "PERCENTAGE OF COMMON WORDS" AS AN INDEX OF DIFFICULTY DEPENDS ON THE LENGTH OF THE COMMON-WORD LIST. WHEN THE LIST IS SHORT, THAT INDEX HAS A NEAR-ZERO CORRELATION WITH FREQUENCY OF OCCURRENCE IN THE LANGUAGE. FURTHER, SHORTHAND DICTATION ON THE BASIS OF THE "STANDARD WORD" (EVEN AT SYLLABIC INTENSITY 1.54) IS AN INSUFFICIENT EQUALIZER OF DIFFICULTY. THE ADDITION OF "PERCENTAGE OF WORDS AMONG THE 1,500 (OR 2,000) COMMONEST" IS RECOMMENDED. THE PERCENTAGE DISTRIBUTION OF CUMULATIVE SEGMENTS OF THE BUSINESS VOCABULARY IS GIVEN, AND THE PROBABLE IMPROPRIETY OF USING A BUSINESS VOCABULARY AS A BASIS FOR TRAINING AND TEST MATERIALS IN PERSONAL TYPING COURSES IS DISCUSSED. (AUTHOR)

DIVISION OF TEACHER EDUCATION OF THE CITY UNIVERSITY OF NEW YORK
535 EAST 80TH STREET, NEW YORK, N. Y. 10021

Office of Research and Evaluation
RESEARCH REPORT

67-10

THE VOCABULARY OF INSTRUCTIONAL MATERIALS FOR TYPING AND STENOCRAPHIC TRAINING--RESEARCH FINDINGS AND IMPLICATIONS

by

Leonard J. West

THE CITY UNIVERSITY OF NEW YORK
Office of Research and Evaluation
Division of Teacher Education

October 1967

The Vocabulary of Instructional Materials for Typing and Stenographic Training--Research Findings and Implications

Leonard J. West

Division of Teacher Education
The City University of New York

The issue treated here is that of optimal vocabulary in the instructional materials for typing and stenographic skills. The fundamental requirement is for the use of vocabulary that will best prepare the individual for the words he will encounter later, either occupationally or in personal uses. A necessary corollary is the use of test materials that will most validly assess the trainee's readiness for employment or the adequacy of his personal skills. The pertinent underlying concept or principle is that of maximum positive transfer (from school training to later life performance), an outcome that requires as close as possible a match between the materials of the training and those of later life. What are the true characteristics of the vocabulary used by typists and stenographers? The answer to that question identifies appropriate instructional materials.

Details are given on the writer's re-analysis of Silverthorn's basic vocabulary of written business communication, and the findings are discussed within the larger framework of earlier research findings. In the light of pertinent educational and occupational data on a nationwide basis, the propriety of basing instructional materials on a business vocabulary is reconsidered. In addition, the applicability of various indices for characterizing vocabulary to the construction and selection of training and test materials for stenographers and typists is discussed.

Modes of Characterizing Vocabulary

In typewriting and stenographic training, there are three commonly used modes or indices for describing the vocabulary of the practice or test materials: (a) per cent of frequent words, (b) stroke intensity (average number of typewriter strokes per dictionary word), and (c) syllabic intensity (average number of speech syllables per dictionary word). These indices may be illustrated via the sentence Please arrange to de-

liver the goods today. The italicized sentence contains 7 dictionary words, 42 typewriter strokes (including interword space and the period), and 11 speech syllables. If the 1,000 most frequently occurring words in the language are arbitrarily designated the "common" ones, then, 5 of the 7 words, or 71 per cent of them, are "common." The stroke intensity of the illustrative sentence is 6.0 (42/7), and its syllabic intensity is 1.57 (11/7).

It is important to understand what underlies these indices. The real variable or factor that determines difficulty is the amount of practice that has been given to the words. Other things being equal, the easy words are those that have been subjected to much practice. Since, in ordinary prose, the words that will tend to have been heavily practiced are the ones that commonly occur in the language, "per cent of common words" is a measure of the probable amount of practice that has been given to the materials. The other two indices, syllabic and stroke intensity, arise from the supposition that the commoner words in the language tend to have fewer letters and syllables than less common words; the higher its frequency of occurrence in the language, the shorter the word. Accordingly, average stroke or syllable length of words has conventionally been assumed to be an indirect measure of frequency of occurrence in the language, which is, as earlier stated, a measure of probable amount of practice given to the word.

In any event, "per cent of frequent words" as a means of describing the probable difficulty of prose materials requires an arbitrary decision as to how many words will be considered frequent or common. Shall it be the commonest 100? 500? 1,000? 2,000? 5,000? Obviously, a passage 75 per cent of whose words are among the 500 commonest in the language differs greatly from one in which three-fourths of the words are among the 2,000 commonest in the language. The former passage will tend to contain a far narrower vocabulary.

Concerning stroke intensity, since the adoption, in 1924, of the five-stroke word as the basis for scoring typewriting performance, it has been assumed that the average word in the language contains four letters (plus a space before the next word). Syllabic intensity was first advocated as a descriptive index for shorthand materials by Dr. Gregg and announced by Leslie (1931), who offered 1.40 as the average syllabic intensity of the language without specifying the source of that estimate in his published article; however, Leslie's later textbook (1949, p. 198) explains that 1.40 was the average syllabic intensity used in shorthand speed-contest material. Shorthand speed-contest material, it should be clear, is shorthand speed-contest material and may not be assumed to be representative of the vocabulary of written business communication. In fact, the findings of the present investigation show 1.40 to be a serious underestimate of that vocabulary. In any event, for years now, syllabic intensity has been used to characterize shorthand materials, and dictation speeds have been based on a "standard word" of 1.40 syllables; a dictation rate of 50 wpm means 70 speech syllables per minute, not 50 dictionary words. Both syllabic and stroke intensity have been commonly applied to typewriting materials, and the latest edition of at least one major typwriting textbook also uses per cent of frequent words as an index of difficulty.

The existence of a count of the vocabulary of written business communication (Silverthorn, 1955) makes possible re-examination of the conventional suppositions about instructional and test materials for stenographers and typists, suppositions that should have been called into question by data that were available years ago. For example, Dewey's estimate (1923) was 5.7 typewriter strokes for the average English word (including spacing and punctuation). Miller (1951, p. 798) gave 1.7 as the average syllabic intensity for English. More recently, Mellinger (1964) recomputed the syllabic intensity of the Silverthorn list of 11,564 different words found among 300,000 words of written communication as 1.56, using an exact count for most of the list and an estimate for the remainder of the list. The present study provides stroke intensity and syllabic intensity information for the Silverthorn list, using a modified definition of a "word," one that is in accordance with the practices of specialists in vocabulary study (Lorge and Chall, 1963).

National Data on the Use of Typing and Stenographic Skills

Assessing the vocabulary characteristics of typing and stenographic materials requires an immediate distinction between personal and occupational uses of these skills. Occupational uses no doubt swamp personal uses of stenographic skills--so that instructional materials based on a business vocabulary are probably appropriate in shorthand classes. The reverse appears to be true of typing skills, for which personal use appears to swamp occupational use, as inferred from national data on employment, registration in typing classes, and typewriter sales.

Consider the facts. In 1960, 2.3 million persons (3.5 per cent of the full-time labor force) were employed as "stenographers, typists, and secretaries" (Rutzick and Swerdloff, 1962). In the same year more than a fifth of all public day secondary school students were enrolled in a typing class, 70 per cent of them in one-year courses (Wright, 1964, 1965). If the 1960 typing registration percentage (21.17 per cent) is applied to the total Fall, 1967, public secondary school enrollment of 12.3 million students (as reported by the Saturday Review of Literature on the basis of U.S. Office of Education figures), the result is a typing enrollment of 2.6 million high school students in 1967. Assuming, for illustrative purposes, stability in high school enrollment, percentage registration in typing classes, and percentage of typing enrollment in one-year courses, in a four-year period 8.8 million different high school students are taught to type: 4(.7 x 2.6 million)+ 2(.3 x 2.6 million) = 8.84 million (exclusive of nonpublic schools). In contrast, employment projections through the mid-1970's for stenographers, typists, and secretaries do not exceed 3 to 3½ million persons (U.S. Department of Labor, 1963). Clearly, many more persons are taught to type than are or will be gainfully employed in occupations that make major use of the typewriter. In fact, the writer of an economics column for a metropolitan newspaper (Porter, 1966) estimated that 35 million Americans use the typewriter in one fashion or another--about one in every six persons from newborns through centenarians.

Consider, next, that in 1966 there were 140 portable typewriters sold for every 100 standard machines (U.S. Bureau of the Census, 1967). Granting the probable use of portable typewriters for occasional job-related tasks among persons who are not themselves stenographers, typists, or secretaries, the portable typewriter is primarily a personal-use, not an occupational-use machine. That is, the vocabulary of business communications is probably not especially prominent in personal-use typewriting. On this score, Featheringham's survey of personal typing activities (1965) shows letters, manuscripts and speeches to be the three commonest personal typing activities of out-of-school adults who had a personal-typing course in high school. Whether the letters are business or personal letters was not specified. But surely the manuscripts and speeches are not ones whose vocabulary is of the "Dear Sir: Thank you for your order of April 18" variety.

Taken together, the data on occupations, school enrollments, and typewriter sales make it apparent that in this country typewriting is more a personal than an occupational skill. That inference calls into question the use of a business vocabulary as a primary basis for instructional and test materials for typists; for there are important differences between specialized and general vocabularies, between narrowly and broadly based ones.

Some Characteristics of Vocabulary Studies

As Godfrey Dewey pointed out in his classic "Relativ Frequency of English Speech Sounds" (1923), the number of different words and their frequency of occurrence found in any vocabulary study will depend on the diversity of the source materials and the size of the sample of words drawn from the source materials. In his own study of 100,000 words of "diversified" materials (newspaper editorials and news columns, modern fiction, speeches, personal and business correspondence, religious English, popular scientific English, popular magazine articles and editorials), Dewey found 10,019 different words, 118 of which occurred at least once in every 1,000 words. Silverthorn (1955) found 11,564 different words in his examination of 300,000 words of written business communication, 107 of which occurred at least once in every 1,000 words. Thorndike and Lorge found 30,000 different words in their tally of approximately 4½ million words (1944).

Differences also exist in how a "word" is defined. Dewey lists separately root words and "particular" words (variants of root words, e.g., seem, seemed). Silverthorn counted all variants as separate words, but Thorndike and Lorge did not do so (e.g., look, looks, looked, looking, were counted under the "main word" look). Silverthorn's study differs uniquely from all other studies in counting as single words compound expressions that do not occur in the dictionary as such but represent the particular style of the writer (e.g., easy-to-use, do-it-yourself). In fact, that hard-to-defend feature of Silverthorn's original word was a major stimulus for the writer's re-analysis of the Silverthorn list.

Another general characteristic of vocabulary studies is the near impossibility of agreement among studies on even the 100 commonest words;

even they will vary depending on the source materials examined. Strikingly, the three studies mentioned above do not even agree on the 10 commonest words. Perhaps the sharpest illustration of the differences that arise between diversified and narrow source materials is the presence among the 100 commonest words in Silverthorn's business list of *Mr.*, *dear*, *sincerely*, *truly*, *order*, *business*, *please*, *service*, *office*, *sales*, *price*, *gentlemen*, *enclosed*, *manager*, *department* precisely because such words occur often in business communication. But *Mr.*, *sincerely*, *sales*, *gentlemen*, *manager* are not included in the first 1,000 from Dewey's greatly more diversified source materials, while the others range from the 200 to the 1,000 level in Dewey's list. Accordingly, there is grave risk in too heavy a concentration of typing training materials around a business vocabulary when, as it appears, only a modest proportion of all those taught to type will in fact use the skill occupationally.

However, for stenographic training and for vocational typing, the characteristics of a business vocabulary are pertinent, and the findings from the writer's re-analysis of Silverthorn's "Basic Vocabulary of Written Business Communication" (1955) are presented next.

Re-Analysis of the Silverthorn Business Vocabulary

The Silverthorn vocabulary is based on examination of 300,000 words of written business communication, using a proportional, stratified sample in order to give appropriate representation to the specialized vocabularies that might exist in various types of business. That is, the number of words examined from each type of business was in proportion to the number of the nation's "stenographers, secretaries, and typists" employed in that type of business. For example, with .275 per cent of the nation's secretarial personnel employed in "agriculture, forestry, and fishery," .275 per cent of the 300,000 words examined (825 words) were from communications in that area of business. Words were tallied in cumulative blocks of 10,000 until the rank order of the commonest 100 words was unchanged by the addition of a new block. A total of 300,000 words was found necessary before that criterion was reached. Thus, the Silverthorn list is accurate for the first 100 words; the rank order of all remaining words is an approximation. In all, Silverthorn listed 11,564 different words found among the 300,000 tallied.

Procedures

However, the original Silverthorn list contains occasional ungrammatical expressions, and it counts as separate words each different compound expression and each differently spelled version of the same word (e.g., *do-it-yourself* as one "word"; *easy-to-use* is another; *enc*., *inc*., *encl*., *incl*., *enclosure* were counted by Silverthorn as five different words). But a language, as Dewey pointed out more than forty years ago, is "a tongue, a speech, a collection of sound patterns" (Dewey, 1926, p. 18). *Enclosure* is one "word," however spelled and whether abbreviated or not. For that reason and also because the stenographer and typist do not respond to compounds as if they were single words (the stenographer lifts his pen and the typist makes an appreciable pause at the point of a hyphen), it was judged preferable to define a "word" in the fashion used in the many dozens of other vocabulary studies that have been carried out. Accordingly, in the present investigation all compounds that do not occur as such in the third edition of Webster's unabridged dictionary were broken up and the frequencies added to those for each of the components making up the compound. In this fashion the 33 occurrences of *to* in 16 different compounds (e.g., *easy-to-use*) were added to the original frequency of *to* of 9,704 to make a new frequency of 9,737. Abbreviations were treated in analogous fashion; e.g., the frequencies for each of the various forms of *enclosure* were summed for the one lexical unit originally represented in five different ways orthographically. However, for the purpose of computing average stroke length of words, abbreviations and spelled-in-full words were credited separately. The result of these various regularizations was to reduce the number of different words from Silverthorn's reported 11,564 to 11,055. Another result was to increase from 107 to 109 the number of different words that occur at least once in every thousand words of business communication.

A second procedural feature arises from the ties in frequency often found at the borders of the class intervals commonly used for describing segments of a vocabulary. For example, the four words that occurred 124 times in Silverthorn's list were originally reported in alphabetical order as occupying rank positions 299-302. In order to determine which two of the four belong in the 201-300 interval and which two in the 301-400 interval, it was assumed that if Silverthorn had examined a larger amount of business communication, these ties would have been broken and the frequencies would approximate those found in larger vocabulary studies, specifically in the Thorndike-Lorge study of $4\frac{1}{2}$ million words leading to their list of 30,000 words (1944). Illustratively, the different frequencies shown by Thorndike and Lorge for each of the four words shown as tied in frequency by Silverthorn were used to determine the probable rank order of these four words and, in turn, the class intervals to which they should be assigned. The same procedure was followed for all ties in frequency at the borders of class intervals.

Concerning the class intervals into which the entire vocabulary of 11,055 words were divided (the commonest 100, words 101-200, etc.), for lower frequency words it was not possible to employ round-number class intervals. The breaks were made at the point of changes in frequency. For example, words 2001-2446 were those that occurred from 16 to 12 times in 300,000 words, words 2447-2944 occurred from 11 to 9 times in 300,000 words, and so on.

For the new list of 11,055 different words, treated as described, each word's frequency, number of letters, and number of syllables were punched on tabulating cards. Via computer, number of letters and syllables was multiplied by frequency for the words in each interval (words 1-100, 101-200, and so on), and the sum of these products was divided by the sum of the frequencies for the words in the interval.

Results and Discussion

For the 11,055 different words of the re-analyzed Silverthorn vocabulary, Table 1 shows (at the left) the number of different words accounting for various percentages of all business communication. At the right, the same information is shown in reverse fashion: the percentage of all usage accounted for by various numbers of words.

The impressiveness of the Table 1 testimony to the economy with which communication is carried out can make one lose sight of the fact that practice confined to a 1,000-word vocabulary will leave every fifth word untouched. Practice at a 2,000-word vocabulary will leave every tenth word unpracticed. Dewey (1923, p. 6) has pointed to the "interesting and significant fact that some of the commonest sillables of the language scarcely occur among the 500 or even 1,000 commonest words, but owe their importance rather to occurrence in many different words each relativly infrequent.[1] Among these are such suffixes as -ance, -ment, -ity, -less, -ness, -ful, and others.

Table 1

Percentage Distribution of Cumulative Segments of the Vocabulary of Written Business Communication

Per Cent of Communication	Number of Different Words	Number of Common Words	Per Cent of All Usage
10	2-3	5	16.8
20	6	10	24.8
25	10	25	37.3
30	14	50	45.4
33-1/3	18	100	53.1
40	31	200	61.0
50	76	500	72.4
60	183	1,000	81.4
66-2/3	317	2,000	89.5
70	413	2,500	91.7
75	608	5,000	96.9
80	895	11,055	100.0
90	2,096		
95	3,727		
100	11,055		

Table 2 shows the unweighted and weighted stroke intensity and syllabic intensity of various segments of the re-analyzed Silverthorn vocabulary, successively and (below the dashed line) cumulatively. The unweighted values (or, strictly speaking, the unit-weighted values) give each word equal weight regardless of differences in frequency of occurrence. The weighted values are according to frequency and are the pertinent ones for instructional materials.

[1] Dewey was an advocate of simplified spelling; thus: sillables, relativly, hav, fonetic, publisht, and many others.

Table 2

Unweighted and Weighted Syllabic and Stroke Intensity of Successive and Cumulative Portions of the 11,055 Different Words in Re-analysis of Silverthorn's Vocabulary of Written Business Communication

Interval	Occurrences in 300,000 Words[a]	Syllabic Intensity		Stroke Intensity	
		Unweighted	Weighted	Unweighted	Weighted
1- 100	16252 - 330	1.36	1.09	4.01	3.03
101- 200	325 - 179	1.81	1.52	5.51	4.96
201- 300	179 - 124	2.12	1.86	6.25	5.73
301- 400	124 - 96	2.17	2.02	6.72	6.45
401- 500	96 - 77	1.97	1.94	6.23	6.18
501- 750	77 - 51	2.05	2.00	6.48	6.32
751- 1000	51 - 38	2.32	2.23	7.21	7.00
1001- 1500	37 - 23	2.18	2.14	7.07	6.95
1501- 2000	23 - 16	2.31	2.26	7.30	7.21
2001- 2446	16 - 12	2.37	2.35	7.55	7.51
2447- 2944	11 - 9	2.49	2.45	7.69	7.61
2945- 3824	8 - 6	2.45	2.43	7.66	7.62
3825- 4917	5 - 4	2.46	2.46	7.66	7.67
4918- 5822	3	2.50	2.49	7.84	7.83
5823- 7403	2	2.55	2.55	7.96	7.96
7474-11055	1	2.59	2.59	8.13	8.13
First 500		1.88	1.27	5.74	3.73
First 1000		2.04	1.36	6.29	4.04
First 2000		2.14	1.44	6.74	4.31
All 11055		2.46	1.54	7.70	4.67

[a] As explained in the "Procedures" section, ties in frequency at the borders of class intervals were broken by recourse to the larger vocabulary study of Thorndike and Lorge (1944).

It may be noticed in Table 2 that fully one-third of the words in the business vocabulary occur only once in 300,000 words and that more than half of them occur fewer than four times in 300,000 words. At the same time, as was mentioned earlier, certain parts of words (syllables and letter sequences) occur with high frequency in substantial numbers of words, each of which, as a word, is of low frequency. The enormous effect of the shortness of the very commonest words is also apparent in the differences between the weighted and unweighted indices. A mean of 7.70 letters per word is reduced to 4.67 when frequency is taken into account; syllabic intensity is comparably reduced from 2.46 to 1.54.

In the weighted values in the bottom row of Table 2 lie the consequential values of 1.54 and 4.67: the true ones for the vocabulary of written business communication. The conventional estimate of 1.40 as the average number of syllables per word applies to something between the first 1,000 and 2,000 commonest words in business communications. If about the first 1,500 commonest words may be taken as an approximation of the number whose mean syllabic intensity is 1.40, then it is self-evident that nearly seven-eighths of the vocabulary of written business communication lies beyond that level. In the same fashion, the conventional estimate of 5.0 typewriter strokes per word applies only to the 1,000 commonest words (4.04 letters plus 1.0 spaces); the remaining 10,000+ words lie beyond that level. Materials that are fully representative of business vocabulary have a syllabic intensity of 1.54. Average stroke intensity approximates 6.0: 4.67 letters plus 1 space bar stroke plus .3 for the incidence of punctuation marks, as estimated by Dewey in his shorthand dissertation (1926) on the basis of his earlier (1923) investigation of the sounds, syllables, and words of written English prose from diversified sources. The conventional 1.40 and 5.0 figures are substantial underestimates of the 1.54 syllabic intensity and 6.0 stroke intensity that are the true values for the vocabulary of written business communication as originally collected by Silverthorn and re-analyzed here.

Correlational Data. The assumption that underlies the use of syllabic or stroke intensity as vocabulary indices is that they are highly correlated (inversely) with frequency of occurrence, the commoner words presumably having fewer letters and syllables. While words with fewer letters strongly tend to have fewer syllables (r = .792), letters and syllables show hardly any correlation with frequency of occurrence. For the re-analyzed Silverthorn vocabulary, r = -.080 was found between number of syllables and frequency of occurrence; for strokes (i.e., letters) and frequency, r = -.113 was found. There is only the tiniest tendency for the words that have fewer letters and syllables to be the more frequent ones. Superficially, it might seem that both syllabic and stroke intensity should be discarded as acceptable substitutes for the frequency of occurrence in the language that is the real determinant of the probable difficulty levels of instructional materials for typists and stenographers. However, the near-zero correlations are an artifact of dealing with the actual frequencies of each word among the 11,055 different ones found. The clear tendency for increases in letter and syllable length with decreases in frequency for the commoner words on the list is swamped by the thousands of words above the 5,000 level that vary in letter and syllable length but that share the same frequencies of 3 or 2 or 1 occurrence in 300,000 words. The forest is lost sight of for the trees, so to speak. The low correlations reported here were paralleled in Hillestad's shorthand investigation (1962). When she divided the Silverthorn vocabulary into nine frequency levels, her obtained correlation between "vocabulary level" and syllables was .109.

The commonly used "per cent of common words," it turns out, is not at all the same as "frequency of occurrence" as an index. The former yardstick depends heavily on how many words are considered "common." For example, Bell (1949) used the 472 words found in common in three other lists of the thousand commonest words as her "common word list." With so modest a list, she found for 38 samples of copy from typewriting textbooks a correlation of -.84 between stroke intensity and per cent of frequent words and one of -.74 between syllabic intensity and per cent of frequent words. When you expand the definition of common words to those within the first 1,500, the correlation with syllables drops, as in Hillestad's study (1962), to .059. The number of words defined as "common" determines the extent to which "per cent of common words" will be correlated with syllable- (and probably stroke-) length of words; the shorter the list, the higher the correlation.

These correlational findings might be thought to call into question the validity of "per cent of common words"--and, in turn, of syllabic and stroke intensity as equivalents or substitutes--as useful indices of the probable difficulty levels of practice and test materials for typists and stenographers. However, as Table 1 shows, a mere 500 to 1,000 words account for about 70 to 80 per cent of business usage. The commonly used indices of difficulty work because so few words make up so much of the business vocabulary. The correlations drop drastically as the common word list is expanded. Performance hangs so heavily on relatively few words that size of "common word" list is a crucial consideration in using "per cent of common words" as an index. The preference for syllabic or stroke intensity as indices (over "per cent of common words") lies in two considerations: (a) syllabic and stroke intensity are vastly more economical to compute and (b) they do not require an arbitrary decision as to how many words will be considered "common."

Effects of Vocabulary Differences on Performance. For both typewriting and stenographic skills the two criteria of performance are speed and errors. For stenographic skills, the further question is one of whether dictation or transcription performance is the appropriate feature to examine. On that score, since transcription performance depends so heavily on what might be called "word sense" (the ability to infer missing or poorly written outlines from contextual clues), dictation rather than transcription performance seems the appropriate thing to examine in relation to differences in the vocabulary. Here, there has been no research whatever on whether writing speed varies with vocabulary; we do not know whether a person who can write materials at a specified vocabulary level at 80 wpm can write materials at another specified level at, say, 70 or 90 wpm. The only available information is on errors in the notes written from dictation at a presumably appropriate speed (e.g., from dictation at 80 standard wpm to high school students completing a second year of shorthand instruction).

For typewriting, on the other hand, information on both speed and errors as a result of differences in the vocabulary of the test materials is available. Because the decision-making processes that go with placement of so-called "problem" materials are irrelevant to questions of vocabulary, the data for typewriting are for straight copy skills, ordinary copying of printed prose. On the question of the effects on performance of differences in the copy as measured by (a) per cent of common words, (b) stroke intensity, and (c) syllabic intensity, the available information is for speed and errors in straight copy typing and for errors in shorthand notes written from dictation.

For shorthand, Hillestad (1962) found a correlation of .758 between note errors and the number of words in the dictation above the 1,500 level in Silverthorn's list and one of .813 between note errors and "vocabulary level index" (essentially a measure of the rank position of the words in Silverthorn's list). Clearly, the larger the number of less frequent words in the dictation, the higher the incidence of errors in the notes. However, Hillestad's _r_ of .494 between syllables and note errors shows that syllabic intensity is only a moderate index of frequency of occurrence in the language and calls into question basing dictation rates solely on a syllabic basis, as represented in the concept of the "standard word." The r = -.080 between number of syllables and frequency of occurrence in the present study is further and even more powerful testimony to the dubiousness of the conventional mode of marking materials for shorthand dictation at given speeds. The "standard word," by itself, is insufficient. Consider two pieces of copy both marked for dictation at 80 wpm in the conventional way: (a) a 150-word letter 15 of whose words are above the 1,500-word level and (b) a 150-word letter 35 of whose words are above the 1,500-word level. The latter letter will be far more difficult. It appears that copy for dictation should be accompanied by information about the percentage of words in it that are common (i.e., within the first 1,500 or 2,000 or 5,000 words). Even so, the findings of the present re-analysis of the Silverthorn list dictate that the "standard word" be considered to have 1.54, not 1.40, syllables. Fifty words means 77, not 70, syllables; dictation at 80 standard words per minute means 123, not 112, syllables in a minute: three quarter-minutes of 31 syllables each and one quarter-minute of 30 syllables. Since the exact figure (at 80 wpm) is 123.2 syllables per minute, five minutes' worth of dictation (at a standard word of 1.54 syllables) requires 616, not 560, syllables: about 6 more _dictionary_ words per minute than under the present erroneous assumption that the average "business" word contains 1.4 syllables.

The effect of the present use of 1.40 as average syllabic intensity--especially when unaccompanied by information about the vocabulary level of the dictation materials--is to overestimate the true skill of stenographic students and to underprepare them for the vocabulary of real life. The implications are clear enough: use 1.54 as the "standard word," marking materials for dictation accordingly, and accompany the materials with some appropriate measure of vocabulary level, say, percentage of dictionary words above (or within) the 1,500 level in Silverthorn's list.

For straight copy typing, the findings are simply stated. No investigator, among dozens, has been able to show any clear and consistent effect of copy differences on stroking errors. Stroking accuracy is entirely unaffected by differences in syllabic intensity, stroke intensity, or percentage of common words in the copy. The effects are all on gross speed--and quite modest in an absolute sense. Bell's early findings for straight copy errors in relation to stroke intensity, syllabic intensity, and per cent of frquent words (using a 462-word list of frequent words) consisted of correlations of -.23, -.16, and -.10 respectively. Correlations with gross speed were -.61, -.47, and .68 respectively. Of the two easy-to-compute indices (stroke and syllabic intensity), stroke intensity seems slightly the more powerful one. Bell further found that gross speed decreases significantly with increases of .1 in syllabic intensity or of .5 in stroke intensity or with a decrease of 25 per cent in percentage of common words. Differences of these sizes (or greater) make a difference in typing speed.

More recently, Robinson (1967) administered five pieces of copy--varying concomitantly in stroke intensity, syllabic intensity, and percentage of frequent words (withing the first thousand on Silverthorn's list)--to all the several thousand first-year typists in the public high schools of Indianapolis, at 6-week intervals from Week 12 through Week 36. As should be expected from the fact that the beginner is a letter-level typist who has not yet developed any chained responses for the commonly occurring letter sequences in the language, differences in the vocabulary of the copy made hardly any difference in performance during testing at Week 12. The notion of artificially simplified (mostly monosyllabic) practice materials for beginners is a fiction. Since all the novice's stroking is at the same elementary level, it makes no difference what copy is used; he performs as well on long or uncommon words as he does 9n short or common ones. Thereafter (by Week 18 and afterwards) differences in the copy begin to make a difference--increasingly as skill increases with amount of training, as the typist begins to develop more and more chained responses for the commonly occurring letter sequences that appear and reappear in all words, common and uncommon. For example, at Week 18 gross stroking speed on copy at a stroke intensity of 4.00, a syllabic intensity of 1.00, 85 per cent of whose words were within the thousand commonest was four wpm faster than on copy at 7.20/2.00/45%. By Week 36, there was a five-wpm difference in speed on these two pieces of copy. In general during the second semester of training, copy that differed by .8 in stroke intensity, .25 in syllabic intensity, and 10 per cent in percentage of words among the thousand commonest tended to lead to at least one-wpm differences in gross stroking speed for each change in the copy of the sizes mentioned.

Summary

Re-analysis of the Silverthorn vocabulary of written business communication for the purpose of furnishing accurate indices for the vocabulary used in instructional and test materials for typing and stenographic skills identified 11,055 different words occurring in 300,000 words of materials, of which 109 occur at least once in every thousand words. As contrasted with the conventional assumptions of 1.40 speech syllables and 5.0 typewriter strokes per average word, re-analysis (weighted for frequency) showed mean syllabic intensity to be 1.54 and mean stroke intensity to be 6.0 (4.67 letters plus 1.0 spaces plus .3 for the incidence of punctuation). Syllabic and stroke intensity have nearly no correlation with frequency of occurrence (r's of -.080 and -.113) because of the very large number of words in the list that have the same frequency. Further, the propriety of percentage of common words as a vocabulary index depends heavily on the number of words considered "common"; syllabic and stroke length of words are appreciably correlated with a short common word list (472 words), but negligibly correlated with a longer list (1,500 words). Data on occupations, typing registration, and typewriter sales suggest that personal uses swamp occupational uses of the typewriter and call into question the basing of typewriting materials on a business vocabulary.

However, for shorthand and for vocational typing, the implications of the findings are these: (1) If percentage of common words is used as a vocabulary index, a 1,500-word list is greatly preferable to a 1,000-word list. (2) Practice materials with a syllabic intensity of 1.54 or a stroke intensity of 6.0 will most closely match the vocabulary of real-life uses of typing and stenographic skills, and test materials at those levels will most validly assess achievement of occupational objectives; the conventional 1.40 and 5.0 estimates seriously underprepare the trainee and overestimate his true proficiency. (3) Marking of copy for shorthand dictation should use a standard word of 1.54, not 1.40, syllables and the copy should be accompanied by a statement of the percentage of common words contained in it (within the first 1,500 words); by itself, the standard word is not a sufficient equalizer of difficulty.

References

1. Bell, Mary L. Some Factors in Typewriting Difficulty. Unpublished doctor's dissert., U. of Oklahoma, 1949.

2. Dewey, Godfrey. *Relativ Frequency of English Speech Sounds*. Cambridge: Harvard U. Press, 1923

3. Dewey, Godfrey. A Sistem of Shorthand for General Use. Unpublished doctor's dissert., Harvard U., 1926.

4. Featheringham, Richard D. The Validity of Personal-Use Typewriting Courses as Determined by an Analysis of the Practical Application of this Subject Over a Fifteen-Year Period (1950-1964). Unpublished doctor's dissert., U. North Dakota, 1965.

5. Hillestad, Mildred C. Factors that Contribute to the Difficulty of Shorthand Dictation Material. *Delta Pi Epsilon J.*, 1962, Vol. 4, No. 4, 2-18.

6. Leslie, Louis. Providing Controlled Dictation. *Amer. Shorthand Teacher*, 1931, Vol. 12, 21-[illegible].

7. Leslie, Louis. *Methods of Teaching Transcription*. New York: McGraw-Hill, 1949.

8. Lorge, Irving and Chall, Jeanne. Estimating the Size of Vocabularies of Children and Adults: An Analysis of Methodological Issues. *J. Exper. Educ.*, 1963, Vol. 32, 147-157.

9. Magnitude of the American Educational Establishment. *Saturday Review of Literature*, October 21, 1967, p. 67.

10. Mellinger, Morris. Has the Syllabic Intensity Yardstick Lost Its Magic? *Bus. Educ. World*, 1964, Vol. 45, No. 3, 9-11.

11. Miller, George A. Speech and Language. In S.S. Stevens (Ed.) *Handbook of Experimental Psychology*. New York: Wiley, 1951. Pp. 789-810.

12. Porter, Sylvia. Typewriter Boom. *New York Post Magazine*, June 22, 1966, p. 2.

13. Robinson, Jerry W. The Relation of Copy Difficulty to Typewriting Performance. Unpublished doctor's dissert., U. of California at Los Angles, 1965.

14. Rutzick, Max and Swerdloff, Sol. The Occupational Structure of U.S. Employment. *Monthly Labor Rev.*, 1962, Vol. 85, 1209-1213.

15. Silverthorn, James E. The Basic Vocabulary of Written Business Communication. Unpublished doctor's dissert., Indiana U., 1955.

16. Thorndike, Edward L. and Lorge, Irving. *The Teacher's Word Book of 30,000 Words*. New York: Bur. of Publ., Teachers College, Columbia U., 1944.

17. U.S. Bureau of the Census. Typewriters, Summary for 1966. In *Current Industrial Reports*, Series M35C(66)-13, June 8, 1967.

18. U.S. Department of Labor. Employment Projections by Industry and Occupation, 1960-1975. *Monthly Labor Rev.*, 1963, Vol. 86, 240-248.

19. Wright, Grace S. *Summary of Offerings and Enrollments in High-School Subjects, 1960-61*. (Preliminary Report) U.S. Office of Education, OE-24010, 1964.

20. Wright, Grace S. *Subject Offerings and Enrollments in Public Secondary Schools*. U.S. Office of Education, OE-24015-61, 1965.