JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 6, 903–909 (1967)

Hesitations in the Speaker's Production and Listener's Reproduction of Utterances[1]

JAMES G. MARTIN

Chico State College, Chico, California

Twenty-four college Ss (encoders) described TAT pictures in short utterances. Each was yoked unsystematically with one of 24 listener Ss (decoders) who heard his recorded utterances and attempted to reproduce them. Words were classified as content or function. While encoders and decoders yielded about the same proportion of content words (41%), encoders yielded a relatively higher proportion of repeats, unfilled pauses, and total hesitations before content words (which have greater uncertainty) than did decoders. Decoders placed relatively more of their hesitations at sentence breaks than did encoders. Apparently, while encoder pauses reflect uncertainty, decoder pauses tend more to mark grammatical boundaries. The selection of semantic–syntactic structure precedes selection of individual words during encoding but follows during decoding.

While there is no reason to imagine different models of the speaker's and listener's grammatical competence (Miller and Chomsky, 1963), the psychological processes mediating their language performance surely differ. The speaker begins with some unitary thought to be expressed by words unreeling linearly in time. While (at least part of) the structure of his utterance is determined before the utterance begins, some words in the structure chosen may yet remain to be selected. Then hesitations in his utterance mark choice points where the number of word alternatives is large (Goldman-Eisler, 1958; Maclay and Osgood, 1959). While these hesitations mark speaker uncertainty they have probably little utility for the listener. The listener's task is to take as input the left-to-right sequence of sounds and organize them perceptually, presumably on the basis of their grammatical and semantic structure.

[1] Supported in part by grant MH 10400, National Institute of Mental Health, U. S. Public Health Service.

We assumed in this experiment that the listener, given the task of reproducing utterances he hears, will reproduce them as he perceptually organized them. Thus, hesitations should differ in the speech produced by a speaker (hereafter the encoder) and reproduced by a listener (decoder). Encoder hesitations reflecting uncertainty should appear before content words (nouns, verbs, adjectives, and adverbs) rather than before function words to a greater degree than decoder hesitations. In the experiment encoders described Thematic Apperception Test (TAT) pictures in short utterances. Each encoder was yoked with one decoder who heard his utterances and attempted to reproduce them. It was predicted that encoders would yield a higher proportion of hesitations before content words than decoders.

METHOD

Encoding Procedure. The S who served as encoder was fitted with a microphone held by a wire neck-loop. His instructions were, in effect,

to describe in one sentence each TAT card presented to him. Two cards were shown for practice. The one-sentence instruction was not mentioned again, either in the practice session or later.

Next the 20 experimental TAT cards were presented to S at 15-sec intervals, and his utterances were taped magnetically on a portable Bell and Howell recorder. An utterance was taken to include the *total* interval between onset and end of vocal response to one card. Later the encoder's utterances were re-recorded on Ampex equipment to eliminate long stretches of silent interval *between* utterances, that is, the silent interval between the end of responding to one card and the onset of responding to the next. Typewritten transcripts were prepared.

Decoding Procedure. After the microphone had been placed on his shoulders, the S serving as decoder was instructed to listen to the speech presented by tape recorder and to repeat what he had heard upon signal from *E*. Then *E*, following along on the typewritten transcript, started the tape recorder, stopping it 3 sec after the end of each utterance, which cued *S* to attempt to reproduce it. When the decoder had finished, the next utterance was presented. The decoder utterances were transcribed also.

Subjects. There were 24 Ss who generated utterances (encoders). The recordings of eight additional Ss were unusable due to unintelligibility, apparatus failure, language handicaps, etc., and were not transcribed or used further. Each encoder was yoked with one decoder S. The pairing was haphazard; the first scheduled decoder was paired with the first scheduled encoder, etc. All 48 Ss were extra-credit volunteers from psychology classes.

Word Classification. The list of 363 function words drawn up by Miller, Newman and Friedman (1958) was used as the basis for word classification. Their list includes what are traditionally called articles, prepositions, pronouns, numbers, conjunctions, and auxiliary verbs, plus certain irregular forms, and was determined by starting with the function words listed by Fries (1952) and augmenting the list on the basis of their grammatical intuitions. The important characteristic of this classification for Miller *et al.* (and the present study as well) was that the function words were far more common than the content words in the printed text they analyzed; the 363 function word types (6.4%) accounted for 59% of the word tokens in their sample while 5180 different content words (93.6%) accounted for the remaining 41% of the words.

Analysis of Hesitations. Two scorers independently of each other and without any knowledge of (a) the purpose of the experiment, (b) the word classification, or (c) the encoder–decoder classification, listened to each utterance and marked hesitations on copies of the typewritten transcripts, alternating between encoders and decoders in the approximate order encoder 1, decoder 24, encoder 2, decoder 23, ..., etc. Following Maclay and Osgood (1959), hesitations were classified as Repeats, False Starts, Filled Pauses, and Unfilled Pauses. Briefly, Repeats were all nonsemantic repetitions, False Starts were incomplete or self-interrupted utterances, and Filled Pauses were occurrences of the common hesitation devices in English ɛ, æ, r, ə, m. Unfilled Pauses were hesitations between words judged to be abnormal for the speaker yielding the utterance.

To familiarize the scorers with their task the author collaborated with each in marking the utterances of one speaker. Then each scorer marked the remainder, listening to each utterance at least twice.

Results

Words. Table 1 shows that while encoders yielded more words than decoders, $F(1,23) = 229.58$, $p < .01$, and both produced more function than content words, $F(1,23) = 36.73$, $p < .01$, there was no interaction; the overall percentage (and mean percentage) of each kind of word was about the same for both groups. While these percentages cannot be compared directly to Maclay and Osgood's sample of spontaneously encoded speech since those investigators used Fries' classification, they are strikingly similar to those in the

TABLE 1

Mean Words (M) and Percent Words (P) Uttered in Each Classification

		Encoders	Decoders
Content	*M*	122.4	100.3
	P	40.8	39.4
Function	*M*	177.8	154.1
	P	59.2	60.6

TABLE 2

CORRELATIONS BETWEEN SCORERS

	Repeats	False starts	Filled pauses	Unfilled pauses
Encoder utterances	.93	.70	.89	.90
Decoder utterances	.84	.88	.93	.91

printed text analyzed by Miller *et al.* (1958), who found 41% content words in their sample. It is clear that differences in the distribution of content and function words cannot account for whatever differences between encoders and decoders are found in hesitation phenomena.

Scorer Agreement. Table 2 shows that most of the correlations between the two scorers for the total number of each kind of hesitation marked were fairly high. Scorer judgment does not seem to be an issue in this experiment; however, since scorers agreed on the direction of differences between encoders and decoders in the data reported below, their data were pooled.

Hesitations. The analysis and presentation of results in this section and the next follow the same form throughout. Each hesitation measure appears in a 2 × 2 classification of type of coder (encoder-decoder) vs. type of word following a hesitation (content-function). Hesitations are represented in the tables in terms of hesitations per word, e.g., hesitations before a type of word divided by opportunities to hesitate before that type of word (× 100 = percentages), over all Ss. The statistical tests, however, are based upon 2 × 2 repeated measures analyses of variance of the four proportions (arcsine transformation) for each pair of Ss, where both numerator and denominator in each proportion were obtained from the data of the S concerned.

Table 3 shows the percentages for each type of hesitation and for Total Hesitations. Note that for most types of hesitations, encoder hesitations occur relatively more often before content than function words. For analysis of these data, proportions were calculated for each S for each type of word, e.g., number of Filled Pauses before content words given by the first encoder S divided by the number of content words he uttered. Since some decoders gave few, if any, hesitations, only scores from pairs of Ss in which each S in the pair gave at least one hesitation of the type being analyzed were included in each analysis.[2] The *N*s in Table 3 refer to the number of S pairs included in each analysis.

The five 2 × 2 analyses of variance gave the following significant results. There were more Repeats per word before content than function words, $F(1,16) = 19.05$, $p < .01$. Encoders yielded more Unfilled Pauses than decoders, $F(1,22) = 76.91$, $p < .01$, and more Total Hesitations, $F(1,22) = 63.34$, $p < .01$.

Of greater interest were the interactions between type of coder and type of word. Encoders gave relatively more Repeats before content words than did decoders, $F(1,16) = 5.72$, $p < .05$, more Unfilled Pauses, $F(1,22) = 10.92$, $p < .01$, and more Total Hesitations, $F(1,22) = 16.29$, $p <$

[2] This procedure was followed since including all 24 pairs of Ss would seem to bias the results in favor of the experimental hypothesis when decoders give two zero scores. Statistical conclusions about the interactions between coder type and word type, however, remain unchanged whether or not all pairs are included.

TABLE 3

PERCENTAGES OF ENCODER (E) AND DECODER (D) WORDS PRECEDED BY HESITATIONS

	Repeats (N = 17)			False starts (N = 18)			Filled pauses (N = 11)			Unfilled pauses (N = 23)			Total hesitations (N = 23)		
	E	D	Both	E	D	Both	E	D	Both	E	D	Both	E	D	Both
Content words	1.43	0.75	1.12	0.73	0.69	0.71	1.21	0.62	0.94	11.01	4.49	8.05	14.38	6.54	11.06
Function words	0.52	0.43	0.48	1.22	0.92	1.08	1.16	1.04	1.10	8.57	4.83	6.85	11.53	7.22	9.52
Both	0.89	0.56	0.74	1.02	0.83	0.93	1.18	0.88	1.04	9.59	4.69	7.34	12.68	6.95	10.05

TABLE 4

PERCENTAGES OF ENCODER (E) AND DECODER (D) WORDS PRECEDED BY BREAKS AND HESITATIONS AT BREAKS

	Breaks per word (N = 17)			Hesitations at breaks per break (N = 7)			Hesitations at breaks per hesitation (N = 15)		
	E	D	Both	E	D	Both	E	D	Both
Content words	0.78	0.46	0.64	45.83	31.82	42.65	2.60	2.22	2.50
Function words	4.55	5.57	5.02	30.41	24.27	27.25	12.02	18.73	14.38
Both	3.01	3.55	3.26	32.11	24.65	28.39	7.66	12.60	9.23

.01. These shifts in the distribution of hesitations support the hypothesis that encoder and decoder hesitations reflect different psychological processes.

Hesitations at Sentence Breaks. Copies of the typewritten transcripts, unmarked for hesitations, were examined for breaks setting off complete sentences or independent clauses from the remainder of an utterance. Only segments of the utterances which were judged to be clearly a sentence or an independent clause were so marked. There were relatively few of these breaks for at least two reasons. First, most of the (spontaneous) encoder utterances and their successful decoder imitations were quite ungrammatical. Second, the encoders who followed instructions, which were to give only one sentence to each card, tended to give only a short sentence or sentence fragment to each; breaks in their protocols and likewise in the decoder reproductions were few in number. Four encoders and seven decoders yielded no breaks at all.

Since there were few hesitations at the breaks (about 9% of all hesitations) the four types of hesitation at the breaks were combined. Breaks and hesitations at breaks were analyzed in three ways: (a) breaks per word, e.g., breaks followed by a content word over number of content words; (b) hesitations at breaks per break, e.g., hesitations at breaks followed by a content word over number of breaks followed by a content word; and (c) hesitations at breaks per hesitation, e.g., hesitations at breaks followed by a content word over number of hesitations followed by a content word. These results are reported as percentages in Table 4.

The significant findings from the three 2×2 analyses were as follows. The left side of Table 4 shows that breaks marking sentences or clauses were more likely to be followed by a function word, $F(1,16) = 74.68$, $p < .01$. But, as the middle of Table 4 shows, breaks containing a hesitation were more likely to be followed by a content word, $F(1,6) = 12.99$, $p < .05$. More interestingly, note in the right side of Table 4 that relatively more of the decoders' hesitations in their utterances could be found in the sentence breaks. This difference, adjusted by covariance with breaks per word as the covariate since decoders gave relatively more breaks, was reliable, $F(1,13) = 6.18$, $p < .05$. These findings, while they apply only to the relatively few breaks marking sentences or independent clauses, are quite in line with the experimental hypothesis. They can be summarized as follows. (a) Function words are more likely to begin a sentence or independent clause. (b) Hesitations appearing at these breaks, however, are more likely to precede a content word. (c) While decoders compared

to encoders hesitate less under the conditions of this experiment, when they do they are more likely than encoders to hesitate at sentence breaks.

Discussion

Decoders were less likely to hesitate before content words when asked to reproduce encoder utterances than the encoders which they attempted to imitate. If some of the decoders regarded reproduction of hesitations as part of their task, the data indicate little success at it. Rather, the distribution of decoder hesitations suggests that they were more likely to mark the boundaries of grammatical units than to mark uncertainties in the selection of words. The conclusion that decoder hesitations mark grammatical boundaries while encoder hesitations mark uncertainty agrees with the results of other experiments. The findings of Goldman-Eisler (1958) and Maclay and Osgood (1959) using (encoded) spontaneous speech are consistent with ours in showing little correspondence between pausal phenomena and grammatical boundaries. At the same time many recent experiments have attempted to demonstrate in various ways the psychological reality of grammatical units (e.g., Fodor and Bever, 1965; Johnson, 1965; Martin, Davidson, and Williams, 1965; Martin and Jones, 1965). These latter experiments all tap decoding processes; Ss are required to learn or otherwise react to prescribed variations in grammatical inputs.

Implicit in the above experiments and explicit here is the assumption that the decoder's reproductions reflect his active perceptual organization of verbal inputs. The present data indicate that the assumption is likely. If the higher proportion of encoder hesitations before content words reflected merely less practice in uttering these words (which is certainly the case), then decoders as well would be expected to hesitate before the content words. The same argument supports another assumption of this experiment, that encoder hesitations identify unresolved word choices as well as grammatical boundaries. If encoder hesitations represented only the latter, the distribution of decoder hesitations should be about the same.

These data suggest a communication model in which words follow structure during encoding but precede structure during decoding. In encoding the intention of the speaker determines the semantic and syntactic structure (or part structure) of utterances before the individual words are chosen. Once the structure is determined, however, the number of alternatives to each word which could have been chosen determines the choice latency of each word which finally appears. Thus in the utterance *in the house*, a pause before *house* rather than before either of the other two words reflects a greater cognitive latency based upon the alternates to *house*. Given the intention to express the content of *in the x*, however, there are few if any selection alternatives in the language to *in* and *the*.

The decoder's comprehension, on the other hand, depends upon hearing sufficient words to provide the structure of the speaker's intention. These words are heard in a channel filled with noise, of which hesitations are only a part. The most efficient mechanism, then, for hearing speech would be one which is biased against the noisy elements of a speech signal.

References

Fodor, J. A., and Bever, T. G. The psychological reality of linguistic segments. *J. verb. Learn. verb. Behav.*, 1965, 4, 414–420.

Fries, C. C. *The structure of English:* New York: Harcourt Brace and World, 1952.

GOLDMAN-EISLER, F. Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, 1958, **10**, 96–106.

JOHNSON, N. F. The psychological reality of phrase-structure rules. *J. verb. Learn. verb. Behav.*, 1965, **4**, 469–475.

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

MARTIN, J. G., AND JONES, R. L. Size and structure of grammatical units in paired-associate learning at two age levels. *J. exp. Psychol.*, 1965, **70**, 407–411.

MARTIN, J. G., DAVIDSON, JUDY R., AND WILLIAMS, MYRNA L. Grammatical agreement and set in learning at two age levels. *J. exp. Psychol.*, 1965, **70**, 570–574.

MILLER, G. A., AND CHOMSKY, N. Finitary models of language users. In R. D. Luce, R. R. Bush, and E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2, New York: Wiley, 1963. Pp. 419–492.

MILLER, G. A., NEWMAN, E. B., AND FRIEDMAN, A. E. Length-frequency statistics for written English. *Information and Control*, 1958, **1**, 370–398.

(Received April 22, 1966)