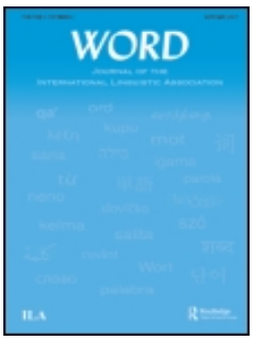


WORD

ISSN: 0043-7956 (Print) 2373-5112 (Online) Journal homepage: http://www.tandfonline.com/loi/rwrd20

# Hesitation Phenomena in Spontaneous English Speech

Howard Maclay & Charles E. Osgood

**To cite this article:** Howard Maclay & Charles E. Osgood (1959) Hesitation Phenomena in Spontaneous English Speech, WORD, 15:1, 19-44, DOI: 10.1080/00437956.1959.11659682

**To link to this article:** https://doi.org/10.1080/00437956.1959.11659682

*HOWARD MACLAY and CHARLES E. OSGOOD*——

# Hesitation Phenomena in Spontaneous English Speech

[This paper reports an exploratory investigation of hesitation phenomena in spontaneously spoken English. Following a brief review of the literature bearing on such phenomena, a quantitative study of filled and unfilled pauses, repeats, and false starts in the speech of some twelve participants in a conference is described. Analysis in terms of both individual differences and linguistic distribution is made, and some psycholinguistic implications are drawn, particularly as to the nature of encoding units and their relative uncertainty. A distinction between non-chance statistical dependencies and all-or-nothing dependencies in linguistic methodology is made.]

## INTRODUCTION

Although references to pause and silence occur frequently in the literature of structural linguistics, very little research of the kind undertaken here has been carried out. A brief examination of three important works will illustrate the customary treatment of pause and hesitation in language description.

In his list of the phonemes of Chicago Standard English, Bloomfield (1933, p. 92)* includes [,], which he describes as being ". . . placed between primary symbols, the pause, often preceded by rising pitch, that promises continuation of the sentence: *John, the older boy, is away at school. . . .*" This is a *pause-pitch* or *suspension-pitch* phoneme which consists of a rise in pitch before a pause within a sentence. A later comment (p. 185) by Bloomfield suggests that the pausal component of this phoneme is a useful, but not a necessary part, of the criterion for its identification.

Harris (1951, p. 174) uses pause in the same general way, as an index of juncture:

> 12.52. Intermittently Present Pause. We may also find in some languages that loose contact and division between breath-groups of phonemes occur sometimes (though not always) at morpheme boundaries, but practically never within a morpheme . . . except as intermittently present features in the utterance. They are free variants,

* All bibliographic references are identified at the end of the article.

and do not occur every time morpheme boundary occurs. But in some occurrences of the morpheme sequence those pauses would constitute observable evidence of morpheme boundary.

In many cases, however, these pauses come at points containing phonemic junctures. At these points in the utterance we find segments which occur only at utterance boundary or at points of intermittently present pause, and which are phonemicized into junctures or into sequences of some phoneme plus juncture. We can then say that the pause (when it occurs) is an occasionally-occurring free variant of the phonemic juncture.

Probably the most explicit and systematic employment of pausal phenomena as an element in linguistic analysis is seen in Bloch's work on Japanese (1946), where he states (p. 201):

1.2. All pauses within a sentence are preceded by level intonation. Every such pause is *faculative:* repeated utterances of the same sentence (by the same speaker or different speakers) will show the pause sometimes present, sometimes absent, without any change in meaning. The presence or absence of pauses within a sentence depends partly on stylistic factors (with more pauses in emphatic or affective speech), partly on the tempo and care of utterance.

However, some facultative pauses are more constant than others, appearing more consistently in repeated instances of the same sentence. It is enough to distinguish two *ranks* of facultative pauses: *higher* (more constant), marked with a semicolon; and *lower* (less constant), marked with a comma.

Bloch also defines a "minimal pause group" which provides one of the criteria for defining words in Japanese. An interesting aspect of Bloch's discussion is the ranking of pauses on the basis of frequency in a given environment and the implication that they are not randomly distributed with respect to linguistic forms.

It seems fair to conclude that pausal phenomena, while they are often mentioned by linguists, function essentially as non-significant events which may serve to identify linguistically relevant units, such as junctures located at the boundaries of phonemes, morphemes, words, phrases, and sentences. The physical character of pause as the term is used by linguists is seldom specified. While it is likely that many of the instances of pause that are referred to in linguistic description overlap with the hesitation types defined below, there is one basic difference between pause and hesitation: hesitations, as we define them, refer to events that are relatively gross and easily observable, while the juncture-pauses of linguistic theory are quite short in duration and much harder to observe and record. In addition, hesitations often interrupt the flow of speech while juncture pauses do not.[1]

[1] This distinction is taken from Lounsbury, in Osgood and Sebeok (1954), p. 98.

Thus, under "hesitation" we also include the phenomena which classical rhetoric called aposiopesis and anacolouthon. To quote Bloomfield again (p. 186):

> Often enough non-linguistic factors interfere with construction; what the speaker has said is nevertheless meaningful, provided he has already uttered a free form. In *aposiopesis* the speaker breaks off or is interrupted: *I though he* ——. In *anacolouthon* he starts over again: *It's high time we—oh, well—I guess it won't matter.* When a speaker hesitates, English and some other languages offer special parenthetic *hesitation-forms*, as [r] or [ɛ] in *Mr.—ah—Sniffen* or *Mr.—what you may call him—Sniffen* or *that—thingamajig—transmitter.*

This statement defines two of the hesitation types described below—*False Start* and *Filled Pause.*

Bloomfield's observation that non-linguistic factors are operating here agrees with our results, and leads us to the psychological side of the question. Although possessing a long tradition of research on verbal behavior, experimental psychologists have usually preferred to rely on normalized written material and have therefore shown little or no interest in the hesitations present in spoken language. Various sorts of speech deviation have, however, been the concern of clinical psychologists, especially psychoanalysts, for many years. Freud has devoted a good deal of attention to the interpretation of speech slips, which include the forgetting of words, the replacement of intended words by others, and transpositions of sounds and words (see, e.g., Freud, 1938). He has offered psychoanalytic explanations of these phenomena, and psychotherapists have often observed that a sudden onset of hesitations, extended silences and speech disturbances in general indicates that an anxiety-provoking area in the patient has been touched.

In a recent series of studies, George Mahl (1956*a*, 1956*b*, 1957)[2] has attempted to specify more exactly the relation between the speech disturbances and level of anxiety in both clinical interviews and stressful role-playing situations. He has established eight "disturbance" categories: (1) "ah"; (2) sentence correction; (3) sentence incompletion; (4) repetition of words; (5) stutter; (6) intruding incoherent sound; (7) tongue slip (neologisms, transpositions, etc.); (8) omission of words or parts of words.[3]

[2] See also G. Mahl (1958).

[3] These equate as follows with our hesitation types: "ah" is equivalent to our Filled Pause; sentence correction equates with our Retraced False Start; sentence incompletion is equivalent to our Non-Retraced False Start; repetition plus stutter corresponds to our Repeat. Categories 6, 7, and 8 have no counterparts in our analysis, and Mahl used nothing strictly equivalent to our Unfilled Pause (although he has studied extended silences).

He has also developed three speech disturbances ratios: *general*, which is the number of disturbances divided by the number of words in an utterance; *ah*, which is the number of instances of category 1 divided by the number of words; and *non-ah*, which is the number of instances of categories 2–8 divided by the total number of words. He finds that the *non-ah* ratio correlates positively with degree of anxiety as judged by a trained observer and the *ah* ratio correlates negatively, whereas the *general* ratio is independent. He presents anecdotal material which indicates that "the vast majority of the speech disturbances are 'unintended' and escape the awareness of both speakers and listeners, in spite of their very frequent occurrence." He also finds "striking individual differences in the frequency of speech disturbances, and in the relative predilections for the individual categories of disturbance." We have preferred to call our measures "hesitations" rather than "disturbances," but the overlap with Mahl's categories is obvious, and our results accord very well with his where direct comparisons are possible.

Another research program bearing closely on our work is that being conducted by Frieda Goldman-Eisler (1954*a*, 1954*b*, 1955, 1957, 1958*a*, 1958*b*). She has studied the rate of speech production, hesitation pauses, speech-breathing and their interrelationships under various conditions, particularly the psychiatric interview. Among her significant observations are the following: (1) Variability in total speech rate proves to be mainly a function of the time spent in hesitation pauses and not in time spent in articulation. (2) Short utterances are much more variable in speech rate than long utterances, a high degree of stability within the individual being found for utterances longer than 100 syllables. (Our samples are mostly 100 *words* or longer, hence well within the range of stable rate.) (3) Interruptions in the flow of speech are occasioned by both breathing pauses and by hesitation pauses; the former seem to vary more with emotionality and the latter more with cognitive processes, e.g., prevocal formulation of utterances, word finding and the like. (4) Most relevant to this study is Goldman-Eisler's report (1958*b*) that hesitation pauses anticipate sudden increases in information or uncertainty in the message being produced (1958*a*, p. 67):

> . . . the close relation found to exist between pauses and information on the one hand and fluency of speech and redundancy on the other, seems to indicate that the interpolation of hesitation pauses in speech is a necessary condition for such an increase. . . . Fluent speech was shown to consist of habitual combinations of words such as were shared by the language community and such as had become more or less automatic. Where a sequence ceased to be a matter of common

conditioning or learning, where a speaker's choice was highly individual and unexpected, on the other hand, speech was hesitant.

Our own findings point to very much the same general conclusion.

Lounsbury has discussed hesitation phenomena along highly similar lines (in Osgood and Sebeok, 1954, pp. 98–101). Three of his hypotheses are worth repeating since our results bear directly on them.

HYPOTHESIS 1: Hesitation pauses correspond to the points of highest statistical uncertainty in the sequencing of units of any given order.

HYPOTHESIS 2: Hesitation pauses and points of high statistical uncertainty correspond to the beginning of units of encoding.

HYPOTHESIS 3: Hesitation pauses and points of high statistical uncertainty frequently do not fall at the points where immediate-constitutent analysis would establish boundaries between higher-order linguistic units or where syntactic junctures or 'facultative pauses' would occur.

Hypotheses 1 and 3 are testable, and we offer some evidence on them in this paper. Hypothesis 2 has an element of circularity, in that no independent method of defining encoding units has been developed. However, the implication of the hypothesis is that it should be possible to identify functional psycholinguistic units of encoding through analysis of hesitation phenomena.

## PROCEDURE

This study is based on analysis of a sample of slightly over 50,000 words taken from a conference held at the University of Illinois in 1955. The selection of the sample was non-random in that it included almost all of the relatively longer utterances in the text, those under 80 words being omitted. The final corpus consists of 163 utterances by 13 male speakers, all professional people, with a mean utterance length of 309 words. The data were tape recorded and transcribed in normal English orthography by secretaries who were instructed to produce as literal a transcription as possible. They were told to avoid any "cleaning up" of the text and asked to preserve everything said by each speaker regardless of their opinion as to its grammatical accuracy. This resulted in what might be called a "semi-literal" text, in that it lies somewhere between a rigorous phonemic transcription and the usual normalized orthographic version of spoken English. This compromise arose primarily from the nature of the study. It seemed wise to sacrifice depth for breadth in an exploratory investigation, with the hope that any significant relations that might have been affected by this procedure could be checked in future studies using a fully adequate linguistic transcription.

Four Hesitation Types were defined.

1. REPEATS (R): All repetitions, of any length, that were judged to be non-significant semantically. In the utterence, *I I saw a very very big boy*, *I* is repeated but *very* is not, since in the latter case the repetition intensifies *big* and thus changes the meaning, while this is not the case for *I*. We assume that the utterance with or without the repeated *I* would be judged the "same" in meaning by English speakers, while utterances with one as against two occurrences of *very* would be judged "different." A REPEAT can vary from a single phoneme to an extended stretch that could, theoretically, be of any length but actually does not exceed four or five words in this corpus.

2. FALSE STARTS (FS): All incomplete or self-interrupted utterances. *I saw a very . . .* is an incomplete utterance with FS following *very*, while *I saw a very big // a very small boy* is a self-interrupted utterance with FS following *big*. The second case represents an instance of RETRACED FS and the first an instance of NON-RETRACED FS. This distinction is made on the basis of whether or not the speaker backed up in an attempt to correct one of the words he had already used. Intonational patterns and word order are the major linguistic cues in the identification of this type.

3. FILLED PAUSES (FP): All occurrences of the English hesitation devices [ɛ, æ, r, ə, m]. Of these alternatives [ə] is by far the most frequent in our data.

4. UNFILLED PAUSES (UP): These were marked when there was judged to be an abnormal hesitation in speech that could not be referred to the three previous categories. UP has two major forms: silence of unusual length and non-phonemic lengthening of phonemes. This is necessarily a matter of judgment on the part of listeners (here, the authors) familiar with the pace and style of a particular speaker. What may be clearly noted as an instance of Unfilled Pause for one speaker would not be so judged for another speaker, say, with a slower rate of delivery.

On the whole, these four categories refer to variations in speech that have not been the explicit concern of linguists, although Unfilled Pauses would certainly include many instances of Bloch's facultative pause. False Starts seem to cover much the same ground as is included in the terms aposiopesis and anacolouthon.

A word of caution should be entered concerning the references to intonation, non-phonemic length and other linguistic correlates of the hesitation types. These are after-the-fact interpretations; the actual scoring involved judgments that were not explicitly based on these considerations. The authors, as native speakers of English, attempted to categorize and locate the hesitations that seemed to be present in the text. Scoring was based on

the tape recording with the written text being used as a guide and a framework within which the scores might be placed. Preliminary preparation of the judges was limited to a general discussion of the types and a one-hour session in which a sample text was jointly scored.

After each author had independently scored a semi-literal transcription, a final text was obtained by combining the two ratings and counting only those hesitations where both judges agreed as to type and location. This means that some hesitations which actually occurred were not included, but it also makes it highly probable that all hesitations upon which the analysis is based really did occur at the points indicated. Very high product-moment correlations ($r$) were found between the independent ratings for utterance totals (R: +94; FS: +98; FP: +99; UP: +88). Specific correlations based on agreements and disagreements at each point presumably would have been lower. These latter, and more accurate, reliability estimates could not be obtained due to the impossibility of stating the number of cases where *neither* judge had indicated a hesitation. This uncertainty led to the conservative decision to consider only instances of complete agreement. In addition to the frequency of each type of hesitation per utterance, the speed in words per minute of each utterance was calculated.

In the following sample taken from the final text, R is indicated by underscoring the items repeated, FS by //, FP by /ª and UP by /.

As far as I know, no one yet has done the // in a way obvious now and interesting problem of / doing a // in a sense a structural frequency study of the alternative / syntactical ///ª in a given language, say, like English, the alternative /ª possible structures, and how // what their hierarchal / probability of occurrence structure is. Now, it seems to me you w-w-will need that kind of data as base line from which to judge or study deviations—in particular style in the clinical situation and so on. If we get this // now in other words, if you find that in the clinic, say in the /ª protocol of a patient, that the distribution of these /ª alternative structures are precisely what they are in ordinary /ª communications, then there's no evidence that this, at least i-is a / relevant variable for the clinical situation. On the other hand, if, knowing that standard sort of thing for English / speakers, /ª then you find you get oscillations about that in different stages of therapy and so on, then it becomes, I think, very very very relevant. For example, a // if under // I was speaking to you, George, the // at supper last night, I think. If /ª you find that /ª at each point, say, in English structure you have an alt- // several alternatives, some of which are highly probable / in ordinary English sentences, syntactical sequences of, for example, /ª adjective, noun, verb /ª vs. a a de-dependent phrase inserted, and so on. But these have / differing probabilities of occurrence in ordinary English. Now, if you find, for example, that in high points of anxiety and stress in the therapy situation that he / shifts into very simple ones—I mean ones that are of a higher // much higher probability, whereas when he's being intellectual, if you will, and / resisting, you tend to get /ª relatively low probability / sequences, you see. (Subject D, Utterance 10.)

## RESULTS

### LINGUISTIC DISTRIBUTION OF HESITATION TYPES

METHOD OF ANALYSIS. We have made a preliminary attempt to state the distribution of the four hesitation types with respect to English word classes as described by Fries (1952).[4]

Fries classifies English words on the basis of their occurrence in relation to other words rather than on semantic grounds. For example, one class of words can be established by first listing the following three sentence contexts or frames:

Frame A: The ____ is/was good.
Frame B: The ____ remembered the ____.
Frame C: The ____ went there.

We then place in one class all the words which can be inserted in one or more of the blanks and still form a grammatical English sentence. The following sentences illustrate this procedure: The *coffee* was good. The *man* was good. The *clerk* remembered the *tax*. The *husband* remembered the *food*. The *team* went there. The *dog* went there. We would therefore conclude that *coffee*, *man*, *clerk*, *tax*, *husband*, *food*, *team* and *dog* are members of the same class, along with a great many other English words, of course, that could be inserted in these frames. A word such as *the*, however, cannot be placed in these frames and thus must be placed in a different class established by other frames. Each English word class is similarly defined in terms of occurrence within a set of frames.

Obviously, the class described has much in common with the traditional class of nouns defined semantically. For those who may be unfamiliar with Fries' system, we present below the nearest traditional equivalent of his classes. It cannot be too strongly emphasized that the overlap is far from complete, and our analysis depends entirely on the description by Fries and not on the traditional classification. We have also followed Fries in making a distinction between the four major classes (which we call "lexical") and the "function" word classes.

| | *Fries' Classes* | *Traditional Classes* |
|---|---|---|
| Lexical Words | 1 | Nouns |
| | 2 | Verbs |
| | 3 | Adjectives |
| | 4 | Adverbs |
| Function Words | A | Articles, Possessive Pronouns, Numbers, etc. |
| | B | Verb Auxiliaries |
| | C | "not" |
| | D | Adjective and Adverb Modifiers |
| | E | Connectives |
| | F | Prepositions |
| | J | Words uniting phrases in a single sentence. |

[4] The summary of Fries' system which follows is for the benefit of readers who have not been trained in linguistics.

In some of the tables which follow, Fries' Class 1 has been arbitrarily split into 1p, which contains subject personal pronouns (*I*, *you*, *he*, *she*, *it*, *we*, *they*), and 1, which includes the remaining members of this class. In addition, the Function Word classes G, H, K, L, M, N, and O are grouped in one category entitled "other function words" in Table 1.

If we are to describe the distribution of hesitation types we must first locate them relative to the sequences of words in our corpus. The overwhelming majority of Filled and Unfilled Pauses are located at word boundaries and their occurrences can therefore be described as falling generally between words of any class.

"... you tend to get /ª relatively low probability / sequences ..." (Subject D, Utterance 10.)

The Filled Pause (/ª) here is located between a class 2 word, *get*, and a class D word, *relatively*, while the Unfilled Pause (/) falls between two class 1 words, *probability* and *sequences*.

Repeats cannot reasonably be located between words, and we have therefore associated each Repeat with the class of the words or parts of words repeated.

"... a measure of the / of the structure of English, the probabil-probabilistic structure of /ª English language ..." (Subject D, Utterance 4.)

The first Repeat involves two words, *of* (class F) and *the* (class A), while the second is associated with the class 3 word *probabilistic*. In the scoring of this part of the utterance, classes F, A, and 3 would each be given one association with the Repeat hesitation type.

The most convenient way of describing the distribution of False Starts is in terms of the class of the word immediately preceding the break. We adopt this convention for purposes of comparison with the other types. Most False Starts do occur at word boundaries, but many occur at morpheme boundaries within words and some even within morphemes, as the following examples illustrate:

*False Start* (//) at word boundary:

"... And the next one would // might be that the the receiver of the action becomes initiator ..." (Subject 1, Utterance 3.)

*False Start* at morpheme boundary within word:

"... whenever anyone mentions communism-capital // capitalism, communism comes up ..." (Subject A, Utterance 3.)

*False Start* within morpheme:

"... if you were to select your samples differently and to /ª to consider the conte- // the larger context from which you select this 250 word sample ..." (Subject B, Utterance 7.)

In the last example we concluded that the False Start follows the class 1 word, *context*. We determine that the phonetic sequence [kánte?] is class 1 by the presence of the class A word *the* preceding it and the use of the full word *context* in the retracing which follows the False Start. A clear decision as to the class of the words preceding False Starts is often not possible when the break comes early in the word and no retracing follows, for example:

". . . We disregard the structure; we simply take /ª as our unit /ª w- // more or less the whole interview . . ." (Subject E, Utterance 5.)

On the basis of our semi-literal transcription we might guess that the speaker intended *whole* before the False Start, but an examination of the tape shows that the sound preceding the break was the voiced semivowel /w/ rather than the phoneme /h/ which begins *whole*. In cases of this kind no score was recorded.

A second type of ambiguity arises when the phonemic shape of the word preceding the False Start can be recognized but its class affiliation is uncertain due to a lack of following context, e.g.,

". . . I think I have // this is a tool . . ." (Subject 7, Utterance 1.)

The *have* could be either class B (verb auxiliary) or class 2 (verb). In this case one score was given to each of the possible classifications.

FALSE STARTS AND REPEATS. The hesitation types differed distributionally in several respects. Table 1 compares the total distributions of False Starts

TABLE 1. DISTRIBUTION OF FALSE STARTS AND REPEATS BY WORD CLASSES

| Fries Word Class | 1p | 1 | 2 | 3 | 4 | A | B | C | D | E | F | J | Other Function Words | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FS | 53 | 200 | 150 | 73 | 42 | 121 | 57 | 10 | 15 | 7 | 54 | 18 | 17 | 817 |
| R | 123 | 67 | 87 | 37 | 6 | 247 | 66 | 7 | 9 | 38 | 172 | 67 | 46 | 972 |

Probability that difference in distribution of FS and R is due to chance by $x^2$: < .001.

and Repeats with respect to 11 of Fries' classes. The distributions of the two hesitation types are significantly different from each other by Chi-square test over the 13 divisions of Table 1 at a level beyond .001. Repeats typically involve class 1p words (subject personal pronouns), class A (articles, possessive pronouns, numbers), class F (prepositions), and class J (words uniting phrases); False Starts typically involve class 1 words (nouns), class 2 (verbs), class 3 (adjectives), and class 4 (adverbs). Table 2

TABLE 2. DISTRIBUTION OF FALSE STARTS (FS) AND REPEATS (R) BY LEXICAL VS. FUNCTION WORDS

| | FS | R |
|---|---|---|
| Lexical Words | 518 | 320 |
| Function Words | 299 | 652 |
| *Total* | 817 | 972 |

Probability that differences are due to chance (chi-square test):
FS vs. R $<.001$
FS vs. Chance $<.01$
R vs. Chance $<.001$

presents a direct test of the hypothesis that Repeats and False Starts are reciprocally distributed with respect to lexical vs. function words. The Chi-square value for this test is significant at the .001 level. Using a chance estimate based on the fact that 42% of a random sample of our corpus were function words, we find that False Starts tend to occur with lexical words (.01 level) and Repeats with function words (.001 level) more often than would be expected by chance. One aspect of these distributions warrants special comment: subject personal pronouns (1p, classified by Fries with lexical nouns) appear to function with respect to hesitation phenomena like function words rather than lexical items.

The instances of Retraced False Starts offer some interesting data on speech behavior. Every retracing that involved a correction of the word immediately preceding the break was examined. The word corrected was classified by Fries' system, with only those cases where the class determination was clear being included. The question is this: will the amount of retracing vary with the class of the corrected word, so as to include in the retrace function words just antecedent to corrected lexical items? A two-way distinction was established: retracings involving only the corrected word (I) and retracings involving words preceding the corrected word which were not themselves corrected (II). Given the sentence, *I never saw such a man // woman*, the False Start occurs after the class 1 word, *man*, and the retracing includes only the word corrected (case I). In the sentence, *I never saw such a man // a woman*, the False Start occurs in the same place, but the retracing includes the function word, *a*, which is repeated without correction (case II). Table 3 presents the relevant data. It is clear that there is a statistically significant tendency (.001 level) for retraced corrections of lexical items to include antecedent words (almost always function words), whereas retraced corrections of function words seldom include antecedent items. Two exceptions may be noted: again, subject personal pronouns operate more like function words; and class B words (verb auxiliaries) operate more like lexical items. If we assume that speakers will retrace to a

TABLE 3. FALSE STARTS: AMOUNT OF RETRACING BY WORD CLASSES

| Fries' Word Class | I Retracing Includes Only Corrected Words | II Retracing Includes Preceding Words | $\frac{I}{I+II}$ |
|---|---|---|---|
| 1p | 8 | 4 | .667 |
| 1 | 10 | 39 | .204 |
| 2 | 15 | 34 | .306 |
| 3 | 2 | 12 | .143 |
| 4 | 0 | 4 | .000 |
| *Lexical Totals* | *35* | *93* | *.273* |
| A | 48 | 7 | .873 |
| B | 2 | 13 | .133 |
| D | 5 | 0 | 1.000 |
| F | 28 | 1 | .966 |
| *Function-Word Totals* | *83* | *48* | *.634* |

Probability that lexical vs. function-word difference is due to chance ($x^2$): $< .001$.

boundary of an "encoding unit,"[5] these results suggest one kind of a unit as containing a function word followed by a lexical word.

REPEATS. These tend to involve function words rather than lexical items, as has been shown. We may also ask what classes of words they just precede: do speakers tend to repeat items just antecedent to function words or just antecedent to lexical items? Combining all repeats for speakers A through E, we find that 74% of them just precede a lexical item; this proportion is significantly different from chance, based on the relative frequency of lexical items (58%). It thus appears that Repeats tend to be distributed in a way similar to that of pauses (see below). A final question we ask about Repeats is this: How much of the utterance do they typically include? Again combining the data from our major speakers (A through E), we find that 7% of all repetitions include only a single phoneme, 4% a single syllable, 1% a single morpheme (which is not a word), 71% a single word, and 17% two or more words. It is clear that Repeats characteristically involve the single word, occasionally several words, but only rarely units smaller than the word. This, again, seems to have implications for the nature of encoding units.

HESITATION PAUSES. The first question with which we are concerned is whether or not hesitation pauses of both types are distributed at random in spontaneous speech and, if not, what are the determinants of their distribution. A second question concerns the relation between Filled Pause and Unfilled Pause: are they essentially in free variation, or are their distributions complementary to each other? Rather than attempting to collect

[5] On the problem of "encoding units" see pp. 40f. below; see also Osgood and Sebeok (1954), pp. 50–73, and Sarah C. Gudschinsky, "Native Reactions to Tones and Words in Mazatec," *Word* XIV (1958), 338–345.

and analyze all of the pauses occurring in this large corpus, a sample of sixteen word sequences or frames that occurred relatively frequently in the data was selected for detailed analysis. These sequences would usually be defined as phrases in a linguistic description. Table 4 lists the sequences chosen, according to Fries' classification of the word classes included, and

TABLE 4. DISTRIBUTION OF FILLED (FP) AND UNFILLED PAUSES (UP) OVER SELECTED SEQUENCES

| Sequence Type* | Example | | Positions | | | | | Totals |
|---|---|---|---|---|---|---|---|---|
| | | | *1* | *2* | *3* | *4* | *5* | |
| A1 | */*the*2*house*3* | FP | 98 | 61 | 55 | | | 214 |
| | | UP | 50 | 74 | 61 | | | 185 |
| | | total | 148 | 135 | 116 | | | |
| F1 | */*at*2*home*3* | FP | 37 | 80 | 112 | | | 229 |
| | | UP | 44 | 64 | 42 | | | 150 |
| | | total | 81 | 144 | 154 | | | |
| B2 | */*will*2*go*3* | FP | 42 | 52 | 50 | | | 144 |
| | | UP | 18 | 73 | 46 | | | 137 |
| | | total | 60 | 125 | 96 | | | |
| A11 | */*the*2*manor*3* houses*4* | FP | 7 | 6 | 5 | 17 | | 35 |
| | | UP | 5 | 13 | 7 | 4 | | 29 |
| | | total | 12 | 19 | 12 | 21 | | |
| AA1 | */*the*2*three*3* houses*4* | FP | 7 | 3 | 3 | 5 | | 18 |
| | | UP | 4 | 3 | 5 | 3 | | 15 |
| | | total | 11 | 6 | 8 | 8 | | |
| A31 | */*the*2*red*3* houses*4* | FP | 24 | 24 | 19 | 35 | | 102 |
| | | UP | 17 | 36 | 35 | 19 | | 107 |
| | | total | 41 | 60 | 54 | 54 | | |
| FA1 | */*to*2*these*3* houses*4* | FP | 40 | 32 | 35 | 57 | | 164 |
| | | UP | 50 | 34 | 66 | 33 | | 183 |
| | | total | 90 | 66 | 101 | 90 | | |
| F11 | */*of*2*manor*3* houses*4* | FP | 3 | 14 | 2 | 16 | | 35 |
| | | UP | 5 | 6 | 4 | 1 | | 16 |
| | | total | 8 | 20 | 6 | 17 | | |
| F21 | */*of*2*going*3* home | FP | 3 | 7 | 1 | 2 | | 13 |
| | | UP | 3 | 7 | 5 | 2 | | 17 |
| | | total | 6 | 14 | 6 | 4 | | |
| F31 | */*across*2*wide*3* streets*4* | FP | 11 | 35 | 12 | 22 | | 80 |
| | | UP | 7 | 18 | 13 | 6 | | 44 |
| | | total | 18 | 53 | 25 | 28 | | |
| BB2 | */*may*2*have*3* gone*4* | FP | 7 | 1 | 9 | 10 | | 27 |
| | | UP | 2 | 1 | 14 | 9 | | 26 |
| | | total | 9 | 2 | 23 | 19 | | |

* In terms of Fries' classes.

TABLE 4, CONCLUDED

| Sequence Type | Example | | Positions 1 | 2 | 3 | 4 | 5 | Totals |
|---|---|---|---|---|---|---|---|---|
| A331 | /the2big3fat4 man5 | FP | 2 | 3 | 3 | 1 | 4 | 13 |
| | | UP | 1 | 3 | 4 | 7 | 5 | 20 |
| | | total | 3 | 6 | 7 | 8 | 9 | |
| AD31 | /one2very3big4 man5 | FP | 5 | 5 | 0 | 4 | 2 | 16 |
| | | UP | 1 | 5 | 1 | 7 | 0 | 14 |
| | | total | 6 | 10 | 1 | 11 | 2 | |
| FA11 | /under2the3 manor4house5 | FP | 4 | 3 | 4 | 4 | 7 | 22 |
| | | UP | 3 | 4 | 5 | 3 | 2 | 17 |
| | | total | 7 | 7 | 9 | 7 | 9 | |
| FA31 | /under2the3 big4house5 | FP | 14 | 8 | 13 | 11 | 16 | 62 |
| | | UP | 14 | 16 | 32 | 17 | 7 | 86 |
| | | total | 28 | 24 | 45 | 28 | 23 | |
| F331 | /under2heavy3 white4sheets5 | FP | 1 | 5 | 2 | 3 | 2 | 13 |
| | | UP | 3 | 2 | 2 | 2 | 3 | 12 |
| | | total | 4 | 7 | 4 | 5 | 5 | |

presents the frequencies of Filled Pauses (FP), Unfilled Pauses (UP), and their totals occurring in each of the possible inter-word positions, both at the boundaries of these phrases and within them. Illustrative words are given for each sequence. These data are summed over speakers; individual differences in the distribution of pauses have not been analyzed, but casual inspection does not indicate that they are of any magnitude.

Lounsbury's first hypothesis was that hesitation pauses will tend to occur at points of highest uncertainty in spontaneously produced utterances. Since, as Fries notes, there are many more members (alternatives) in his lexical classes than in his function-word classes, we should expect pauses of both types to occur more frequently before lexical words than before function words. Table 5 provides a test of this hypothesis. According to chance, it is assumed, pauses would occur equally in each possible position; actually, both Filled Pauses and Unfilled Pauses are found to occur

TABLE 5. DISTRIBUTION OF FILLED AND UNFILLED PAUSE BY WORD CLASS FOLLOWING PAUSE

| | Filled Pause | Unfilled Pause |
|---|---|---|
| Lexical Words | 418 | 525 |
| Function Words | 360 | 290 |
| *Total* | 778 | 815 |

Probability that differences are due to chance (chi-square tests):
Filled Pause vs. Unfilled Pause < .001
Filled Pause vs. Chance < .02
Unfilled Pause vs. Chance < .001

significantly more frequently before lexical words than before function words (.02 and .001 level, respectively, by Chi-square test). Lounsbury's third hypothesis was that hesitation pauses would often occur at points within phrases where immediate-constituent analysis would not establish syntactical junctures. Inspection of the data in Table 4 shows this to be clearly the case. If we assume that syntactical junctures do not occur within a phrase, then it is obvious that many pauses of both types do not function syntactically. As a matter of fact, approximately half (47%) of all pauses occur within the phrases rather than at their boundaries. Table 6 provides further evidence for the non-chance distribution of hesitation pauses—in this case, for constructions of specific types for which sufficient data for testing were available.

Are Filled and Unfilled Pause in free variation or in complementary distribution with respect to each other? The answer is, clearly, neither. Although both types are about equally likely to occur in any of the phrase sequences studied (*Totals* column, Table 4), they are not equally likely to occur in any of the positions within phrases. Table 5 demonstrates that Filled Pause is relatively more likely to appear before function words, while Unfilled Pause is relatively more likely to appear before lexical words, this difference in distribution being significant at the .001 level by Chi-square test. Table 6 indicates that in particular sequence types,

TABLE 6. CHI-SQUARE TESTS ON DISTRIBUTION OF FILLED AND UNFILLED PAUSES OVER EIGHT SELECTED SEQUENCES

| Sequence* | Filled Pause/Chance | Unfilled Pause/Chance | Filled/Unfilled |
|---|---|---|---|
| A1 | .001* | .001 | .10 |
| F1 | .001 | .001 | .10 |
| B2 | .01 | .70 | .001 |
| A11 | .10 | .02 | .10 |
| A31 | .01 | .20 | .01 |
| FA1 | .01 | .05 | .01 |
| F31 | .20 | .01 | .05 |
| FA31 | .05 | .50 | .001 |

* In terms of Fries' classes.
Each cell indicates probability that difference is due to chance.

where sufficient data were available for analysis, distributions of Filled and Unfilled Pauses may be significantly different. In phrases of the type *will go* (Fries' B2), we find FP before *will* and UP before *go*; in phrases of the type *the red houses* (Fries' A31), we find FP at the external boundaries and UP within the phrase; in phrases like *to these houses* (Fries' FA1), FP again tends to occur at the external boundaries and UP most often before the lexical items; in F31 phrases like *across wide streets*, the same trend is noted, as is also the case for FA31 phrases like *under the big house*. In other

words, for those constructions that can be analyzed statistically, Filled Pauses occur more frequently at phrase boundaries and Unfilled Pauses at word boundaries within phrases. It should be noted that these are statistically significant tendencies, not cases of absolute complementary distribution in the linguistic sense.

### Individual Differences in Hesitation Phenomena

Differences between speakers can be described in terms of both variation between individuals on particular measures and variation within individuals across all measures (e.g., profiles). Table 7 summarizes the data

TABLE 7. SPEED OF SPEAKING (WORDS/MINUTE) AND RATES PER 100 WORDS OF HESITATION TYPES

| Speaker* | N of Utterances | Mean Words per Minutes | Mean Rates/100 Words R | FS | FP | UP | Totals (R, FS, FP, UP) |
|---|---|---|---|---|---|---|---|
| 3 | 4 | 181 | .19 | .47 | 1.88 | 2.53 | 5.35 |
| D | 38 | 177 | 1.39 | 2.12 | 2.48 | 3.58 | 10.20 |
| 8 | 10 | 173 | 1.55 | 1.16 | 3.30 | 1.84 | 8.77 |
| 7 | 8 | 163 | 2.17 | 1.65 | 3.10 | 2.01 | 9.34 |
| 4 | 5 | 162 | .60 | .94 | 1.54 | 2.65 | 6.16 |
| 6 | 3 | 153 | 1.70 | 2.13 | 2.77 | 5.74 | 13.19 |
| A | 16 | 149 | 1.79 | 1.91 | 3.80 | 3.40 | 11.64 |
| 2 | 8 | 144 | 2.50 | 1.85 | 2.96 | 5.27 | 13.22 |
| B | 22 | 143 | 2.21 | 1.28 | 6.41 | 1.80 | 12.26 |
| 1 | 7 | 140 | 3.26 | 1.34 | 5.35 | 2.91 | 13.61 |
| E | 15 | 136 | 1.38 | 1.90 | 7.21 | 2.68 | 13.60 |
| C | 21 | 134 | 1.45 | 1.30 | 3.06 | 4.51 | 10.78 |
| 5 | 6 | 122 | 1.68 | 1.18 | 6.41 | 4.44 | 14.50 |
| Grand Means | | 152 | 1.68 | 1.48 | 3.87 | 3.34 | 10.97 |

R—Repetitions; FS—False Starts; FP—Filled Pauses; UP—Unfilled Pauses.

* Speakers having 15 or more utterances of sufficient length are designed by letters, the others by numbers.

used for these comparisons; speakers are ordered according to their mean rates of speaking in words per minute. The numbers of utterances on which the mean speeds and hesitation types are based varied considerably, as can be seen in the table. We shall limit our discussion mainly to those speakers (A through E) who contributed 15 or more utterances of sufficient length for analysis. It may be noted that for the group as a whole, with a few minor exceptions, Filled and Unfilled Pauses occur more frequently as hesitation types than do Repeats or False Starts.

Significant differences between speakers appear for all of the measures made. SPEED OF TALKING: Only one of the 21 utterances by the slowest speaker (C) is as fast as the fastest speaker's (D) median; reciprocally,

only one of the 38 utterances by the fastest speaker is as slow as the median rate for the slowest speaker. REPETITIONS: Only three of the 15 utterances by the least repetitious speaker (E) exceed the median of the most repetitious speaker (B), and reciprocally two of the 22 utterances by the most repetitious speaker are below the median for the other. FALSE STARTS: Only two of B's utterances have as many false starts as D's median; eight of D's 38 utterances are below B's median, however. FILLED PAUSES: None of D's 38 utterances have as many filled pauses as E's median level; reciprocally, none of E's 15 utterances have as few filled pauses as D's median. UNFILLED PAUSES: None of B's 22 utterances have as many unfilled pauses as C's median rate and, reciprocally, only one of C's 21 utterances shows as few unfilled pauses as B's median. These median tests indicate that, for the five speakers on whom we have sufficient data, differences on these measures are highly characteristic.

Table 8 gives the correlations across all subjects for these measures. The low and non-significant correlations between the hesitation types indicate

TABLE 8. PRODUCT-MOMENT CORRELATIONS OF VARIABLES ACROSS SUBJECTS†

| | R | FS | FP | UP | Total Hesitation |
|---|---|---|---|---|---|
| Speed | −.121 | −.078 | −.712** | −.385 | −.795** |
| R | | +.027 | +.112 | +.032 | +.181 |
| FS | | | +.032 | +.111 | +.148 |
| FP | | | | −.157 | +.720** |
| UP | | | | | +.478* |

* = .05 > p > .01 ** = .p < .01

† For meaning of abbreviations, see Table 7.

that Repetition, False Start, Filled Pause, and Unfilled Pause are essentially independent variables across subjects. The higher correlations of total hesitations with the two types of pauses than with the other phenomena merely reflects the greater frequencies of the former. The fact that there is a very significant negative correlation between speed of talking and total hesitations might be explained away as simply reflecting the fact that speakers who pause more often must produce fewer words per minute. This proves not to be the case, however—at least not within the range of variation in our speakers. For when we recompute words per minute, counting all hesitations as words (i.e., "ah"-pause equaling a word, etc.), the correlation between original and recomputed speeds of speaking is +.98. Therefore we conclude that fast speakers tend to be "better" speakers in the sense that they make significantly fewer hesitation errors of the types measured here.

It is always possible that correlations based on a group of subjects may wash out important relations within subjects taken individually. As a check on this possibility, rank correlations (*rho*) were computed for the five subjects on whom we had enough utterances to make this procedure

TABLE 9. RANK CORRELATIONS OF VARIABLES WITHIN SUBJECTS*

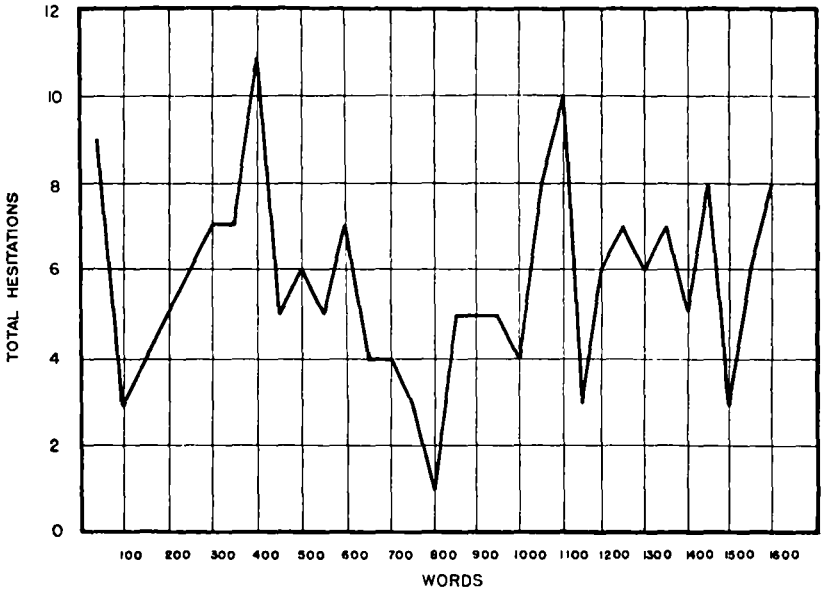
| Variables | A | B | C | D | E |
|---|---|---|---|---|---|
| Speed/R | +.13 | −.11 | +.07 | +.11 | −.09 |
| Speed/FS | +.06 | +.39 | +.08 | +.03 | −.01 |
| Speed/FP | −.70 | −.59 | −.24 | −.63 | −.54 |
| Speed/UP | −.46 | +.17 | −.48 | −.65 | −.52 |
| R/FS | +.36 | +.11 | −.13 | +.09 | +.34 |
| R/FP | +.22 | +.12 | +.07 | −.09 | −.28 |
| R/UP | −.19 | .00 | −.17 | −.17 | +.02 |
| FS/FP | −.12 | +.08 | −.19 | −.21 | +.21 |
| FS/UP | −.40 | +.15 | −.08 | −.01 | −.56 |
| FP/UP | +.23 | −.24 | +.09 | −.23 | −.03 |
| Speed/Total | −.56 | −.39 | −.40 | −.55 | −.74 |

* For meaning of abbreviations, see Table 7.

meaningful. Inspection of Table 9 shows that these variables are related within individuals pretty much as they are across subject averages. The correlations among the hesitation types (R/FS, R/FP, R/UP, FS/FP, FS/UP, and FP/UP) tend to be small in magnitude, different in sign, and averaging toward zero, whereas correlations between speed and pauses of both types (and hence also total errors) are consistently negative as before.

Within these generally consistent relations across speakers there may be uniquely individual tendencies, however. Subjects A and E, for example, show negative correlations of considerable magnitude between False Starts and Unfilled Pauses, as if a sufficient amount of pausing eliminated the conditions for making false starts. Subject B shows a definite tendency to increase False Starts and Unfilled Pauses as his speed of talking increases in utterances, this being in contrast to the other speakers. Furthermore, as Mahl has shown for psychiatric interviews, there may be considerable variation within an extended utterance for a single speaker. Figure 1 plots total hesitations in successive 50-word units for a 1600-word utterance by speaker E; variations from as low as 1/50 to as high as 11/50 are shown, and they seem to follow a non-random course. However, we have not in this study attempted to relate such variations either to the content of utterances or to the states of speakers.

It can also be shown that the relative "preference" for hesitation phenomena of different types varies consistently between individuals. Figure 2 compares profiles for speakers B and D. B's utterances are characterized by a relatively large number of filled pauses and repeats (and




SUBJECT E, UTTERANCE 2

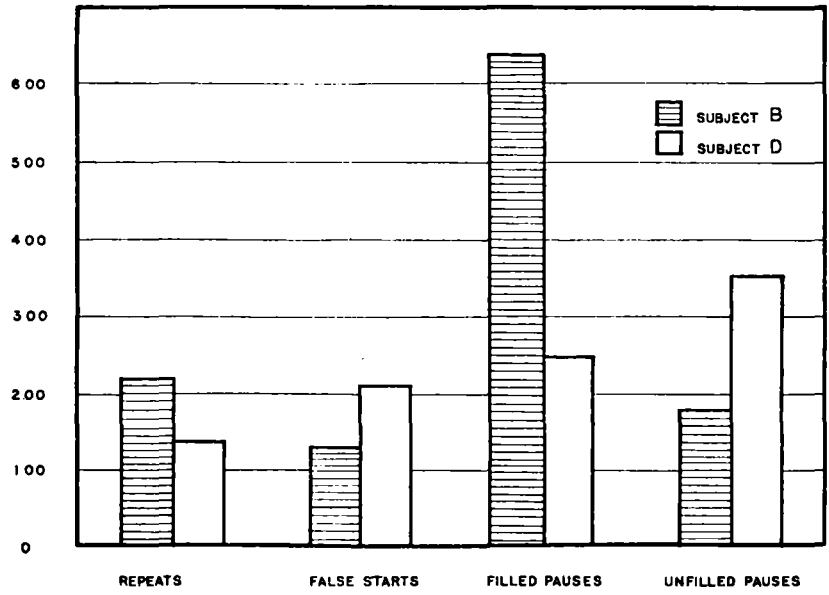
FIGURE 1

a rather slow pace), whereas D's utterances show comparatively more unfilled pauses and false starts (and a rapid pace). Table 10 presents coefficients of concordance (Kendall's *w*) for the rankings of hesitation types by the five major speakers. This statistic, which ranges from zero to 1.00, measures the extent to which the ranking of the hesitation types over his various utterances is consistent within the individual speaker. *W* is significant beyond the .01 level in every case, indicating that individuals—however they may order the types—tend to do it the same way for all of their utterances. Table 10 (like Figure 2) also shows that different speakers

TABLE 10. COEFFICIENTS OF CONCORDANCE (*w*) FROM RANKS ASSIGNED TO HESITATION TYPES

| | | Probability that *w* is due to chance | Average ranking of Hesitation Types* | | | |
|---|---|---|---|---|---|---|
| Subject | *w* | | 1 | 2 | 3 | 4 |
| A | .33 | <.01 | FP | UP | FS | R |
| B | .73 | <.01 | FP | R | UP | FS |
| C | .64 | <.01 | UP | FP | R | FS |
| D | .40 | <.01 | UP | FP | FS | R |
| E | .60 | <.01 | FP | UP | FS | R |

* For meaning of abbreviations, see Table 7.




MEAN RATES PER 100 WORDS OF HESITATION TYPES
(SUBJECTS B & D)

FIGURE 2

rank these hesitation types in different ways. In other words, the relative "preference" for hesitation phenomena of different types may be considered an aspect of individual *style* in speaking.

## Discussion

### Interpretive Summary of Findings

The spontaneous utterances analyzed in this study display consistent differences between speakers, in both the absolute frequencies of False Starts, Repeats, Filled Pauses, and Unfilled Pauses, and in the relative "preference" for these types of hesitation phenomena. However, types of hesitation phenomena are uncorrelated (i.e., independent) both across speakers and across utterances for the same speaker. The negative relations obtained between speed of talking and hesitation phenomena, even when these phenomena are treated as words in counting words per minute, suggest that there are overall differences between individuals in ability to speak spontaneously. The consistency of individual differences demon-

strates that hesitation phenomena may be studied as an attribute of individual style in spontaneous encoding.

Analysis of the linguistic distribution of these hesitation phenomena shows that whereas False Starts typically involve lexical items (blocking after a lexical choice and returning to correct it), Repeats typically involve function words and occur antecedent to lexical items. Repeats thus tend to occur in the same locations as pauses and presumably serve the same function—providing time for selection among diverse lexical alternatives. Repeats are most frequently (71%) of a single word or several words, but rarely (only 12% of cases) of smaller units than the word. When a False Start is retraced, the retracing usually includes the function word or words immediately prior to and associated with that lexical item. Although both Filled Pauses and Unfilled Pauses occur more frequently before lexical than before function words—testifying to the greater uncertainty at these points—and occur within phrases as often as at their boundaries, these two types of pauses are not free variants with respect to positions of occurence. Filled Pauses occur relatively more frequently before function words and at phrase boundaries; Unfilled Pauses, on the other hand, occur more frequently before lexical words and within syntactic phrases. However, these two types of pauses are not in complementary distribution linguistically, since either type *can* occur in any position where the other occurs and does so frequently.

## Linguistic and Psycholinguistic Criteria of Structure

Linguists have typically regarded the data of their science as being confined to those aspects of vocal activity that can be precisely described with a finite number of discrete categories. (See Joos, 1950, for a discussion of this point.) According to this view hesitations fall in a non-linguistic (phonetic) area. We may add, however, that they are "outside" language in a way that other phonetic data are not. These latter, while they may be regarded as continuous phenomena, are systematically related to language structure, since continuous variations in phonetic quality may be translated into discrete phonemic categories by means of linguistic discovery methods. Hesitations are not pre-linguistic in this sense; they function as auxiliary events which help to identify and circumscribe linguistic units, rather than as part of the raw data for which a structural statement must account. The fact that they serve this function shows a recognition of their non-random relation to linguistic forms. We may ask whether a structural statement should account for statistically relevant hesitations as it now does for continuously varying phonetic qualities. A definition of hesitations as part of the raw data of a linguistic description would focus attention on

a number of problems concerning the role of statistical methods in a field that has preferred to rely on a discrete 'either–or' approach.

If we define *structure* as a list of elements and statements about the permissible relations among them it may be that the most acceptable compromise will involve reliance upon discrete assumptions to define units (this would exclude hesitations from the list of elements) and continuous statistical assumptions to state their rules of combination and distribution. It is also possible that statistical methods and criteria could be extended to the identification of phonemes and morphemes and this procedure would include hesitations as linguistic units. Yet another resolution would be to consider as linguistically determined only phenomena that are obligatory in a language code, all other non-randomness being considered to be psycholinguistically determined. Linguistic determinants will inevitably control a substantial part of the message—phonetic content, ordering of phonemes and morphemes, and sequences of word classes and phrase types are largely dictated by the rules of grammar. On the other hand, the selection of alternative constructions and of particular lexical items must be explained primarily on non-linguistic grounds. Hesitation phenomena are clearly related to the dynamics of grammatical and lexical selections. Such a classification does not imply that the formal properties of language are independent of psychological processes in language users; to the contrary, obligatory features represent the extreme case of cultural control in behavior.

## Implications for the Nature of Psycholinguistic Units

Although the final output of spontaneous speech is necessarily linear or unidimensional through time, this does not mean that the psycholinguistic process underlying it is a simple unit-by-unit sequencing. Carroll (1953) suggested a hierarchical process, in which selection of larger units at "higher" levels is followed by selection of smaller units at "lower" levels, the smaller units filling in the larger constructions. More recently, Chomsky (1957) has challenged the entire notion of the unreeling of speech as a Markov process. Does the study of pausal and other hesitation phenomena provide us with any insight into the nature of psycholinguistic units and their selection?

Depending on which measure we pick, we seem to get a different picture of the functional units of encoding. Repeats characteristically involve whole, single words (suggesting the word as a unit at some level of analysis), yet they usually affect a function word and appear before lexical choices. When False Starts are retraced, it is usually a lexical item that is corrected, but the speaker also includes the closely associated func-

tion words in his retracing. This would suggest that at some level of organization the encoding unit is phrase-like, a lexical core with its tightly bound grammatical context. In at least superficial contradiction to this, however, is the fact that pauses of both types tend to occur just before lexical choices and thus often *between* the function-word context and the lexical core (e.g., *in the / house*). Yet—finally—statistical analysis of the distribution of Filled vs. Unfilled Pauses clearly shows that the former tend to coincide with the syntactical junctures at phrase boundaries, while the latter fall within phrases.

What implications for the nature of the "encoding" (speaking) process can we draw from these data? In the first place, the data as a whole suggest that the speaker is operating with units at least as large as the word, not smaller. But words always involve grammatical as well as lexical selection: e.g. (*walk/-* vs. *walk/s* vs. *walk/ing* etc.) versus (*play/-* vs. *play/s* vs. *play/ing* etc.). We must therefore infer something like a "mixer" just prior to the final common path of output. It is as if we had available at some lower level of encoding a "pool" of heavily practiced, tightly integrated word and phrase units, but selection from this pool requires simultaneous lexical and grammatical determinants. Secondly, the evidence as a whole suggests at least two levels of organization in encoding, which we may call lexical (or semantic) and grammatical (or structural). The best evidence for two such levels or classes of determinants is the fact that the speaker often begins a phrase and then pauses just before the lexical choice within that phrase, as if he had made a structural choice before fully selecting the precise lexical item from the many alternatives available. The statistically significant, if imperfect, complementary distribution between the "ah" and silence pauses also implies some such distinction between levels of selection.

What about the distinction between Filled Pause and Unfilled Pause? We suggest that the main distinction lies in *the duration of the non-speech interval*. Let us assume that the speaker is motivated to keep control of the conversational "ball" until he has achieved some sense of completion. He has learned that unfilled intervals of sufficient length are the points at which he has usually lost this control—someone else has leapt into his gap. Therefore, if he pauses long enough to receive the cue of his own silence, he will produce some kind of signal ([m, ər], or perhaps a repetition of the immediately preceding unit) which says, in effect, "I'm still in control—don't interrupt me!" We would thus expect Filled Pauses and Repeats to occur just before points of highest uncertainty, points where choices are most difficult and complicated. We have also noted that Filled Pauses tend to occur at the junctures of larger syntactical units, presumably where constructional decisions as well as decisions as to *what* to say, content-wise, are

being made. This assumption that "ah"-type pauses are reactions of the speaker to his own prolonged silences at points of difficult decision is consistent with our finding that these two pause-types are merely statistically, not absolutely, different in distribution. A difficult choice *can* occur almost anywhere, although it is more likely at points where both structural and content alternatives are being juggled.

We have the picture, then, of the speaker operating simultaneously on two levels of choices, lexical and grammatical. These are merged in the final selection of word and phrase units that are highly practiced skills, reflecting both types of determinants. The larger the unit being "programmed," the more complex the transformation in Chomsky's sense, or the less probable the sequence, the more prolonged the non-speech interval and hence the greater the tendency for an "ah" or a repetition. Since structural choices typically involve fewer alternatives than lexical choices, the speaker will often initiate a construction before he has completed his lexical decisions—with the result that he may pause slightly in the middle of phrases before such lexical items. Since he is monitoring the sense of his utterances more than their structure, and because errors of sense are more likely than errors of form, he will often halt after a lexical "miss," retrace, and correct it—but the unit of retracing is that of the skill sequence, and this typically includes the function-word context along with the lexical core. Speakers differ in how they manage the complex variables in spontaneous speech. Some slow down to avoid False Starts and syntactical disorder, using "ah's" and repetitions to hold on to the conversational line; others pause (silently) only where lexical difficulties force them to, rather recklessly committing False Starts and only extricating themselves from syntactical impasses by sheer disregard for the rules. This overall picture may be quite wrong, of course, but it seems consistent with the hesitation phenomena we have observed.

## Directions for Further Research

Our findings are consistent both with Lounsbury's predictions about where hesitation phenomena should occur and with Goldman-Eisler's reported observations and interpretations. However, the interpretation that such phenomena occur at points of high uncertainty, because they appear more before lexical items than before function words, is rather inferential. Also unsupported is our view that more Filled Pauses and Repeats occur before points of higher uncertainty than do Unfilled Pauses. These matters can be tested by comparing the Thorndike-Lorge frequencies of the lexical items following pauses with a random sample of lexical items following structurally similar points where pauses did *not* occur, and by

comparing the frequencies of lexical items following Unfilled Pauses with those following Filled Pauses and Repeats. We have not yet made the necessary analyses.

It is clear from the work of Mahl that the motivational and emotional states of speakers will influence the frequency of hesitation phenomena and perhaps their distribution. Apart from the interview situation, clinical and otherwise, in which Mahl has worked, it should be possible to study the effects of conversational stress (e.g., implied support or criticism from the audience), of status differences between speaker and listener, of the amount of planning or rehearsal of an utterance, and so forth. These phenomena can also be studied from the point of view of the listener. Presumably the hesitations that slow the speaker at points of uncertainty also permit the listener to catch up, as well as serving to stress the less predictable items. Debate coaches will often encourage their charges to "rough up" their presentations when the season is about half over. It is also possible that natural-appearing pauses and other hesitation phenomena influence the listener's connotative judgment of the speaker, e.g., of the speaker's "sincerity." Similarly, novelists will sometimes introduce pauses (symbolized orthographically by *um*, *er*, etc.) into the speech of a character to convey impressions that are not easily transmitted by the lexical content.

The periodicity of hesitation rate suggested by Figure 1 is of some interest, although we can at present do no more than speculate on the variables at work. Mahl has pointed to anxiety as one source of variation. If anxiety is a tension-producing phenomenon, vocalization is likely to be tension-reducing. This implies that a subject in an anxiety state as he begins to talk will show an initially high rate of hesitations followed by a decreasing rate as speech continues. The level of hesitations may increase again as a result of external factors or as a consequence of the content of his own utterances. An example of this, to which the authors can testify, is the extreme difficulty of talking about hesitation phenomena without having one's own disturbance rate rise rapidly. Changing transitional probabilities during an extended utterance would be another factor. The alternatives available to the speaker are likely to become more and more dispersed as he exhausts the familiar things to be said about any particular topic, and as he moves into areas of low probability words and sequences he will need more time for selection. When he switches topics, he may gain a new statistical foothold. This tendency is probably reinforced by the stylistic requirement of educated English to search for synonyms rather than to repeat previously used words. On the basis of comparison of a few curves like that in Figure 1, we suspect that subjects differ considerably in the periodicity of their hesitation rates, but what the correlates of these

differences may be (e.g., intelligence, personality variables, communication strategy), we do not know.

## REFERENCES


BLOCH, B. (1946). "Studies in Colloquial Japanese. II: Syntax," *Language* 22.200–248.

BLOOMFIELD, L. (1933). *Language* (New York).

CARROLL, J. B. (1953). *The Study of Language* (Cambridge, Mass.).

CHOMSKY, N. (1957). *Syntactic Structures* (The Hague).

FREUD, S. (1938). "The Psychopathology of Everyday Life," in *The Basic Writings of Sigmund Freud* (New York).

FRIES, C. G. (1952). *The Structure of English* (New York).

GOLDMAN-EISLER, F. (1954*a*). "A Study of Individual Differences and of Interaction in the Behavior of Some Aspects of Language in Interviews," *Journal of Mental Science* 100.177–197.

GOLDMAN-EISLER, F. (1954*b*). "On the Variability of the Speed of Talking and on Its Relation to the Length of Utterances in Conversations," *British Journal of Psychology* 45.94–107.

GOLDMAN-EISLER, F. (1955). "Speech-Breathing Activity—a Measure of Tension and Affect During Interviews," *ibid.*, 46.53–63.

GOLDMAN-EISLER, F. (1957). "Speech Production and Language Statistics," *Nature*, Vol. 28, December, p. 1497.

GOLDMAN-EISLER, F. (1958*a*). "Speech Production and the Predictability of Words in Context," *Quarterly Journal of Experimental Psychology* (in press).

GOLDMAN-EISLER, F. (1958*b*). "Speech Analysis and Mental Processes," *Language and Speech* 1.59–75.

HARRIS, Z. S. (1951). *Methods in Structural Linguistics* (Chicago).

JOOS, M. (1950). "Description of Language Design," *Journal of the Acoustical Society of America* 22.701–708.

MAHL, G. (1956*a*). "Disturbances in the Patient's Speech as a Function of Anxiety." Paper read to the Eastern Psychological Association.

MAHL, G. (1956*b*). "'Normal' Disturbances in Spontaneous Speech." Paper read to the American Psychological Association.

MAHL, G. (1957). "Disturbances and Silences in the Patient's Speech in Psychotherapy," *Journal of Abnormal and Social Psychology* 42.3–32.

MAHL, G. (1958). "The Instrumental Model and Relevant Research Methods," in *Trends in Content Analysis*, I. D. Pool, ed. (in press).

OSGOOD, C. E. and SEBEOK, T. A., eds. (1954). *Psycholinguistics: a Survey of Theory and Research* (Baltimore).


*Institute of Communications Research*
*University of Illinois*
*Urbana, Illinois*