HESITATION AND GRAMMATICAL ENCODING

DONALD S. BOOMER*
National Institute of Mental Health, Bethesda

The occurrence of filled and unfilled pauses was examined with respect to their location in phonemic clauses. Both types of hesitation were most frequent after the first word in the clause, regardless of length. These data are regarded as directly challenging the transitional probability theory of hesitations. The phonemic clause is proposed as the encoding unit of speech at the grammatical level.

This study is concerned with two varieties of hesitations in spontaneous English speech : silent pause and "filled pause" (*uh* /ə/, *ah* /a/, *um* /əm/, and similar variants). The data to be presented concern the location of these hesitations in extended utterances, but the basic theoretical issue involved is the nature of the grammatical encoding process in speech. The data and the theoretical issue are inferentially related. The linking hypothesis is that hesitations in spontaneous speech occur at points where decisions and choices are being made. On this basis, the patterning of hesitations should provide clues as to the size and nature of the encoding units which are operative.

If the encoding units are single words then hesitations should occur more frequently before those words which involve a difficult decision ; i.e., a choice among many alternatives. If the encoding unit is a sequence of several words then the hesitations should predominate at the beginnings of such sequences, rather than occurring randomly wherever a difficult word choice occurs. The present inquiry has sought to establish a critical confrontation of these two alternatives.

BACKGROUND OF THE PROBLEM

In general, previous investigations have considered hesitations to be a function of the statistical transition probabilities which obtain in the sequencing of words. Lounsbury (1954, p. 99), in an extended discussion of this issue, puts the case in the form of a hypothesis :

> *Hypothesis 1 : Hesitation pauses correspond to the points of highest statistical uncertainty in the sequencing of units of any given order.* (High statistical uncertainty = high transitional entropy.) The observations which lead us to formulate this hypothesis have been focussed on the sequencing of words. We are relatively hopeful for the substantiation of the hypothesis when the units are of this order.

* *Laboratory of Psychology, National Institute of Mental Health, National Institutes of Health, Department of Health, Education and Welfare.*

Goldman-Eisler (1958) tested this hypothesis experimentally on a small language sample with positive results. A set of 12 spontaneously uttered sentences containing 60 unfilled pauses were transcribed and subjected to the Shannon guessing technique for estimating transitional probabilities between successive words. The words which had been preceded by a pause when originally uttered proved to be significantly harder to guess than were the words which had been spoken fluently. The sample sentences, however, were a markedly biased sample of spontaneous speech, since the experimental design dictated that only grammatical, well constructed sentences be chosen. The difficulty Goldman-Eisler reports in finding such utterances in her corpus attests to their unrepresentativeness. The function of pauses in this restricted context cannot safely be generalized to ordinary spontaneous speech.

In a larger corpus of 50,000 words tape-recorded in a work conference, Maclay and Osgood (1959) studied the incidence of both filled and unfilled pauses relative to lexical words and function words. About 59% of all hesitations occurred before lexical words and 41% before function words, a finding which, although statistically significant, leaves sufficient unexplained variance to warrant further investigation.

The present study was based on the belief that a systematic account of hesitation phenomena may require a unit larger than the word. Previous empirical studies based on this idea have turned to formal grammar for the definition of the larger units. Maclay and Osgood (1959) used some of the grammatical test frames of Fries (1952), and Little (1963) used a structural classification system described by Francis (1958). Both investigations found that hesitations occur within many syntactic structures as well as at their boundaries. Some systematic variability was found in both investigations; that is to say, hesitations were significantly more probable at some syntactic positions than at others. Although both studies are enlightening, neither provides any systematic basis for defining speech encoding units in terms of formal structural linguistics.

In another hypothesis Lounsbury (1954, p. 100) also considered units larger than the word :

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

This could serve as the basic hypothesis of the present study if the words " and points of high statistical uncertainty " were deleted. The reason for this qualification will be clarified later in the paper in the course of an argument against the assumption that pauses occur at points of low word-to-word transition probabilities.

In discussing Lounsbury's hypotheses Maclay and Osgood (1959, p. 23) say, " Hypothesis 2 has an element of circularity in that no independent method of defining encoding units has been developed." This is the point of departure of the present study. The *phonemic clause,* as defined by Trager and Smith (1951), has been adopted as a provisional encoding unit. This unit is defined phonologically and does not depend on the presence or absence of hesitations. The incidence of hesitations

in phonemic clauses, then, may be examined, with the hypothesis that they occur predominantly at the beginnings of these units.

The phonemic clause is a phonologically marked macrosegment which, according to Trager and Smith, contains one and only one primary stress and ends in one of the terminal junctures /|, ||, #/. In practice, the instances of terminal junctures are determined from the tape recording and marked on the transcript. The boundaries so marked determine successive phonemic clauses in the corpus.

Procedure

Speech samples. Tape recordings of spontaneous speech of 16 male native American speakers of English had been made for an earlier study (Boomer and Dittmann, 1964). These recordings were used in the present investigation. After a "warm-up" chat each subject had been asked to speak extemporaneously to an interviewer for about three minutes on any subject he chose; hobbies, sports, summer vacations, or the like. Every effort was made to provide a relaxed atmosphere in which the subjects could talk naturally.

Speech was recorded on an Ampex 350 tape recorder, and a simultaneous oscillograph record was made for the purpose of locating silences. An automatic timer registered signals at 30 sec. intervals on both the oscillograph record and the audio tape to facilitate subsequent matching of the records. Careful typewritten transcripts were prepared in conventional orthography. The location of each silence exceeding 200 msec. was determined from the oscillograph records and marked on the transcripts. This is just slightly below the 250 msec. cutting point proposed by Goldman-Eisler (1958) for distinguishing hesitation pauses from non-pausal phonetic effects. All filled pauses were also located and underlined.

Linguistic analysis. The phonemic clause boundaries were established without knowledge of the purpose of the study by a colleague[1] who had been taught suprasegmental analysis by Trager. His task was to listen to the recordings and to record terminal junctures on an otherwise unmarked transcript. Repeat reliability of 0.93 has been reported by Dittmann and Wynne (1961) for the location of terminal junctures.

Hesitations. The object of analysis was to establish the location of each hesitation form in the phonemic clause in which it occurred. Our experience accords with statements made by previous investigators (Maclay and Osgood, 1959; Little, 1963) that filled and unfilled pauses occur almost without exception at word boundaries.[2]

[1] *Allen T. Dittmann, whose interest in psycholinguistics preceded and stimulated my own.*

[2] *In order to sidestep the prolonged linguistic dispute about this unit it should be emphasized that "word" is used here in its most traditional sense, as defined by a typist when she presses the space bar.*

The successive word boundaries in a given phonemic clause, then, can be regarded as an ordered series of opportunities for hesitation. The possible locations in a five-word clause are illustrated below :

ˌandˌtheˌweatherˌwasˌhot #

The position after *hot* is not considered as a part of the clause, since the utterance is closed by the terminal contour over *hot*. More generally, the presence or absence of a pause after the terminal contour is irrelevant to the determination of juncture and is thus not linguistically a part of the preceding clause. A pause in this position, then, must be assigned to the following clause. Whether or not it functions as a true hesitation is a separate issue.[3]

In general, there will be as many possible locations as words in the clause, each location being labelled with the ordinal number of the word it precedes. Occasional arbitrary exceptions were made in this study for multiple-element proper nouns such as *Bill Smith* and *San Francisco,* for combinatory groups like *thank you* and *what-you-may-call-it,* and for certain " tags " such as *you know* and *you see.* These were counted as single words, as were syntactically superfluous repetitions of words, as in *I took the . . . the train.* Filled pauses themselves and word-fragments were also excluded from the count.

The corpus contained a total of 1593 phonemic clauses of which 713 contained one or more hesitations. Hesitations totalled 1127, 749 unfilled pauses and 378 filled pauses. Each hesitation in the transcript was tabulated by its location number within the clause. The location and tabulation were done independently by two people and cross checked to minimize error.

Results

The hypothesis that hesitations tend to occur at the beginning of phonemic clauses was strongly supported with, however, an unpredicted reversal. As may be seen in Table 1,[4] the greatest frequency of hesitations is not at the outset but at position 2, after the first word of the clause. This is true for all nine of the array distributions

[3] *The pause following terminal juncture is mentioned in the systems of Bloomfield (1933), Harris (1951), and Hill (1958). Despite terminological differences, there is general agreement that this pause is linguistically determined, an intermittently occurring free variant of the terminal juncture. Hesitation pauses, that is, pauses occurring at other than structural boundaries, when mentioned at all are attributed to non-linguistic or extra-linguistic factors.*

[4] *A few clauses, 30, to be exact, longer than 10 words were omitted from this tabulation because there were too few in any length category to permit statistical treatment. In order to exceed 6 or 7 words a clause must usually include one or more extended anacolutha, a fact which may account for the " humps " near the ends of the distributions for the 8 and 9 word clauses. The data tabulated here include 97% of the hesitations in the original corpus.*

TABLE 1

Frequencies of hesitations located at successive word boundaries in phonemic clauses classified by length.

NUMBER OF WORDS IN CLAUSE	BOUNDARY LOCATIONS										
	1	2	3	4	5	6	7	8	9	10	Totals
2	18	28									46
3	16	68	29								113
4	34	81	39	27							181
5	28	80	37	22	21						188
6	21	65	22	26	16	17					167
7	14	67	25	20	13	19	8				166
8	9	38	18	11	11	14	13	7			121
9	7	19	13	15	4	6	6	4	13		87
10	4	15	4	8	5	7	5	3	4	3	58
Totals	151	461	187	129	70	63	32	14	17	3	1127

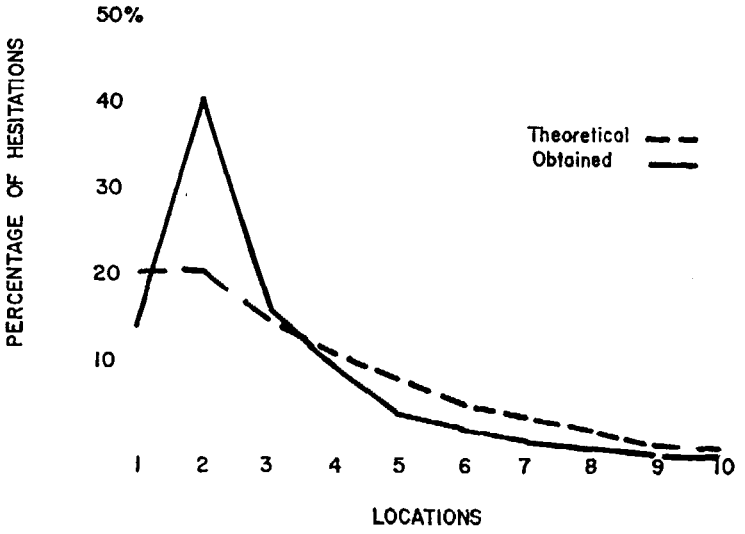

Fig. 1. Comparison of theoretical and obtained total percentages of hesitations occurring at successive boundary locations.

TABLE 2

Frequencies of unfilled pauses located at successive word boundaries in phonemic clauses classified by length.

NUMBER OF WORDS IN CLAUSE	BOUNDARY LOCATIONS (J)	1	2	3	4	5	6	7	8	9	10	Totals
2	(65)	6	22									28
3	(99)	5	46	18								69
4	(101)	16	49	29	21							115
5	(109)	10	50	29	15	17						121
6	(98)	9	47	16	18	11	12					113
7	(76)	5	45	19	14	11	13	7				114
8	(55)	4	27	14	9	9	12	8	4			87
9	(36)	2	14	10	12	3	3	4	4	7		59
10	(20)	1	10	3	4	5	6	5	3	3	3	43
Totals	(659)	58	310	138	93	56	46	24	11	10	3	749

representing clause lengths from two to ten words. An interpretation of this result will be offered in the final discussion.

Tables 2 and 3 show the separate distributions of unfilled and filled pauses respectively. This breakdown was included to demonstrate that the distributions of both kinds of hesitation conform essentially to the total distribution. Of the 18 arrays in the two tables, 16 show the same pattern with the highest frequency at position 2. The two discrepant arrays, clause lengths 2 and 9 in Table 3, may represent sampling error resulting from the small number of entries.

In order further to clarify the data in Table 1, they are presented graphically in Fig. 1 against a hypothetical chance distribution. The dotted line represents the expected total distribution of hesitations in the entire set of phonemic clauses if their occurrence were governed by chance. The solid line represents the distribution of column totals from Table 1. Both sets of values are represented in percentage terms to facilitate comparison. As may be readily seen, only the frequency at position 2 exceeds chance. All other frequencies are at or below chance expectancy.

One other aspect of these data requires discussion at this point. In Table 2 the frequency of juncture pauses is separately tabulated in column (J). In column 1 are tabulated the unfilled pauses in position 1 which are not juncture pauses, that is to say, those pauses which were preceded by an initial *ah* or a word fragment which opened the clause but was not counted as a word. Had the juncture pauses been simply included in column 1 the totals would be greater than for position 2 in each array. The initial hypothesis that hesitations occur at the beginning of phonemic

TABLE 3

Frequencies of filled pauses located at successive word boundaries in phonemic clauses classified by length.

NUMBER OF WORDS IN CLAUSE	BOUNDARY LOCATIONS										
	1	2	3	4	5	6	7	8	9	10	Totals
2	12	6									18
3	11	22	11								44
4	18	32	10	6							66
5	18	30	8	7	4						67
6	12	18	6	8	5	5					54
7	9	22	6	6	2	6	1				52
8	5	11	4	2	2	2	5	3			34
9	5	5	3	3	1	3	2	0	6		28
10	3	5	1	4	0	1	0	0	1	0	15
Totals	93	151	49	36	14	17	8	3	7	0	378

clauses would be even more strongly supported, but the interesting reversal between 1 and 2 would be obscured. We have therefore chosen to maintain the distinction between these types of pause, but to present all of the data for full consideration. Some of the theoretical issues involved will be discussed in the concluding section of this report.

Additional findings. Two minor issues regarding hesitations were examined in this body of data. Although peripheral to the central hypothesis of the study they are at least tangentially of interest.

These data permitted a test of a psycholinguistic interpretation of *ah* offered by Maclay and Osgood (1959, p. 41). These authors say :

> 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, r], or perhaps a repetition of the immediately preceding unit) which says, in effect, " I'm still in control—don't interrupt me."

This hypothesis is subject to quantitative test although Osgood and Maclay were unable to make such a test because of the limitations of their data. The oscillographic recordings available in the present study permitted the measurement of unfilled pauses immediately preceding instances of *ah*. The results of this test do not support the

Maclay-Osgood hypothesis. More than 72% of the *ahs* were preceded by no pause. An additional 11% were preceded by pauses which were shorter than the subject's mean pause length. Thus only 17% of the *ahs* in this study were preceded by longer than average pauses.

The other issue concerns a possible durational distinction between juncture pause and hesitation pause. This widely cited distinction was originally formulated by Lounsbury (1954, p. 98) :

> Even if 'junctures' sometimes consist of short pauses, the [hesitation] pauses under consideration here are not the same. For one thing, there is a difference in duration. Juncture pauses which we have seen in spectrographic analyses of speech were in the order of a hundredth of a second or less in length. The pauses referred to here, however, are appreciably longer.

In the present data about half the junctures were followed by a pause. According to Lounsbury these pauses should be much shorter than the hesitation pauses, that is, pauses which occur at locations other than following junctures. This does not seem to be the case. The juncture pauses as measured are, in fact, somewhat longer than the hesitation pauses. The overall means are 1,027 msec. and 747 msec. respectively, a difference which is significant, $p < 0.001$. Furthermore, 15 of the 16 subject mean comparisons are in the same direction.

Discussion[5]

This study has used the location of hesitations in spontaneous speech as evidence in support of the theory that speech encoding at the grammatical level operates with units larger than the word. When sustained speech is marked off in successive phonemic clauses, bounded by terminal junctures, it has been shown that hesitations are not randomly located throughout such clauses, but tend to occur near the beginnings. This tendency characterizes both types of hesitation studied, all clause lengths from two to ten words, and each of the 16 individual subjects.

If words were in fact the operative encoding units there would be no reason to expect such a finding. In fact, a strict word-transitional model should predict just the opposite. The argument is as follows : primary stress typically occurs toward the end of a phonemic clause ; almost invariably the last or next to last word in the clause receives the stress. And, as Berry (1953) has shown, primary stress is negatively related to word frequency ; that is, the less frequent a word, the more likely it is to receive phrase stress when uttered in sustained speech. Thus the high-information lexical words tend to occur toward the end of phonemic clauses, and this is where a word-unit model should predict the most hesitations to occur. That, in fact, they

[5] *The author gratefully acknowledges the skilled linguistic counsel of C. I. J. M. Stuart. Prof. Stuart, of course, bears no responsibility for any views expressed herein.*

do not argues strongly for the phonemic clause rather than the word as the molar encoding unit of speech.

Let us now consider the unexpected finding that the most frequent pause location is after the first word in the clause, rather than before it, as had been predicted. This fact may have some additional implications for the psycholinguistic process of speech encoding.

Carroll (1953) suggested a two-stage, hierarchical encoding process with larger units chosen first, followed by the selection of smaller components. The larger units may be thought of as involving grammatical decisions and the smaller units as involving lexical choices. The reality of these separate aspects of speech production is strongly supported by evidence from clinical research (Wepman and Jones, 1964) that in aphasia either function can be impaired, leaving the other relatively intact. A medico-linguistic distinction was drawn earlier by Jakobson (1956) between " contiguity " and similarity " disorders.

In the neurologically intact speaker these processes are so smoothly co-ordinated that they seem to be a single process. The temporal ordering of these two encoding stages, if indeed there is an ordering, remains a subject of dispute. Skinner (1957) has speculated that lexical selection is the fundamental process and that such features as syntactic arrangement, inflection and qualification are added as intraverbal, " auto-clitic " responses, that is, verbal responses to previously uttered verbalizations.

In a carefully reasoned review of Skinner's work, Chomsky (1959, p. 54) argues that the opposite is just as likely—" . . . that the ' key responses ' are chosen only after the grammatical frame."

The present data suggest that Chomsky's view is the more correct one. The initial word in a phonemic clause sets certain constraints for the structure of what is to follow. The selection of a first word has in greater or lesser degree committed the speaker to a particular construction or at least a set of alternative constructions, and has also foreclosed the possibility of other constructions.[6]

According to this view, then, the hesitations in phonemic clauses are most likely to occur *after* at least a preliminary decision has been made concerning its structure and *before* the lexical choices have been finally made.

This interpretation rests on the preponderance of hesitations at position 2, which in turn rests on our rather arbitrary decision not to consider juncture pauses as true hesitations. There is some theoretical and empirical basis for our decision. Previous experimentation (Boomer and Dittmann, 1962) has shown that listeners do not distinguish between minimal-pair utterances contrasting a closed juncture and a juncture followed by a 500 msec. pause, whereas a hesitation pause of 200 msec. is discriminated. This empirical finding supports the generally accepted theoretical view that juncture pauses are facultative accompaniments of terminal junctures and essentially linguistic, while hesitation pauses are extra-linguistic.

To be sure, neither the evidence nor the theoretical distinction is compelling, but

[6] *Cf. Hockett, C. F. (1960).*

pending better information we will continue to regard these two pausal phenomena as functionally different.

In passing, it is suggested that the above-mentioned experiment may have some bearing on the generally-held opinion that juncture pauses are vanishingly short as compared to hesitations. This is perhaps true for closed junctures, and about half the junctures in this study were not followed by pauses exceeding 200 msec., our cut-off point. Half, however, were followed by juncture pauses which were significantly longer than hesitation pauses. Even at these lengths, however, the unaided ear fails to register them. It is likely that impressionistic judging procedures and even stop-watch pause timing miss many of these pauses.

Hesitations and transition probabilities. The prevailing view that hesitations reflect low transition probabilities at the point of their occurrence has been directly challenged by this study. The transitional-probability hypothesis contains the unexamined assumption that the speech encoding process does not involve any anticipatory planning beyond the next word. The speaker is, according to this view, not aware that he is at a loss for a word until he is on the point of uttering it, at which point he hesitates long enough to make the choice. This would seem to be an unnecessarily primitive assumption in view of the demonstrable capacities of the human brain for short-term storage, delay, and "feed-back" control of rapid, complex sequential motor activities.

Our data imply a rather more complex process in which planning ranges forward to encompass a structured "chunk" of syntax and meaning. As a given clause is being uttered the next one is taking shape and focus. At the terminal juncture the next clause may be ready, in which case it will be uttered fluently, as were more than half the phonemic clauses in this corpus.

If, however, the emerging clause has not yet been subjectively formulated, speech is suspended until the entire pattern is clarified. This suspension may be manifested as either a pause or a vocalized hesitation sound. Our data do not support any functional distinction between these hesitation signals.

This hypothetical description, of course, fails to account for all the data. Our distributions show that some hesitations occur at each successive position in some clauses, although with declining probability overall. It would be surprising if the case were otherwise. The psychological monitoring processes are undoubtedly delicate, sensitive, and subject to modification at any point. Our data argue only that most decisions in most clauses occur at the first internal word boundary.

References

BERRY, J. (1953). Some statistical aspects of conversational speech. In *Communication Theory*, ed. W. Jackson (London).

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

BOOMER, D. S. and DITTMANN, A. T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, 5, 215.

BOOMER, D. S. and DITTMANN, A. T. (1964). Speech rate, filled pause, and body movement in interviews. *J. nerv. ment. Dis.*, 139, 324.

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

CHOMSKY, N. (1959). Review of 'Verbal Behavior' by B. F. Skinner. *Language*, 35, 26.

DITTMANN, A. T. and WYNNE, L. C. (1961). Linguistic techniques and the analysis of emotionality in interviews. *J. abnorm. soc. Psychol.*, 63, 201.

FRANCIS, W. N. (1958). The Structure of American English (New York).

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

GOLDMAN-EISLER, FRIEDA (1958). Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, 9 - 10, 96.

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

HILL, A. A. (1958). Introduction to Linguistic Structures (New York).

HOCKETT, C. F. (1960). Grammar for the hearer. *Proc. Symposia in Applied Mathematics, Vol. 12, New York*, 220 (Providence, R. I.).

JAKOBSON, R. and HALLE, M. (1956). Fundamentals of Language (The Hague).

LITTLE, L. (1963). The syntax of vocalized pauses in spontaneous cultivated speech. Ph.D. Dissertation, George Peabody College for Teachers.

LOUNSBURY, F. G. (1954). Transitional Probability, Linguistic Structure and Systems of Habit-family Hierarchies. In *Psycholinguistics: a survey of theory and research problems*, ed. C. E. Osgood and T. A. Sebeok (Baltimore).

MACLAY, H. and OSGOOD, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19.

SKINNER, B. F. (1957). Verbal Behavior (New York).

TRAGER, G. L. and SMITH, H. L. (1951). An Outline of English Structure (Norman, Okla.).

WEPMAN, J. M. and JONES, L. V. (1964). Five aphasias : a commentary on aphasia as a regressive linguistic phenomenon. In *Disorders of Communication*, ed. D. M. Rioch and E. A. Weinstein (Baltimore).