JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 4, 134-140 (1965)

Word Predictability in the Environments of Hesitations[1]

PERCY H. TANNENBAUM, FREDERICK WILLIAMS, AND CAROLYN S. HILLIER

Mass Communications Research Center, University of Wisconsin, Madison, Wisconsin

Two experiments were conducted to study the predictability of words in hesitation contexts. The first study focused on a comparison of the first word after hesitations with words sampled from fluent contexts. The second study involved gathering predictability data for all words in a language sample.

Results supported the hypothesis that words subsequent to hesitations tend to be less predictable than words uttered in fluent context. But the associated hypothesis that the word antecedent to hesitations is more predictable than other fluent context was not supported. This led to further analysis of predictability of words in the environments of different hesitations, specifically filled pauses and repeats. The implication drawn was that different types of hesitations index different kinds of encoding decision points.

Various hesitation phenomena apparently occur in almost all speakers of all languages. Generally, they have been taken as interruptions in the "normal" flow of speech and, as such, have been used to index variations in the encoding state of speakers (e.g., Mahl, 1956, 1959; Feldstein and Jaffe, 1962; Stolz and Tannenbaum, 1963). Recently, attention has been focused on the role of hesitations as indices of different encoding decision points (Lounsbury, 1954; Goldman-Eisler, 1958a, 1958b, 1961a), and it was to this issue that the present investigation was directed. More specifically, the present research was concerned with assessing the predictability of words in hesitation vs. fluent encoding contexts, across and between different types of hesitations.

The impetus for such research stems from a series of studies by Goldman-Eisler. Originally interested in language phenomena in psychiatric interviews, she found that speech interruptions involved both breathing and hesitation pauses, but while the former related more to the emotional state of the encoder, hesitation pauses were more cognitive in origin, apparently characterizing conditions of encoder uncertainty (Goldman-Eisler, 1955, 1956).

In subsequent tests of this "cognitive decision point" hypothesis, she found that the first lexical word subsequent to an unfilled pause was less predictable (Goldman-Eisler, 1958a) and took significantly longer to replace (Goldman-Eisler, 1958b) than lexical words uttered in fluent contexts. In addition, the word preceding unfilled pauses tended to be even more predictable than in other fluent contexts. Thus, it was reasoned that hesitation pauses had occurred at junctures representing transitions from relatively high to low redundancy, the original speaker having paused at that point in order to make a particular encoding decision. Further support for this interpretation came from another study (Goldman-Eisler, 1961a) in which the frequency of hesitation pauses was found to vary with the

[1] This research was supported in part by a grant from the Graduate Research Committee, University of Wisconsin, from funds supplied by the Wisconsin Alumni Research Foundation, and in part under Grant GS-296 from the National Science Foundation.

level of cognitive activity required of encoders. Such pauses were more frequent in "interpretations" rather than "descriptions" of subtle cartoons, and diminished with increasing repetition of the same encoding task.

A related descriptive study led to the conclusion that hesitations were non-random in their occurrence and that such phenomena "are clearly related to the dynamics of grammatical and lexical selection" (Maclay and Osgood, 1959, p. 40). An important aspect of this study was the use of a variety of hesitation phenomena; a classification of four different types (false starts, repeats, and filled and unfilled pauses), representing a reduction of Mahl's (1956) eight categories, was employed. In keeping with the Goldman-Eisler and Lounsbury (1954) theoretical positions, unfilled pauses tended to occur antecedent to likely encoding choice points, but so did filled pauses and, to a somewhat lesser extent, repeats. False starts, on the other hand, often involved blocking subsequent to a lexical unit and then retracing to correct it. Goldman-Eisler (1961b) pursued the suggestion that filled and unfilled pauses might characterize similar decision points by re-analyzing the data from her earlier "cartoon" study (Goldman-Eisler, 1961a) for the occurrence of filled pauses. Unlike the findings for unfilled pauses, there were no significant differences between the description and interpretation encoding conditions for filled pauses. This suggested that these two hesitation phenomena reflect different internal processes—unfilled pauses still relating more to the encoder's cognitive decisions, but filled pauses more to his emotional state.

The present research involved two separate but related studies investigating the cognitive decision point theory from a somewhat different methodological standpoint, and leading to the implication that different hesitation phenomena characterize different types of encoding decisions. Because of their somewhat different foci, the two studies are reported separately.

Experiment I

The first phase of the research was a direct test of the hypothesis that words immediately following hesitations are less predictable than those originally uttered in other, more fluent, sequences. As such, this study represented a replication of the earlier Goldman-Eisler (1958a) test of the same hypothesis. However, it involved some distinctive methodological modifications, including the selection of appropriate language samples as test materials, the inclusion of a wider variety of hesitation types, and the use of a different procedure to assess word predictability. These methodological variations were designed to provide for a more sensitive assessment of contextual predictability and, hence, of relating hesitation phenomena to encoder uncertainty.

Method

Language Samples. In order to allow for a closer correspondence between the original encoding behavior and the decoding behavior involving word replacement, the present study employed an entire single message, rather than a variety of independent sentences (cf. Goldman-Eisler, 1958a). Spoken messages are rarely encoded as independent sentences, and the constraints influencing word predictability can certainly exist beyond sentence boundaries.

The message used was encoded as part of another study dealing with the effects of cognitive feedback on oral encoding behavior (Stolz and Tannenbaum, 1963). Under the guise of an examination, students in a psychology of speech course replied to a set of three questions; the instructors' apparent evaluations being provided on a mechanical feedback meter. The first question was answered by all *S*s under the neutral, or no feedback condition, and the remaining two questions were systematically varied between positive (favorable meter reading) and negative (unfavorable) feedback conditions. The message selected for this study was encoded by a male under neutral feedback.

Identification of Hesitations. The recorded message was transcribed, and the location of the various hesitations was noted in the manner suggested by Maclay and Osgood (1959): *Unfilled pauses* were judged by trained coders as "silences of unusual length," rather than the Goldman-Eisler measure of including all silences over .25 sec duration. This admittedly more subjective index was used to assure that true hesitation silences were included without possible contamination from non-hesitation silences, such as those due to different speaker rates, "emphasis" pauses, etc. *Filled pauses* included all non-phonemic vocalizations —usually "ahs." *Repeats* included all repetitions judged to be semantically non-significant—from a phoneme to a series of words, although the most frequent case involved a single word. *False starts* involved all incomplete or "self-interrupted" utterances. The experimental message was of moderate length (273 words) and was marked by a relatively wide-spread hesitation distribution, including three false starts, eight repeats, sixteen filled pauses, two unfilled pauses, and four combinations.

Predictability measures. Another important distinction involved the manner in which the main dependent variable of word predictability or uncertainty was assessed. Goldman-Eisler used a variation of the Shannon (1951) method of estimating entropy in English, allowing for the influence of the antecedent or the subsequent, but not both contexts. Obviously, different semantic and syntactic constraints governing word predictability can operate in the complete context than in its components.[2] Accordingly, cloze procedure (cf. Taylor, 1953, 1954; Osgood, 1959), which involves the systematic deletion of a sample of words from the entire passage, was employed here.

Procedure. Two versions of the transcribed message were prepared. One had a 10-space underlined blank in the place of words which had followed each of the 33 hesitations. The second version had an equal number of blanks in place of words selected on a purely random basis from the remaining corpus.

A total of 82 *S*s was randomly assigned to each of two equally-sized ($N = 41$) groups, each group receiving one of the two message versions. The *S*s were students enrolled in the same course which supplied the original encoder, but in a subsequent semester, thus also providing a closer correspondence between the encoder and response groups than in the Goldman-Eisler procedure.

In the manner characteristic of cloze procedure (cf. Taylor, 1953), all *S*s were instructed to "try to replace the missing words so that the passage makes sense." The *S*s were urged to respond on all blanks, even if it involved guessing.

Results

The responses were first analyzed in terms of a set of measures related to the degree of agreement between the original encoder and the responding *S*s: the proportion of verbatim replacements (V) of the deleted word; a similar measure of agreement of the grammatical form class between the original and replaced words (FC); and a V/FC index, reflecting the predictability of an actual word given that its grammatical classification has been correctly ascertained (*cf.* Fillenbaum and Jones, 1962).

For the present study, the prediction was that V, since it reflects a semantic choice, would be lower in the after-hesitation position than in fluent sequences of the message. It can be argued that the FC score is more related to the syntactic nature of the language than to deliberate semantic decisions and, thus, it should not necessarily discriminate between the two versions to the same degree as does V. Following the above reasoning, V/FC may be regarded as a more sensitive index of cognitive activity, since it adjusts for variations in the form class replacement.

The data for these three measures of encoder-respondent agreement are indicated in Table 1, and conform to the above predictions. The differences on V and V/FC are both in the predicted direction and statistically significant ($p < .02$, one-tailed Mann-Whitney U test, in each case). Although also in the predicted direction, the FC data do not show a significant difference.

TABLE 1
MEAN WORD REPLACEMENT MEASURES IN FLUENT AND AFTER-HESITATION CONTEXTS

Measure	Fluent context	After hesitation
V	.58	.37
FC	.79	.74
V/FC	.73	.50
Rel. H.	.28	.42

Table 1 includes an additional relevant measure relating to degree of uncertainty among the actual responses, independent of the original deleted word. It is indexed by the relative entropy (*Rel. H*) of the response distribution, the measure assuming a range of

[2] Unpublished research by one of the authors (FW) has shown that word predictability in the fully combined context is significantly different from that obtained by Goldman-Eisler's so-called "combined score" of merely averaging the antecedent and subsequent context scores. Much the same point has been made by Garner and Carson (1960) in regard to predicting letters in English.

zero to + 1.00, corresponding to the range of minimum to maximum uncertainty.[3] Thus, for the present study, the prediction would be for a higher value of *Rel. H.* in the after-hesitation condition. Again, the data in Table 1 confirm this prediction, the difference being significant beyond the .05 level.

Experiment II

The second study involved a more detailed attempt to explore word predictability in the environment of hesitations. Instead of a mere dichotomy between "after-hesitation" and "other" word locations, predictability measures were obtained for all words in the sample; and these could then be related to the position of a given word relative to the nearest hesitation.

Such a more detailed approach obviously allows for a more sensitive test of the basic hypotheses derived from earlier work. Again the predictions are for the greatest uncertainty at the position immediately subsequent to a hesitation, while significantly less uncertainty should obtain in all the other positions in the hesitation-environment. Further, in keeping with the Goldman-Eisler (1958a) finding, we would expect the lowest uncertainty of all in the position immediately preceding the hesitation.

Method

For this experiment, two additional messages were selected from the same feedback study that produced the first message. The two passages, totaling 440 words and including 56 instances of hesitation, came from two separate encoders—a male and a female—both encoding under conditions of negative feedback (cf. Stolz and Tannenbaum, 1963).

By systematically rotating the selection of the first word to be deleted and then maintaining an every *n*th word deletion scheme, one can exhaust all the words in *n* versions of the passage. In the present case, Taylor's (1954) procedure was followed, and an every-*fifth*-word deletion pattern was used, five such versions for a passage. Thus, in any one version, there was a four-word context presented both antecedent to and subsequent to every response blank. To keep this pattern constant, the first four words and last four words in each passage were not included in the deletion scheme.

The same 82 *Ss* who served in the first experiment were involved in the present one. They were randomly assigned to one of five groups; each group being exposed to one of the five versions of each of the two experimental passages. The instructions to the *Ss* were the same as those in the first study.

The original tape recordings of the two spoken messages were again scrutinized for the location and type of each hesitation, as in the first experiment. Thus, each of the deleted words could be identified as to its position in the environment of a given hesitation—i.e., by how many words it preceded or followed the hesitation. This was originally done for all words in the samples. However, because of the nature of the distribution of hesitations, it was decided to restrict the analysis to the immediate six-word environment—the three antecedent words (designated as a_3, a_2, a_1) and the three subsequent words (s_1, s_2, s_3)—of a given hesitation. Going beyond this limit would have involved too much confounding of both hesitations and words.

Results

In addition to the dependent variables employed in the first study, another measure, referring only to replacement of *lexical* words in the environment (V_{lex}), was included. This restriction to lexical items stems from Goldman-Eisler's (1958a) reasoning that it is specifically such words that are directly related to the cognitive activities involved.

Table 2 presents the word predictability means for the word positions both antecedent and subsequent to hesitations. Rather uniform results obtain across the various measures.

TABLE 2
MEAN WORD REPLACEMENT MEASURES IN ANTECEDENT AND SUBSEQUENT HESITATION CONTEXTS

	Word position					
Measure	a_3	a_2	a_1	s_1	s_2	s_3
V	.51	.53	.40	.39	.51	.53
V_{lex}	.46	.44	.29	.31	.42	.40
FC	.76	.78	.67	.66	.73	.77
V/FC	.67	.68	.60	.59	.70	.69
Rel. H.	.44	.43	.49	.48	.42	.41

[3] Relative entropy represents the ratio of the obtained entropy of the response distribution to the maximum entropy if all the events were equiprobable, i.e., Rel. H. $= \dfrac{-\sum_{i=1}^{m} p(i) \log_2 p(i)}{\log_2 m}$ where $p(i)$ is the probability of the occurrence of the *i*th event of *m* possible events (cf. Wilson, 1954; Taylor, 1954).

The largest predictabilities are in the more peripheral positions (a_3, a_2, and s_2, s_3). The greatest uncertainty is in the s_1 position, as predicted. But surprisingly, the a_1 position exhibits an equal degree of uncertainty.

These findings are underlined in comparisons between adjacent positions, with the number of words in each category used as the replicates in a Mann-Whitney U test. The differences between the a_2 and a_3 positions fall far short of statistical significance, as does the difference between the s_2 and s_3 positions. Comparison of the a_2 and a_1 locations yields a significant ($p < .05$) difference in all cases, as does the s_1 vs. s_2 comparison. The differences between the a_1 and s_1 positions are statistically negligible. These findings hold for each of the five indices employed, including the *FC* score.

While the results are in general agreement with the predictions, there is one clear and significant exception: *Both the a_1 and s_1 locations are the points of lowest predictability, with no difference between them.* The theoretical expectation was that the a_1 position would be the most redundant, at least more so than the s_1 position.

Discussion

Clearly, a most intriguing aspect of this research is contained in this last finding of no difference in predictability between words immediately preceding and immediately following a hesitation. Why this unexpected finding, in contradiction to Goldman-Eisler's (1958a) conclusions?

One possible answer is that the different results merely reflect differences in what was being compared. Goldman-Eisler restricted her study to unfilled pauses, identifying these as silences beyond .25 sec. The present study not only used a much more stringent criterion for the inclusion of unfilled pauses, but also incorporated these within a larger undifferentiated category involving other hesitations. Thus, unfilled pauses as such were relatively obscured here; hence, the present study should perhaps not be expected to yield the same findings as Goldman-Eisler.

Such considerations may explain the lack of an a_1-s_1 difference, but it does not account for the fact that a_1 represents a position of significantly lower predictability than other (e.g., a_2 and a_3) locations. A plausible explanation for both findings may be that other hesitation types behave differently than do unfilled pauses. Consider the possibility that for one or another hesitation type, an a_1-s_1 difference does in fact exist but in a direction opposite to that for unfilled pauses. Under such circumstances, the final results, since they included all hesitations within a single category, could represent a cancelling-out among several significant but opposing response tendencies in the a_1 and s_1 positions.

Such a hypothesis can readily be tested by comparing the world predictability scores for the a_1 and s_1 positions separately for each hesitation type. Unfortunately, the data from this investigation allowed for only a limited analysis on this point. After elimination of the many instances of mixed hesitations in order to conduct the analysis only on pure cases of each hesitation type, enough cases remained to represent only repeats ($N = 13$) and filled pauses ($N = 20$). But even from such a restricted comparison, important clues emerged.

The data presented in Table 3 clearly indicate a differential pattern of word predictability between the two types of hesitations. The filled pause category exhibits the results originally predicted—greater uncertainty in

Table 3

Mean Replacement Measures of Words Immediately Antecedent to and Subsequent to Repeats and Filled Pauses

	Filled pause		Repeat	
Measure	a_1	s_1	a_1	s_1
V	.42	.30	.37	.43
V_{lex}	.32	.23	.29	.39
FC	.72	.75	.58	.70
V/FC	.58	.40	.64	.62
Rel. H.	.40	.53	.51	.45

the s_1, as opposed to the a_1, position. This is true on all but the *FC* measures where the reversal is relatively small. Examination of the repeat category, however, indicates precisely the opposite trend; the a_1 location being the one of higher uncertainty (again, the one reversal, in *V/FC*, is of minimal magnitude). When the a_1-s_1 difference scores are compared between the two hesitation categories by a Mann-Whitney *U* test, the difference between filled pauses and repeats is significant ($p <$ .05) on all but the *FC* measure.

This evidence is admittedly quite sparse and tentative, but it does raise a legitimate theoretical question: what accounts for the differential distributions observed? One interpretation is that *different types of hesitations index different kinds of encoding decision points.* The fact that Maclay and Osgood (1959) found various hesitation types to be somewhat differentially distributed with respect to syntactic environment lends credence to this viewpoint (although their data do not lead to the particular prediction for repeats and filled pauses). Similarly, while the Goldman-Eisler (1961b) distinction between the role of filled and unfilled pauses does not relate directly to the differential choice-point hypothesis, it does suggest that different hesitation phenomena may differ in the kind of psycholinguistic behavior they characterize.

The theoretical position offered here does not necessarily assume that each hesitation type indexes a distinctively different type of encoding-decision juncture. It may be suggested that two basic types of psycholinguistic phenomena underlie the occurrence of hesitations in encoding. One is a type of *groping* phenomenon, similar to that suggested by Goldman-Eisler. The speaker reaches a point where, for one reason or another, he cannot immediately elicit the "right" word, phrase, or sentence. Accordingly, his speech is interrupted as he gropes for the appropriate term. Under such circumstances, we would expect that the next encoding unit would be the more difficult to replace in cloze procedure—the encoder paused because he was uncertain, and this degree of uncertainty would be reflected in the cloze responses. Both filled and unfilled pauses may particularly characterize such encoding blocks.

An alternate type of hesitation-producing behavior stemming from *feedback* during encoding can also be postulated. The speaker, hearing himself say what he may not intend, interrupts his message production. Under such conditions he may backtrack to correct himself or may be momentarily stunned and repeat himself. This approach implies, then, that such feedback-induced hesitations will be characterized more by false starts or repeats. Moreover, it would predict that the highest uncertainty would be not in the subsequent, but in the preceding word that precipitated the hesitation.

This idea of a contrast in the function of different hesitation phenomena is obviously still speculative at this stage. The theoretical position is more a consequence of the investigation than the study was a test of the theory. The implications, however, are amenable to further exploration—both through more detailed descriptive analysis of the occurrence of hesitations, and through experimental tests of specific hypotheses suggested by the theory.

References

Feldstein, S., and Jaffe, J. A note about speech disturbances and vocabulary diversity. *J. Communication*, 1962, **12**, 166-170.

Fillenbaum, S., and Jones, L. V. An application of "cloze" technique to the study of aphasic speech. *J. abnorm. soc. Psychol.*, 1962, **65**, 183-189.

Garner, W. R., and Carson, D. H. A multivariate solution of the redundancy of printed English. *Psychol. Rep.*, 1960, **6**, 123-141.

Goldman-Eisler, F. Speech-breathing activity—a measure of tension and affect during interviews. *British J. Psychol.*, 1955, **46**, 53-63.

Goldman-Eisler, F. The determinants of the rate of speech output and their mutual relations. *J. psychosom. Res.*, 1956, **1**, 137-143.

Goldman-Eisler, F. Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, 1958, **10**, 96-106. *a*

Goldman-Eisler, F. The predictability of words in context and the length of pauses in speech. *Language and Speech,* 1958, **1**, 226-231. *b*

GOLDMAN-EISLER, F. Hesitation and information in speech. In C. Cherry (Ed.), *Information theory —fourth London symposium.* Washington: Butterworths, 1961, Pp. 162-174. *a*

GOLDMAN-EISLER, F. A comparative study of two hesitation phenomena. *Language and speech,* 1961, **4**, 18-26. *b*

LOUNSBURY, F. G. Sequential psycholinguistics. In Osgood, C. E., and Sebeok, T. A. (Eds.), Psycholinguistics: a survey of theory and research problems. *J. abnorm. soc. Psychol.,* 1954, **49**, suppl., 93-125.

MACLAY, H., AND OSGOOD, C. E. Hesitation phenomena in spontaneous English speech. *Word,* 1959, **15** (1), 19-44.

MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.,* 1956, **53**, 1-15.

MAHL, G. F. Exploring emotional states by content analysis. In I. De Sola Pool (Ed.), *Trends in content analysis.* Urbana: Univer. Illinois Press, 1959, Pp. 89-130.

OSGOOD, C. E. The representional model and relevant research methods. In I. De Sola Pool (Ed.), *Trends in content analysis.* Urbana: Univer. Illinois Press, 1959. Pp. 33-88.

SHANNON, C. E. Prediction and entropy of printed English. *Bell System tech. J.,* 1951, **30**, 50-64.

STOLZ, W., AND TANNENBAUM, P. H. Effects of feedback on oral encoding behavior. *Language and Speech,* 1963, **6**, 218-228.

TAYLOR, W. L. "Cloze procedure": a new tool for measuring readability. *Journ. quart.,* 1953, **30**, 415-433.

TAYLOR, W. L. Application of "cloze" and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, Univer. of Illinois, 1954.

WILSON, K. The information theory approach. In Osgood, C. E., and Sebeok, T. A. (Eds.), Psycholinguistics: a survey of theory and research problems. *J. abnorm. soc. Psychol.,* 1954, **49**, suppl., 35-49.

(Received March 16, 1964)