Hesitation and Semantic Planning in Speech[1]

Brian Butterworth[2,3]

Samples of spontaneous speech were analyzed according to their distributions of phonations and silences. Some of these exhibited cyclic, or "rhythmic," patterns, in the sense defined by Goldman-Eisler. Transcripts of three such samples were subjected to a segmentation procedure carried out by independent judges utilizing a common semantic intuition. Points in the transcripts where agreement was high among the judges were found to correspond with the beginnings of temporal cycles, and agreed semantic segments coincided with sentence or clause boundaries and usually consisted of several clauses and more than one sentence. It is argued that a theory of speech generation must contain provision for semantic integration at the suprasentential level.

INTRODUCTION

The investigation of cognitive planning processes underlying the generation of spontaneous speech has recently focused attention on what might be called "the *macrostructure* of hesitation." Previous studies on the significance of hesitation have dealt with "microstructure"—the location and duration of individual pauses. These latter studies have revealed a relationship between some specifiable pauses and the process of lexical selection (Maclay and Osgood, 1959; Goldman-Eisler, 1958; Butterworth, 1972, 1973; Tannenbaum *et al.*, 1965). By taking not each individual pause, its duration and location, but the overall proportions of pausing in speech, it has been found that

---

[1] A shorter version of this paper was given at the British Psychological Society London Conference, December, 1972.
[2] University of Cambridge, Psychological Laboratory, Cambridge, England.
[3] The experiments reported here were carried out while the author was at the Psycholinguistics Research Unit, University College, London.

several factors are associated with variations in this proportion, and these reflect aspects of the cognitive process which are concerned with content rather than structure (Goldman-Eisler, 1961, 1968, Chap. 4; Lay and Paivio, 1969; Reynolds and Paivio, 1968).

Henderson *et al.* (1966) raised the question of whether content planning takes place sporadically throughout an utterance or at definable periods during it. They found that in samples of spontaneous speech hesitant periods predominantly of silence alternated with event periods predominantly of phonation.

This phased or cyclic speech frequently had the property that the amount of pausing in the hesitant phase (consisting mainly of silence) of a cycle was directly related to the amount of phonation in the succeeding fluent phase of that cycle. They termed speech with this relation "rhythmic" speech, and inferred from the *mathematical* dependency of fluent phase on hesitant phase a *psychological* dependency. They hypothesized that the amount of speech in the fluent phase *required* the planning time given by the pausing in the hesitant phase, and found that hesitant phases exhibited not only a greater proportion of pausing but also more hestiation phenomena of other kinds; for example, there were more filled pauses—"ah's" and "um's"—and the gaps in the hesitant phases were less likely to occur at grammatical junctures than they were in fluent phases. As Goldman-Eisler (1968, p. 83) stated, "These results show that at the time of uttering fluent speech the speaker's pausing is under control and that it is relatively well integrated into the syntactic structure, that pauses serve the function of communication rather than being symptomatic of internal processes." And, conversely, pauses in the hesitant phases are less under control, and reflect the ongoing planning processes.

A number of authors have suggested, not implausibly, that speakers plan out their utterances in terms of well-understood structural units, in particular the (surface structure) sentence (notably Wundt, 1912; Miller *et al.*, 1960). Boomer (1965), in a paper criticizing Goldman—Eisler's (1958) claim that hesitation pauses are used by the speaker to make difficult lexical selections, maintains that speakers plan their utterances in terms of "phonemic clauses" and that pauses are used to organize the phonemic clause immediately following them, and not to select the lexical item immediately following them.

Since it has been found that overall proportions of pause time are related to the semantic properties of spoken texts (Goldman-Eisler, 1961, 1968, Chap. 4), it seems likely that semantic units, which may or may not coincide with surface sentences or (phonemic) clauses, should be the initial area of investigation. An attempt operationally to capture semantic units

independently of the temporal properties of the spoken texts is described, and their relationship to rhythmic cycles is shown.

METHOD

Collection of Speech Samples

The speech samples used in this study were collected as part of a broader research program described in detail elsewhere (Butterworth, 1972). The original corpus comprised nearly 3½ hr of spontaneous speech of eight male subjects.

Subjects were native British English speakers (third-year undergraduates, graduates, professional academics, and a journalist), and were between 20 and 30 years old (mean 24 years). All the subjects were acquainted with the experimenter (the author) before the experiment.

The experimental sessions took place in a comfortable soundproof room which contained only the subject and the experimenter (the author), seated facing each other, and a microphone. The subjects were given a set of propositions on social and political topics and asked to select the one they most agreed with and the one they most disagreed with. Their task was "to make out the best case they could for the proposition chosen." It was pointed out to the subject that, although the experimenter was primarily interested in what the subject had to say, the situation would be very like an ordinary conversation, in that the experimenter would contribute to the ensuing discussion, as he felt appropriate, by questioning what the subject said, or by asking for further clarification, or by adding any other comments of his own. In each instance, the subject started the conversation with what he conceived was a defense of the chosen proposition.

The experimenter interjected as he would in a normal conversation, except that he exercised rather more restraint. When the experimenter judged that the subject had no more to say on the proposition, he asked the subject to begin on the second proposition, repeating that the subject should make out the best case he could for that proposition.

There were thus two conditions—an "agree" condition and a "disagree" condition—which were randomized, roughly half the subjects beginning with the "agree" condition.

A visual analogue of the taped conversation identifying periods of silence and phonation was prepared in a manner described by several authors (e.g., Goldman-Eisler, 1968; Henderson *et al.,* 1966; Butterworth, 1972).

Transcripts were also prepared, as was a version which matched each phonation with what was said in it.

The identification of rhythmic segments (a segment being an uninterrupted utterance by one speaker, bounded by the utterances of the other speaker or by the beginning or end of the condition) was carried out in the manner described by Henderson *et al.*:

> The successive pauses and speech durations were plotted sequentially with speech time along the abscissa and pause time along the ordinate.... The passages could be fitted overall with one straight line revealing a general speech/silence ratio characteristic of each subject. Within a passage the sequential temporal structure revealed an alternating sequence of periods with different speech/silence ratios. Periods in which relatively long pauses and short speech utterances occurred together alternated with periods in which relatively short pauses and long speech periods occurred together. These alternations were apparent as regular changes in the rate of acceleration in series of short straight lines which could be fitted to the overall slope, relatively steep slopes alternating with relatively shallow slopes. The points at which the changes in acceleration occurred were estimated by inspection, and the straight lines were fitted using the method of averages. (Henderson *et al.*, 1966, p. 208)

This procedure was further checked by asking independent judges to decide on changes of slope in the graph. Very substantial agreement on changes of slope was found in the checking procedures; in general, this means that only clear-cut changes of slope are included.

Twenty segments contained three or more complete cycles (hesitant phase plus fluent phase), and ten of these were rhythmic in the mathematical sense described by Henderson *et al.* (1966) and reported above.

Cycle times varied between 10.6 and 39.2 sec, with a mean of 18.0 sec (SD = 5.29). Significantly more phased and rhythmic speech occurred in the "agree" condition, and three of these rhythmic segments from different speakers were selected for further investigation. (These three samples were chosen because they represented a wide range of segment lengths—sample S2A2 was 201 words long, S10A2 was 654 words long, and S12A1 was 884 words long.)

Identifying Semantic Units

In identification of semantic units, the task was to segment texts according to the common, informalized intuitions about semantic structure used in making a précis, paragraphing written output, etc.

The problem was to find the appropriate instructions for the subjects, to find the word (or phrase) that came nearest to referring to these semantic

intuitions. "Idea" was the best approximation. It is not a technical word, and it is reasonably free from obvious structural implications. The most important advantage was that because it put no restriction on the size of the unit, it did not trivially predetermine the outcome. For example, it is permissible to say both that the word "democracy" expresses an idea and that the text of the play *Othello* expresses an idea, namely, jealousy.

Procedure

Typed transcripts of S2A2, S10A2, and S12A1 were prepared. They were in normal orthography with the usual punctuation. Apart from punctuation, no information about the temporal structure of the text was available to subjects. Each transcript was headed by the proposition which the speaker was arguing for. The subject was given a copy of a transcript. The instructions to him were

> This is a transcript of part of a conversation. In it one person is speaking uninterruptedly and arguing a case for the proposition at the head of the page. Please read through the whole text and then divide it into ideas. Mark the end of one idea and the beginning of the next by putting a short vertical line through the appropriate point in the text. You can divide the text into as many or as few ideas as you want. Do you have any questions?

If the subject asked the experimenter (the author) what an idea was, he was instructed to decide that for himself. Each subject was given one transcript, and each of the three transcripts was given to eight subjects. The subjects were all students at University College, London.

RESULTS

Quantitative

There was considerable individual variation in performance on the protocols. In text S2A2, 201 words long, the mean response by the eight subjects was to divide it up into 12 Ideas; but one subject made only five divisions and another as many as 24 (SD = 5.77)—a range of 19. For text S10A2, 654 words, the mean number of Idea divisions was 22.6 (SD = 11.87), ranging from eight to 46. And in text S12A1, 884 words, the mean was 22.4 Idea divisions (SD = 3.96), ranging from 16 to 30.

The total number of divisions in each text—i.e., the number of points in a text where at least one subject put a line—was 28 divisions for S2A2, 55 for
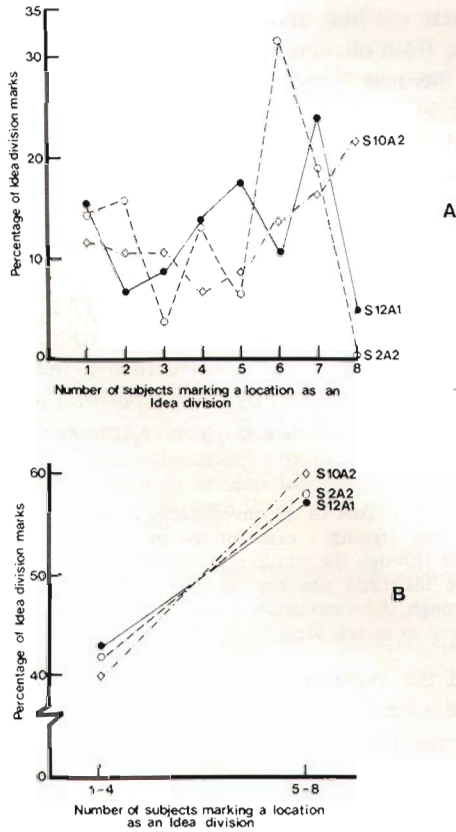
Fig. 1. Percentage of subjects marking a location as an Idea boundary (A) according to the eight categories of agreement and (B) according to grouping of the categories into agreement of fewer than half or agreement of more than half of the subjects.

S10A2, and 56 for S12A1. This gave a mean number of words per division as 7.1 for S2A2, 11.8 for S10A2, and 15.8 for S12A1.

Agreement on the location of subjects' Idea divisions was examined. Each division mark was classified according to how many division marks were at that location. ("Locations" were word boundaries—no subject put a division mark inside a word.) So a division would be classified as "8" if all eight subjects put a mark at that location, as "7" if seven out of eight subjects put a mark at that location, and so on down to "1." The percentage of marks in each category for the three texts is shown in Fig. 1A, and no clear pattern

can be discerned in it. However, if the eight categories are amalgamated into two categories—(1) marks at locations picked by less than half the subjects and (2) marks at locations picked by more than half the subjects—a consistent picture for the three texts is revealed (Fig. 1B). In each text, between 56 and 59% of all divisions fall in categories 5-8, i.e., locations picked by half or more of the subjects.

Since there was wide individual variation among subjects on the number of Idea divisions in a text, the criterion that more that half the subjects should agree on a location was used to define Idea boundaries within the texts.

These criterial divisions were found to coincide significantly with the previously identified temporal cycles in the text.

In text S2A2, there were three cycles and seven Ideas; the beginning of two cycles coincided with Idea boundaries. (Considering that cycle and Idea boundaries could occur between any two words in the text, the probability of this occurring was less than 0.01, calculated by means of a χ^2 test.) In text S10A2, there were 19 cycles and 13 boundaries. Four of these Ideas occurred at the beginning of text before phasing started; of the remaining nine, seven coincided with cycle starts ($p < 0.001$). In S12A2, there were 16 Idea boundaries and 21 cycles. Eight of the Idea boundaries coincided with the beginning of cycles ($p < 0.001$).

Ideas and Syntax

The location of Idea boundaries was also found to coincide with clause boundaries in the texts. In Table I, the locations in the texts are classified as to whether they are chosen by a subject as an Idea division. The vast majority of division locations chosen by subjects were clause boundaries. Of the 25

Table I. Clause Starts Chosen as Idea Divisions: Data from Three Texts[a]

	Clause start	Not clause start	Total
Chosen as an idea division by at least one subject	102	25	127
Chosen as an idea division by no subject	81	1528	1609
Total	183	1553	1736

[a]The locations preceding the first word and succeeding the last word in each text are discounted, so the total number of possible locations is one less than the total number of words for each text.

Table II. Starts of Temporal Cycles with Clause and Sentence Boundaries

	Texts			
	S2A2	S10A2	S12A1	Total
Cycle starts corresponding to sentence boundaries	1	10	11	22
Cycle starts corresponding to other clause starts	2	3	5	10
Other cycle starts	–	6	5	11
Total	3	19	21	43
Total number of sentences	11	19	26	56

division locations which were not also clause boundaries, only five were chosen by more than one subject, but one of these was chosen by more than two subjects. However, a substantial number of clause boundaries were not chosen as Idea divisions; for example, in S12A1 there were more clause boundary locations not chosen than chosen.

The criterial Idea divisions, those chosen by more than half the subjects, were all located at clause boundaries. And these tended to coincide with sentence boundaries marked by a period in the text. In S2A2, there were 11 sentences, six of which were Idea boundaries. In S10A2, all the 15 Idea boundaries were at one of 19 sentence boundaries. In S12A1, 13 of the 16 Idea boundaries were located at one of the 26 sentence boundaries.

The starts of cycles also tended to correspond to clause—in particular, sentence—boundaries (see Table II).

Qualitative

After they had completed the division of a text into Ideas, subjects were asked by the experimenter whether they used any principles in making these divisions, and if so what. Many of them replied that there were no particular principles used consistently in carrying out the task—they referred instead to "intuitions" or "vague ideas." One subject, indeed, said he "tried not to think of a principle." For those who said they had a principle or strategy, informal semantic concepts were generally a part of it. A typical remark was "I gutted the text for content. I tried to get the guts of it out." Several made specific

reference to the argumentative nature of the text: "I looked for points in the argument." "I tried to get points as in making a preçis." "Those propositions worthy for consideration in examining an argument."

Many subjects expressed difficulty with the hierarchical organization of the Idea units, and some used a bracketing convention where they thought that a part should be subsumed under a larger Idea, even though they were not asked to do so in the instructions. Several subjects remarked that they could have divided the text up in a different way—either by coalescing the divisions already made or by subdividing these divisions.

Many subjects claimed not to have used syntax or punctuation conciously in choosing where to divide the text. Others were aware of the syntax and mentioned trying to ignore it. Only two subjects said that syntax was a major determinant in their choices.

DISCUSSION

The temporal cycles described above are here clearly shown to be associated with both large semantic units and major syntactic—especially sentence—boundaries. This decisively refutes the claim by Jaffe et al. (1972) that these cycles are randomly generated. And it supports the thesis put forward by Goldman-Eisler and her colleagues (Goldman-Eisler, 1968; Henderson et al., 1966) that these cycles are not random, but associated with the planning of spontaneous speech.

Clause boundaries appear to be a necessary but not sufficient condition for the onset of both cycles and new Ideas, in that the vast majority of cycles and the Idea divisions given by any subject coincided with clause boundaries but a very substantial number of clause boundaries were not coincident with either cycles or Ideas. There was somewhat better match between sentences, Ideas, and cycles. Taking Idea and sentences first, of 35 criterial Ideas boundaries—i.e., where more than half the subjects agreed on the location of an Idea division—all but four coincided with sentence boundaries; this left 19 sentences which were not judged to constitute an Idea if taken separately. Thus, of clause types, the relevant kind for Ideas seems to be sentences; but Ideas may consist of more than one sentence.

With regard to cycles, about half coincided with sentence starts and three-fourths with all kinds of clause boundaries. This left cycles consisting of more than one sentence in some cases and of parts of sentences in about half the cases, either in conjunction with complete sentences or, in a tiny number of cases, just a part of a sentence. No systematic account of these exceptions could be discovered from these data alone.

The results presented here are consistent with the hypothesis that the cycles represent integral planning units for the speaker, and shed light on what these planning units consist of linguistically. First, the speaker tends to plan ahead in terms of well-understood linguistic units—namely clauses and sentences. Second, he appears to have the ability to chunk together several clauses or sentences as one superordinate planned structure integrated by some kind of semantic unity. Presumably, there will also be subordinate planning units, although, as reported above, only difficult lexical choices have been revealed as meriting sufficient of the speaker's planning time to cause a delay in his output. It would therefore seem that the subordinate units are highly automatic consequents of the longer-range planning evidenced by the temporal cycles. If this is correct, then serious qualifications are required of Boomer's thesis that the main unit of planning is the phonemic clause (Boomer, 1965; Boomer and Laver, 1968). If speakers do encode speech into phonemic clause units, then this will occur well down the hierarchy of encoding processes and will be a process of a quite different kind from the planning of cyclic segments. It will be much more like Hughlings Jackson's (1878) automatic processes than like the creative, "now-organizing" aspects of his duality. There are three different kinds of evidence for supposing this: First is the evidence here presented of superordinate planning processes. Second, insofar as pauses correspond to phonemic clause boundaries (and it has been shown above that they occur widely elsewhere as well), it is always possible that such pauses have a communicative function, since they will serve the listener in helping segment his input into linguistically relevant chunks. Third, there is a relatively small ensemble of phonemic clause types, and it is certain that each speaker will have used each type on thousands of occasions. Hence the production of each type will be a highly overlearned skill sequence. What makes the utterance of strings of phonemic clauses novel, if not unique, in the speaker's linguistic ontogeny are the exact words that constitute the phonemic clauses and the ongoing organization of this string of phonemic clauses into meaningful segments of a spoken text. And it is precisely these two processes—the choice of words and overall semantic planning—that require the hesitations characteristic of high-level creative, cognitive activity.

The wide variation among subjects in the experimental task, taken with their introspective reports that alternative segmentations of the texts were equally permissible, suggests that there are other intuitive semantic units besides those classed as Ideas—certainly smaller units, corresponding perhaps to the segments defined by the subject making the most divisions in the text, and perhaps larger units defined over larger texts. It is likely that some way can be found of organizing the smaller and larger units into a hierarchy and

that the intuitions underlying it can be captured by some semantic formalism. This is a field for further research. That Ideas should not coincide in every case with cycles may also be partly accountable in terms of the wide variation among subjects as to possible and appropriate semantic segmentations of a text, since, in a sense, the speaker can be counted as one such subject. He will be organizing his output text according to his own conception of useful segmentation, and it would be curious indeed if this segmentation should coincide exactly with the common segmentation drawn from a pool of subjects. If there are various possible segmentations, perhaps organized hierarchically, a subject will choose one most appropriate to his task, and the tasks of the spontaneous speaker and the experimental subject are manifestly different. The speaker has constraints of cognitive load, memory, selecting words, interesting the listener, and so on; the experimental subject has none of these constraints. Instead, he has to interpret the experimental instructions to segment a text—which he does in a variety of ways—and this the speaker does not have to do. If there are semantic segments, then they can be expected to coincide in the two tasks only some of the time.

The intimate relation between Idea and syntax boundaries should not be surprising. Pedagogically, a sentence is held to express a single idea; and one subject reported how difficult he found it "to get away from the traditional concept of the sentences as an idea."[4] And it is, in any case, reasonable to see grammar as providing the common framework in which ideas are expressed. It would be surprising if this were not so. Even in the abstract discipline of logic there are rules for the well-formedness of propositions.

Another way of accounting for the data is to suppose that cycles have an origin independent of the planning process: they may reflect underlying rhythms of a biological sort—a kind of natural "cognitive stride"—which planning processes use (rather than create). Planning may lock onto the first phase of the cycle and execution onto the second phase. One way of testing this hypothesis might be to see if cycle durations are constant, for an individual, across different speaking tasks, since it is more likely that biological rhythms will show invariance than the amount of planning required by speech texts. This hypothesis, however, would still leave to be explained why there is not a perfect match between cycles and Ideas, and why some speech shows no rhythm at all.

To sum up, the fact that temporal cycles tend to correspond to whole units of semantic and/or linguistic structure shows that they are not randomly

[4]Goldman-Eisler (1972) found that the distribution of pause durations at sentence boundaries, in both spontaneous speech and readings, was quite characteristic, and unlike distributions for other syntactic transition points.

generated; and it also lends support to the thesis of Goldman-Eisler and her colleagues that these cycles reflect the planning processes of the speaker. Their evidence was that, in "rhythmic" speech, the amount of speech in the fluent phase of the cycle was dependent, mathematically at least, on the amount of pausing in the preceding hesitant phase. They suggested that the dependence was not merely mathematical, but psychological—that an amount of speech time needs an amount of cognitive processing time. For this to be plausible, it needed to be shown that a cycle of planning and execution had a linguistic and semantic integrity, and evidence for this integrity has been presented here.

ACKNOWLEDGMENTS

REFERENCES

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Lang. Speech* 8:148-158.
Boomer, D. S., and Laver, J. D. M. (1968). Slips of the tongue. *Brit. J. Disorders Commun.* 3:2-11.
Butterworth, B. L. (1972). Semantic analyses of the phasing of fluency in spontaneous speech. Unpublished Ph.D. dissertation, University College, London.
Butterworth, B. L. (1973). Lexical choice, grammatical options and hesitations in speech. Manuscript in preparation.
Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quart. J. Exp. Psychol.* 10:96-106.
Goldman-Eisler, F. (1961). Hesitation and information in speech. In Cherry, C. (ed.), *Information Theory* (Fourth London Symposium, 1960), Butterworth, London.
Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*, Academic Press, London.
Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Lang. Speech* 15:103-113.
Henderson, A., Goldman-Eisler, F., and Skarbek, A. (1966). Sequential temporal patterns in spontaneous speech. *Lang. Speech* 9:207-216.
Jackson, H. J. (1878). On affections of speech from disease of the brain. Reprinted in *Selected Writings of Hughlings Jackson,* Vol. II (1958), Basic Books, New York, pp. 155-170.
Jaffe, J., Breskin, S., and Gerstman, L. J. (1972). Random generation of apparent speech rhythms. *Lang. Speech* 15:68-71.
Lay, C. H., and Paivio, A. (1969). The effects of task difficulty and anxiety on hesitation in speech. *Canad. J. Behav. Sci.* 1:25-37.

Maclay, H., and Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word* 15:19-44.

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the Structure of Behavior,* Holt, Rinehart and Winston, New York.

Reynolds, A., and Paivio, A. (1968). Cognitive and emotional determinants of speech. *Canad. J. Psychol.* 22:164-175.

Tannenbaum, P. H., Williams, F., and Hillier, C. S. (1965). Word predictability in the environments of hesitations. *J. Verb. Learn. Verb. Behav.* 4:134-140.

Wundt, W. (1912). Die Sprache. In Blumenthal, A. L. (ed.), *Language and Psychology* (1970), Wiley, New York, pp. 20-33.