

II

THE FORM AND BEHAVIOR OF WORDS

THERE is a decided advantage in beginning our investigation of language by restricting our attention to the form and behavior of words, since the essential phenomena in the dynamics of words are far more clearly apparent and readily apprehended than those of the smaller or larger speech-elements. If it cannot be truthfully averred, as is sometimes felt, that the stream of speech is primarily a stream of words rather than a stream of, say, phonemes or sentences, the word does seem, nevertheless, to occupy a middle terrain between the smaller elements which are its components and the larger phrasal, clausal, and sentence elements of which the word is in turn but a part. In studying the dynamics of words, then, we are studying what represents simultaneously either an aggregate of, or the component of, other speech-elements, and are hence incidentally approaching the dynamic problems of these other speech-elements at their most accessible side.

PART I

THE QUESTION OF FORM

1. THE LENGTH OF WORDS AND THEIR FREQUENCY OF USAGE

Probably the most striking feature of words is difference in length. A word may consist of a single phoneme (e.g. English *a* in *a book*), or it may represent a phoneme-sequence of considerable magnitude (e.g. English *transcendentalism*, *constitutionality*, *quintessentially*). The question naturally arises as to what, if any, is the significance of these observable differences in length.

If there is any connection between the length of a word and its meaning, the nature of the connection is certainly not at once apparent. The same idea

is found adequately conveyed in different languages by words of decidedly different length (e.g. English *trade*, German *Handwerk*; English *work*, German *Arbeit*). Hence, at least for the present, we may disregard considerations of meaning in examining the significance of the factor of length in the form of words.

The question is natural as to the number of different long and short words in the vocabulary of a given person, or of a given dialect. As far as any *a priori* statement is permissible on this subject, it seems deducible only from the law of permutations which clearly gives the presumption in favor of a greater abundance of different longer words than of shorter. For, the possible permutations of a given number of phonemes taken, say, five at a time is far greater than those when only two are taken at a time. If some permutations of phonemes would be too difficult to pronounce with convenience (e.g. *tp* or *tpbgd* in English), unpronounceability is not restricted to long or short words. That is, short permutations may be unpronounceable as well as long permutations. If long permutations offer greater opportunity for unpronounceable arrangements of phonemes than do smaller configurations, they also offer, by the same token, greater opportunity for pronounceable arrangements. Hence, the palpable fact that some combinations of phonemes are impossible to articulate does not in itself invalidate the *a priori* statement which has just been made.

Nevertheless, empirical evidence does preclude the use of this *a priori* statement, for it would be accurately descriptive only if every language availed itself progressively of every possible legitimate short permutation before employing a larger permutation, a condition which is by no means the case. In English, for example, the long permutation *constitutionality* is a meaningful word while the shorter and equally pronounceable permutations *puv*, *za*, *ut* have no meaning. Hence, the one *a priori* statement which it seems possible to make about the length of words may be discarded at the outset.

Interesting light might be shed empirically upon the significance in the length of words if it were possible to make a list of all words in the active-passive vocabulary of an individual or of a speech-community to ascertain the actual number of different words representing each of the different degrees of length. But a list of this type for any single language is impossible. Dictionary lists are generally inadequate because of their inclusion of obsolescent and obsolete words, and because of their exclusion

of highly useful neologisms. And it seems practically impossible ever to make a completely adequate list, even of the vocabulary of an individual person, because so many words exist in the passive vocabulary which are used rarely, if at all, in the stream of speech which is alone perceptible. Then, too, the active-passive vocabulary, whether of individuals or of speech-groups, differs so in size and content and varies so from time to time, that an attempt even to estimate merely the limits of a vocabulary at any time is largely a matter of guesswork.¹ Hence we are obliged at the very beginning of our investigation to restrict our attention exclusively to the objective evidence of the stream of speech itself which, during the entire course of our investigation, will be the sole source of our data.

Now, in so far as data are already available from the stream of speech, it seems reasonably clear that shorter words are distinctly more favored in usage than longer words. That is, however large the stock of short and long words may be, the evidence of language seems to indicate unequivocally that the larger a word is in length, the less likely it is to be used. To illustrate this point, the data gathered by F. W. Kaeding² from samplings of connected written German, totalling 10,910,777 words (or 20 million syllables) in length, are presented. Kaeding selected the syllable as the unit of length, and, in the following tabulation of his results, the left-hand column indicates the magnitude of each class of words as estimated by the number of syllables; the center column presents the number of occurrences (including repetitions) of words of each magnitude; and the column at the right notes the percentage of the occurrences of all the words (including repetitions) of each magnitude to the total number of words (10,910,777).

Number of Syllables in Word	Number of Occurrences (Including Repetitions) of Words	Percentage of the Whole
1	5,426,326	49.76%
2	3,156,448	28.94
3	1,410,494	12.93
4	646,971	5.93
5	187,738	1.72
6	54,436	.50
7	16,993	.22
8	5,038	
9	1,225	
10	461	
11	59	
12	35	
13	8	
14	2	
15	1	
	<hr/> 10,906,235*	

These figures indicate that in German there is a decided preference in usage for short words, and that the magnitude of words stands in an inverse (not necessarily proportionate) relationship to the number of occurrences (including repetitions) of all words possessing that magnitude.

Though these statistics from Kaeding seem clear for all occurrences of all words when arranged in classes according to differences in syllabic magnitude it provides no information about the number of different words in each class. We do not know from the statistics for example whether the five million odd occurrences of words of one syllable represent the single occurrence of that many different words, or that many different occurrences of a single word. While Kaeding gives enough additional information in his entire treatise for us to determine with reasonable accuracy the general tendencies involved, we shall in lieu of this Kaeding material (which may be consulted in the notes ⁴) present the results of three separate investigations, one of Plautine Latin,⁵ one of modern colloquial Chinese ⁶ (Peiping dialect), and one of English,⁷ which are far more precise than the Kaeding material on this subject, and which are corroborated in their main tendencies by the data of the Kaeding material.

The material for the investigation of colloquial Chinese of Peiping consists of twenty different samplings of connected speech, each a thousand

syllables long. The number of different words * occurring in these 20,000 Chinese syllables are arranged in the ensuing table (p. 26) according to the relative frequency of their occurrence. The left-hand column of the Chinese statistics gives the times of occurrences; the center column presents the number of words assignable to a given frequency of occurrence; and the right-hand column, in parentheses, indicates the number of words in each frequency grouped according to the number of syllables (designated by a superior number). Thus, of the 2046 Chinese words of single occurrences, 315 contained only one syllable, 1571 two syllables, etc.

In the investigation of the Latin of Plautus a different procedure was adopted. With all the words of four Plautine plays (*Aulularia*, *Mostellaria*, *Pseudolus*, and *Trinummus*) selected for material, the average number of syllables in each frequency category was computed (p. 27). The respective averages are in parentheses at the right. Thus, the average number of syllables of all words occurring once was 3.23, of those occurring twice, 2.92, etc.

The third investigation was made by R. C. Eldridge of four samples of American newspaper English totalling 43,989 words in length and representing the occurrences of 6002 different words.* The figures in parentheses at the right (p. 28) of the two columns represent the average number of phonemes in each frequency category. For example, the average number of phonemes in all words occurring once in this investigation is 6.656; this figure was derived by dividing the sum total (i.e. 19,809) of all phonemes (estimated according to the phonemic system in use in Cambridge, Massachusetts) of all words occurring once, by the number of words occurring once (i.e. 2976).

Thus in the investigations of the three different languages, three different yet apparently equally valid units of length were employed: the morpheme in Chinese, the average number of syllables in Plautine Latin, and the average number of phonemes in American newspaper English. The difference in the unit of magnitude does not disguise the presence of the prevailing tendency which, as we shall now clearly see, is equally manifest in each of the three languages. From the evidence of these tables it is clear: (1) that the magnitude of words tends, on the whole, to stand in an inverse (not necessarily proportionate) relationship to the number of occurrences; and (2) that the number of different words (i.e. variety) seems to be ever larger as the frequency of occurrence becomes ever smaller. That is, a

statistical relationship has been established between high frequency, small variety, and shortness in length, a relationship which is presumably valid for language in general.

CHINESE OF PEIPING

LATIN OF PLAUTUS

Number of Occurrences	Number of Words	Average Number of Syllables	Number of Occurrences	Number of Words	Average Number of Syllables	
1	5429	(3.23)	31	8	(2.05)	
2	1198	(2.92)	32	3		
3	492	(2.77)	33	4		
4	299	(2.05)	34	6		
5	161	(2.60)	35	3		
6	126	(2.53)	36	5		
7	87	(2.39)	37	7		
8	69	(2.44)	38	2		
9	54	(2.35)	39	4		
10	43	(2.32)	40	3		
11	44	(2.29)	41	3	(1.70)	
12	36	(2.30)	43	4		
13	33	(2.30)	44	1		
14	31	(2.09)	45	1		
15	13	(2.07)	46	1		
16	25	(2.40)	47	3		
17	21	(2.09)	48	1		
18	21	(2.04)	49	1		
19	11	(2.18)	50	2		
20	15	(2.08)	51	2		
21	10		53	4		
22	8		54	1		
23	8		55	1		
24	9		56	2		
25	11		58	1		
26	7		61	3		
27	9		(2.00)	62-514	71	(1.40)
28	12			33,094	8,437	
29	4					
30	4					

LATIN OF PLAUTUS

Number of Occurrences	Number of Words	Average Number of Syllables	Number of Occurrences	Number of Words	Average Number of Syllables
1	5429	(3.23)	31	8	(2.05)
2	1198	(2.92)	32	3	
3	492	(2.77)	33	4	
4	299	(2.05)	34	6	
5	161	(2.60)	35	3	
6	126	(2.53)	36	5	
7	87	(2.39)	37	7	
8	69	(2.44)	38	2	
9	54	(2.35)	39	4	
10	43	(2.32)	40	3	
11	44	(2.29)	41	3	(1.70)
12	36	(2.30)	43	4	
13	33	(2.30)	44	1	
14	31	(2.09)	45	1	
15	13	(2.07)	46	1	
16	25	(2.40)	47	3	
17	21	(2.09)	48	1	
18	21	(2.04)	49	1	
19	11	(2.18)	50	2	
20	15	(2.08)	51	2	
21	10		53	4	
22	8		54	1	
23	8		55	1	
24	9		56	2	
25	11		58	1	
26	7		61	3	
27	9		<u>62-514</u>	<u>71</u>	(1.40)
28	12		33,094	8,437	
29	4				
30	4				

AMERICAN NEWSPAPER ENGLISH
(According to R. C. ELdridge)

Number of Occurrences	Number of Words	Average Number of Phonemes	Number of Occurrences	Number of Words	Average Number of Phonemes
1	2976	(6.656)	31	6	(3.903)
2	1079	(6.151)	32	4	
3	516	(6.015)	33	6	
4	294	(6.081)	34	2	
5	212	(5.589)	35	5	
6	151	(5.768)	36	3	
7	105	(5.333)	37	2	
8	84	(5.654)	39	2	
9	86	(5.174)	40	4	
10	45	(5.377)	41	1	
11	40	(4.825)	42	7	
12	37	(5.459)	43	1	
13	25	(5.560)	44	4	
14	28	(5.00)	45	1	
15	26	(4.807)	46	2	
16	17	(5.058)	47	5	
17	18	(4.166)	48	1	
18	10	(6.100)	49	3	
19	15	(4.733)	50	3	
20	16	(4.687)	51	1	
21	13	(3.455)	52	3	
22	11		54	1	
23	6		55	1	
24	8		56	1	
25	6		58	2	
26	10		60	1	
27	9		61-4290	71	(2.666)
28	6				
29	5				
30	4				

2. THE QUESTION OF A CAUSAL RELATIONSHIP BETWEEN THE LENGTH AND FREQUENCY OF WORDS

The question now arises as to the nature of a possible causal relationship between the length of a word on the one hand, and the relative frequency of its occurrence on the other. If there is a causal relationship between relative frequency and length which accounts for the statistical relationship just discussed, there are only two possible explanations: (1) the length is a cause of the frequency of usage, or (2) the frequency of usage is a cause of the length. That is, for example, the shortness of, say, the most frequent English word, *the*, is either (1) a cause of its high frequency of occurrence, or (2) a result of its high frequency of occurrence.

It seems that on the whole the comparative length or shortness of a word cannot be the cause of its relative frequency of occurrence because a

speaker selects his words not according to their lengths, but solely according to the meanings of the words and the ideas he wishes to convey. Occasionally, of course, out of respect for the youth, inexperience, or low mentality of a particular auditor, a given speaker may seek to avoid long or unusual words. On the other hand, speakers are sometimes found who seem to prefer the longer and more unusual words, even when shorter more usual words are available. Yet in neither case are the preferences for brevity or length followed without respect for the meanings of the words which are selected. Hence there seems no cogent reason for believing that the small magnitude of a word is the cause of its high frequency of usage.

There are, however, copious examples of a decrease in the magnitude of a word which results, as far as one can judge, solely from an increase in the relative frequency of its occurrence, as estimated either from the speech of an individual in which the shortening may occur, or in the language of a minor group, or of the major speech-group. Shortenings of this sort may be termed abbreviations; these are of two types: (1) truncations, (2) substitutions, whether permanent or temporary. A consideration of these two types of abbreviation reveals that they account for practically the entire statistical relationship between magnitude and frequency, and suggest unmistakably that high frequency is the cause of small magnitude.

a. Abbreviatory Acts of Truncation

That truncation occurs primarily with frequent long words, presumably for the purpose of saving time and effort, is a proposition which is too self-evident to require demonstration. When any object, act, relationship, or quality becomes so frequent in the experience of a speech-community that the word that names it develops a high frequency of occurrence in the stream of speech, the word will probably become truncated. A development of this sort is reflected in the histories of the words *movies*, *talkies*, *gas*, which are shortenings of *moving pictures*, *talking pictures*, *gasoline*. The shortenings result from frequent usage, a frequency due to the rapid increase of frequency of movies, talkies, and gas in our daily experience.* Longer words than these, such as *constitutionality*, *quintessentially*, *idiosyncrasy* are not truncated because they are not frequently used.

There are, however, two aspects of truncation which deserve mention at this time: first, the risk of a possible homonymy arising from truncation,

and second, the influence of small speech-groups upon truncations within the larger speech-community. Though the two are essentially unrelated phenomena, yet the influence of the small speech-group in minimizing the risk of homonymy, which may in turn conceivably restrain truncation, justifies their being treated together.

That the truncation of a longer word may result in an abbreviation which is homonymous with another word already in the language is not inconceivable nor will it of necessity lead to a confusion of meaning. One may safely assume that all languages have homonyms, such as English *hole* and *whole*, *hear* and *here*, which are of identical phonetic form but of different usage. The differences in usages seem in most instances sufficient to obviate any serious confusion in meaning.

But though differences in syntactical usage are frequently sufficient to keep separate and distinct two words of like form and different meaning, this is not always the case. The simple statement 'I want some gas,' could in itself signify a desire for illuminating gas, for gasoline, or for 'laughing gas' (nitrous oxide). Another instance is the two homonyms *hypo*, both of which are the results of truncations: *hypo* may be a truncation of a *hypodermic injection* (A below), or it may be an abbreviation of 'hyposulphite of soda' (B below), a name erroneously applied by early photographers to a well-known fixative (hypothiosulphate of soda). Both *hypos* (A and B) are of the same part of speech; in real life a confusion of the two might easily be disastrous. To photographers *hypo* means one thing, to physicians and trained nurses another, to perhaps the majority of English speakers it is without any meaning at all. If we were examining the vocabularies of photographers, physicians, and the general public in respect to *hypo*, we should find at least one salient difference in usage. Photographers use *hypo* (B) in their speech presumably much more frequently than *hypo* (A); physicians use *hypo* (A) more frequently than *hypo* (B); the public uses either rarely if ever. A patient suffering from heavy metal-poisoning would be more likely to receive a 'hypo of hypothiosulphate of soda' than a 'hypodermic injection of hypo' although both amount to the same thing; he certainly would not receive a 'hypo of hypo.'

From the above we may perhaps conclude that a longer word may be truncated if it enjoys a high relative frequency, not only if this high relative frequency obtains throughout the entire speech-community (*movies* for

moving pictures) but if its use (as *hypo A* and *B*) is frequent within any special group inside this large and inclusive speech-community. To the photographer, *hypo* means ‘fixative’ and he very likely calls a *hypodermic injection* by its full name. At a filling station, *gas* means ‘gasoline’; at a plant producing illuminating gas, *gas* means ‘illuminating gas’; in a dental clinic, *gas* means ‘nitrous oxide.’ And the mutually exclusive nature of many groups tend to minimize the danger of confusion which might otherwise arise from homonymy resulting from truncations.

The influence of the special group also doubtless explains the short form of many words of comparatively rare occurrence in the general stream of speech, such as English *volt*, a unit of measure of electricity. Though rare in the general speech it is doubtless of high frequency in the group of electricians and physicists by whom the word *volt* was introduced into the general stream of usage.⁸ To understand its short form we must remember its high frequency in this special group in which the impetus toward brevity took place. Though a speech-community is a unit group in itself, it is also a complex of many different minor social, political, professional, economic, and even geographical groups, in each one of which there are deviations in relative frequency of usage of words from that found in the general vocabulary of the total speech community. Not only do truncations of words occur and persist in these minor groups, as we have seen in *hypo A* and *B*, but a truncated form originating in a minor group may become adopted into the language of the large community. Viewed from the average language of the larger group, the rare word appears to be unjustifiably short; yet viewed from the special group where the word is used, the word has a frequency sufficiently high to justify its shortness. The influence of minor groups, then, as we shall repeatedly observe, must be borne in mind as a possible modifying factor in the behavior of the stream of speech of the general speech-community.

Occasionally homonyms of identical usage but of decidedly different meaning may arise through some process of linguistic change, either in a special group or in the whole speech community. What then? An example of this is the Shakespearean *let* ‘to hinder’ and *let* ‘to allow’ — words of the same part of speech and of almost opposite meaning. Today the verb *let* ‘to hinder’ is obsolete. On the ground of being less frequent and on occasion susceptible to confusion with *let* ‘to allow,’ *let* ‘to hinder’ was presumably

dropped in favor of the synonymous, unambiguous, equally familiar, and incidentally longer *hinder*.

In concluding our brief discussion of the phenomenon of truncation, it may be said that abbreviatory acts of truncation seem to arise on the whole as a consequence of the increased frequency in usage of a word, whether within the entire speech-community or within certain minor groups thereof. The accumulated effects of abbreviatory acts of truncation during the long periods of years in which language has slowly evolved are probably responsible for the shortness of many of the frequently occurring words in speech today, and responsible, as we shall presently see, in many other ways than that which we have just observed.

b. Abbreviatory Acts of Substitution

The substitution of shorter words for longer words, such as *car* for *automobile*, or *it* for *Christmas*, has much the same net effect as truncation on the magnitude of words, and doubtless contributes extensively to the preponderance in usage of short frequent words in the stream of speech. The abbreviatory acts of substitution fall into two types: (1) the more durable substitutions which often involve a change in meaning, and (2) the temporary substitutions which we shall see are largely contextual in nature.

i. Durable Abbreviatory Substitutions

Durable abbreviatory substitutions may occur throughout the entire speech-community (e.g. *car* for *automobile*) or within minor groups within the entire speech-community (e.g. *juice* for *electricity*, *soup* for *nitroglycerine*, *spuds* for *potatoes*). Though one effect of substitutions of this sort may be a more or less permanent renaming (see page 274), we are now chiefly interested in the effect of such substitutions on the frequency-magnitude relationship of words in the stream of speech. If it cannot be directly proved by means of statistics that abbreviatory acts of substitution are the direct result of high relative frequency of occurrence, we can nevertheless apprehend the existence and nature of this causal relationship between high frequency and abbreviatory substitution by viewing typical examples of abbreviatory substitution against the general background of the statistics already presented.

The influence of high frequency upon the more durable substitutions is most readily observable in the substitution of a single word for a complex of words. For example, let us take the two complexes, *sweet potatoes* and *Irish potatoes*, which designate two distinctly different vegetables familiar both in the northern and southern states of the United States. In the northern states, sweet potatoes are called *sweet potatoes*, but Irish potatoes simply *potatoes*; in the southern states the reverse is true. In the South *potatoes*, ‘sweet potatoes,’ has been dialectally and colloquially abbreviated to *taters*; in the North *potatoes*, ‘Irish potatoes,’ has been similarly abbreviated to *spuds*. In the South, the sweet potato is a far more familiar article of diet than the Irish potato, and being more familiar in experience is undoubtedly more frequent in the stream of speech; in the North, the reverse is true. Surely in these two instances where all significant factors are constant except differences in frequency, one cannot but believe that the preponderant frequency in each case has led to the shortening.

In the two transitive phrases, *strike with the chin* and *strike with the foot*, there is no difference in the degree of complexity of verbal arrangement or of clarity of meaning of the concept. Yet is not the greater frequency of the second (*to strike with the foot*) indicated by the very existence of a convenient abbreviatory substitute, *kick*, which is lacking to the first? In the two concepts, *brother* and *uncle’s second wife’s tenth child by her first marriage*, we find the first described by one word, the second by nine. The difference in frequency of these concepts in the normal stream of speech seems alone to account for the differences in length; were the second as frequent in occurrence as the first, we should doubtless possess a single word for it — an abbreviatory substitution caused by high frequency. Though the longer example seems to be an “inherently more complex” concept than that of the single word, yet what we may term “inherent complexity of the concept” does not seem alone capable of preserving the speech-element from abbreviation; few things are more complex in nature than electricity, yet we not only have a single word for the total phenomenon but even a colloquial substitute, *juice*, a substitution presumably made because of the high frequency of usage of the concept without any respect for its complexity whatsoever.

In the minor speech-groups within the large speech-community, abbreviatory substitutions occur more frequently; witness the technical jargon and slang of the various professional, social, political, and

commercial groups.⁹ To the outsider the most striking aspects of the jargon, aside from its picturesqueness, are the shortness of the clique-terms, the frequency of their usage, and the unusualness of the meanings which they convey. But to the insider, the meanings are familiar and the high frequency and short length unnoticed.

It does not, of course, follow that every substitution constitutes a shortening, or that the primary conscious impulse thereto is always one of time-saving. Substitutions may be made for the sake of increased vividness of expression,¹⁰ or of increased articulatedness of meaning. Nevertheless the frequent use of slang and technical terms, words on the whole apparently more convenient than standard language, takes place because it saves time in expression; slang and technical terms save time because these terms represent, by and large, abbreviatory substitutions for frequent concepts which, if fully articulated in standard language, would be excessively long.

The sole point of present concern, however, is not a consideration of the question of change in meaning, which will be treated later (see page 274), nor of the influence of the group upon the speech of the whole community, but rather the fact that many substitutions are shortenings resulting from high frequency. Until it can be shown that lengthenings occur from frequency or shortenings from rarity, we may reasonably presume: (1) that, where frequency and abbreviatory substitution are connected, the frequency is the cause of the abbreviatory substitution; and (2) that the accumulated effect of acts of durable abbreviatory substitution during the evolution of a language is in part reflected by the frequency-magnitude relationship of words today.

ii. *Temporary Abbreviatory Substitutions*

Substitutions of the second type, such as a pronoun for a noun (e.g. *it* for *Christmas*) or a simple adverb for an adverbial phrase, are likewise the result of high frequency. But they differ from the first (more durable) type of substitutions in one salient respect: the more durable substitutions of the first type reflect a general increase in the average relative frequency of a concept within the entire speech-community or within a minor group, while the temporary or transitory substitutions of the second type reflect merely a temporary increase in relative frequency resulting from the topic of

conversation. Thus in the substitution of *car* for *automobile*, there is a high average frequency of occurrence of the concept; but in the substitution of *it* for *Christmas* in the sentence ‘Christmas is a great day, *it* comes but once a year,’ the substitution is the result of only a high temporary frequency which is occasioned by the nature of the context. Similarly with the substitution *there* for *down in Florida* in the sentence ‘They are *down in Florida* because it is so warm *there*.’ If substitutions of the first type are intimately connected with the phenomena of naming, substitutions of the second type will be found closely bound up with questions of syntax and style (see [Chapter V](#)).

Likewise with temporary abbreviatory substitutions one Cannot prove statistically that frequency is the inevitable cause of all substitutions of shorter forms. Nevertheless, our feelings assure us that the substitutions of *it* for *Christmas* and *there* for *down in Florida* in the above typical sentences were made to avoid a too great repetitiveness or frequency of *Christmas* and *down in Florida* within a short period of time. The unusual frequency of occurrence of the concepts precipitated the substitutions which were in fact shorter words. It is unquestionably true that from the point of view of grammar, either the chief or a major function of many of these shorter words, such as pronouns, adverbs, and auxiliaries, is to act as substitutes. But for our present purposes it is sufficient to observe that their use may generally be viewed as abbreviatory substitutions which result from a high though transitory frequency of occurrence of the concepts for which they stand.

3. CONCLUSION: THE LAW OF ABBREVIATION

In view of the evidence of the stream of speech we may say that the length of a word tends to bear an inverse relationship * to its relative frequency; and in view of the influence of high frequency on the shortenings from truncation and from durable and temporary abbreviatory substitution, it seems a plausible deduction that, as the relative frequency of a word increases, it tends to diminish in magnitude. This tendency of a decreasing magnitude to result from an increase in relative frequency, may be tentatively named the Law of Abbreviation.

The law of abbreviation seems to reflect on the one hand an impulse in language toward the maintenance of an equilibrium between length and

frequency, and on the other hand an underlying law of economy as the *causa causans* of this impulse toward equilibrium. That the maintenance of equilibrium is involved is clear from the very nature of the statistics. That economy, or the saving of time and effort, is probably the underlying cause of the maintenance of equilibrium is apparent from the fact that the purpose of all truncations and transitory contextual substitutions is almost admittedly the saving of time and effort. If one cannot argue with complete certainty in favor of economy as the sole cause of the more durable abbreviatory substitutions, one cannot readily advance any other factor as a general precipitating cause, nor escape the inference that the result of durable abbreviatory substitutions is frequently an economy of time and effort, even though this may conceivably not be the purpose. Unquestionably other factors are involved in the general phenomena of abbreviation, some of which will be subsequently discussed in considerable detail as they manifest themselves in the typical behavior of the phoneme, morpheme, and sentence. And from these several angles we shall also observe that the law of abbreviation is by no means restricted in its scope to the length of words.

PART II

THE BEHAVIOR OF WORDS

1. THE FREQUENCY DISTRIBUTION OF WORDS IN THE STREAM OF SPEECH

Manifestations of a tendency toward the maintenance of equilibrium in the behavior of words is not restricted to the relationship between their length and the frequency of their usage; the orderliness of the frequency distribution of the words themselves in the stream of speech suggests an analogous tendency toward the maintenance of equilibrium. But before turning to the evidence in support of this statement, let us momentarily digress in order to define a certain aspect of the term *word*.

a. The Word and the Lexical Unit