was paralleled by differences in the way these two groups rated their experimenters. Masculine subjects rated their experimenters more positively than did the feminine subjects (as more "ambitious," "selfless," "unwary," "frank," and "clever"). It can be argued that in these ratings the feminine male subjects projected less positive self-concepts than the masculine male subjects. To the extent that anxiety is a correlate of negative self-concept, as has been found by Lipsitt (1958) and Mitchell (1959), the differences in postsession ratings of their experimenters made by feminine and masculine subjects support the explanation (presented above) of the performance results associated with sex typing of the subject.

The generality of the present findings is limited by the small number of experimenters employed. Nevertheless, the study suggests the importance of sex typing (or factors associated with it) as an influence on socially reinforced performance.

REFERENCES


(Received June 28, 1966)

THE PHONEMIC CLAUSE AS A UNIT OF SPEECH DECODING

ALLEN T. DITTMANN AND LYNN G. LLEWELLYN

Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland

It is argued here that spoken language is decoded by the listener in word groups called phonemic clauses. Behavioral data, the location of listener responses in relation to the speaker's phonemic clauses, provide the evidence. Pairs of Ss spoke to each other alternately on topics of mutual interest in a situation resembling that of a telephone conversation. The listeners were instructed to let the speakers know that they were listening and interested. Their responses were found to be located almost exclusively at the ends of the speakers' phonemic clauses rather than within them, whether or not there were hesitations at either point. Further, only those types of clause ending which sound most "final" were followed by listener responses.

In this paper we present indirect but strong circumstantial evidence that the phonemic clause is a basic unit which the listener uses in decoding speech. The unit was originally described by Trager and Smith (1951) as necessary to the linguistic analysis of spoken language, applicable to groups of speech sounds larger than the phoneme and morpheme. Boomer and Dittmann...
(1962) found that thresholds for recognizing hesitations are markedly affected by their locations in relation to phonemic clauses. In a further study of hesitation forms and phonemic clauses, Boomer (1965) argued that this unit is fundamental to the grammatical encoding process in speech.

Boomer hypothesized that hesitations are indicators of decision processes as the speaker formulates what he is going to say next and how he is going to say it. Boomer traced his hypothesis to a related one by Lounsberry (1954), who reasoned that hesitations would occur at points of high statistical uncertainty and thus mark the beginning of units of encoding. While Lounsberry did not specify the unit at which these points might be found—in fact he said, “units of any given order [p. 99]”—he did note that “The observations which lead us to formulate this hypothesis have been focused on the sequencing of words [p. 99].” He then went on to propose studies on word transition probabilities as the relevant testing ground for his hypothesis. Boomer used the phonemic clause rather than the word as the unit, 3 minutes of uninterrupted speech from each of 16 subjects as the corpus, and found that hesitations were located predominantly at the beginnings of clauses, where syntactic and lexical decisions may not yet have been made. The statistical uncertainty of early words in phonemic clauses, incidentally, is likely to be low.

This unit thus appeared to be established as a promising one for further work in speech encoding. For both logical and empirical reasons the next step was to look for evidence that the same unit is important in decoding as well: the logical reason is that encoding and decoding processes must be closely related if a communication system is to operate efficiently. Empirically, there are compelling results from experiments by Miller and his colleagues that units larger than the individual word are the functional base of speech perception; for an overview see Miller (1962). That work used sentences which were composed for specific experimental conditions instead of conversational speech, and the units were based on considerations derived from the transformational grammar of Chomsky (1957). A possible line of evidence from a more naturalistic setting became apparent as we were analyzing data for another experiment which used the phonemic clause as the unit of analysis. The protocols for that study were 15-minute interviews with college students in which one of us (ATD) served as the interviewer. These interviews were designed to be as unstressful for the subjects as we could make them, and the interviewer interposed many comments and “um-hum’s,” by which he hoped to keep the conversation going smoothly and show the subjects that he was attentive and interested in what they had to say. In the course of examining these protocols we noticed that the spontaneous vocal listening responses of the interviewer seemed to be inserted almost exclusively at the boundary points, called junctures, between the speakers’ phonemic clauses and almost never at any points within the clauses. Since only one listener’s behavior had been examined, a listener who was very much involved in phonemic clause analysis, we then turned to material representing two other interviewers. The same pattern emerged. In this preliminary work we studied a total of 675 phonemic clauses (which ranged from 2 to 10 words, with a mode of 5), in which the listener responded 240 times at junctures and 25 times at other points in the speech.

The phonemic clause, then, is a unit which influences speakers’ behaviors as Boomer had found, and listeners’ behaviors as well. Boomer related the speech behaviors he studied, and by implication the unit itself, to encoding in speech, as had Lounsberry and others before him, and we have in parallel way related the listener behaviors we have seen in our preliminary results to decoding. Before continuing on to any further data, we shall need to explain our conception of the decoding process, since only then will the use of listener responses as evidence make sense.

While speech is composed of sounds, only combinations of these sounds are ever heard in conversation, or words in the usual sense of that term. Words vary greatly in their information value, as an examination of any language will show: some recur so often and in so many different contexts that it is clear that they cannot convey much information, while others occur so seldom that their appearance marks something quite new. The very frequent words are usually referred to as function words, since they serve to relate the other words, termed lexical items, to one another. In any one conversation the function words will be the same as in any other conversation; they are needed no matter what is being discussed. The lexical items, on the other hand, vary according to the subject matter under discussion.

In an ingenious experiment Berry (1953) showed that these information differences are reflected in the way the words are said. Using recordings of telephone conversations, he marked the transcript to indicate the stress received by each word (strong, secondary, and unstressed), then prepared a word-frequency distribution of
the corpus. He found the familiar curve of decreasing frequency with rank for the unstressed words and a complementary reversed curve for the stressed words. Words of secondary stress were distributed in a flat, bell-shaped curve. Uncommon words, the high-information lexical items, received strong stress, while function words were unstressed. Those of medium information were sometimes stressed and sometimes not.

A phonemic clause must have one strongly stressed word, and it may have other words of secondary stress. Degree of stress is relative within each clause, so that until the entire clause has been uttered the listener cannot identify the word which will receive the primary stress, and therefore convey the greatest information. Thus it is not until the end of the clause that he can tell what the main lexical item of the clause has been, relate it to others he has heard, and listen for what may follow. In the view presented here, the words of the clause cannot be decoded until the main lexical item has been identified. Relating one to another as the series of clauses unfolds in an entire utterance is a more complicated matter, not being studied here, probably best described as "understanding" or "comprehension." 1

By now our conception of the decoding process was clear. The listener must wait until an entire phonemic clause has been uttered before he can discriminate the main lexical item from the other material in the group. Until he has made that discrimination he has not, in effect, heard what the speaker has said during that time. A major clue that a clause has come to an end lies in the rhythmic pattern of stress and pitch in the group of words and the juncture that marks its end. Our hypothesis was that listening responses will occur predominantly following junctures at the ends of phonemic clauses and seldom at any point within a clause. 2 Although analysis of the preliminary material lent strong support to this conception, it raised two specific ques-

1 In this line of reasoning the term decoding is restricted to one level of analysis or unit, the word. In spite of unclarity on some points, there is general agreement that decoding must occur with respect to smaller units as well, such as phonemes and morphemes, but that substituting meanings for sounds, or semantic decoding, begins at the level of the word (see Osgood, 1963, p. 742).

2 Vocalizations are not, of course, the only sort of listener response, though they have been capitalized upon in the work reported here. Other work is currently under way in our laboratory on one of the more visible listener responses, the head nod, and its relationship to the vocal response.

...tions which could only be settled by a more formal study:

1. A simpler explanation of our preliminary results might be that listeners, not wishing to interrupt the speaker, respond after pauses wherever they may occur and not selectively at junctures. According to Boomer's study cited above, about half of all pauses are juncture pauses. These are significantly longer than other pauses and might therefore be considered more conducive to polite response. The difference Boomer found (1.027 seconds for juncture pauses, .747 second for hesitation pauses) is not great enough to account in full for the nearly 10:1 difference in response locations which we had already found. Nevertheless, some more precise way of comparing these two alternative hypotheses was certainly called for.

2. Since we now had a hypothesis firmly in mind, our subsequent work of locating junctures in speech was subject to a strong contaminating influence. In spite of the high reliability of this procedure (see Dittmann & Wynn, 1961) all previous work—including that of the reliability study itself—had been done before hypotheses had been formulated. Now the person locating junctures in a speaker's material had the additional cues of the listener's responses, which are recorded on the same tape, thus opening the way for spurious confirmation of the hypothesis.

The form of the data for the present study was therefore dictated by these considerations: it must include a record of time so that pause length could be measured, and the sound signal from the speaker and the listener must be recorded independently so that junctures could be located without contamination.

While listening to the material in our preliminary data, another idea emerged. There appeared to be systematic selection of the junctures at which the responses occurred. It should be recalled that there were listener responses to about one-third of the junctures we examined. It was our impression that these were more likely to be the most "final" sounding of the three types of juncture. The protocols studied so far had been coded only for presence or absence of junctures, not for type of juncture also, so more completely coded material was needed to check this impression. There was a possibility, too, of interaction with the alternative hypothesis that listeners respond at pauses, since the type of juncture most often responded to might also have longer pauses associated with it. Thus it was doubly important that the data for the more formal study be coded for type of juncture as well as for their presence or absence.
A related question had to do with type of listener response. We reasoned that brief vocal sounds like "um-hmm" might be functionally different from more extended comments and questions and give more information about the nature of the decoding process. More time might be required to formulate an extended response, so that these would more likely be found after longer pauses. In addition, the more final-sounding junctures might elicit comments, the others being followed more often by brief vocal sounds. Thus we wished to separate these listener responses in our analysis so that we would be able to make the necessary comparisons with pause length and juncture type.

**METHOD**

**Subjects**

The subjects were six college-student volunteers, four males and two females who were serving as normal controls during a 3-month stay at the Clinical Center of the National Institutes of Health. All subjects were acquainted with one another before the experiment.

**Procedure**

The experimental conditions were designed to simulate an everyday two-person telephone conversation. Subjects were seen in pairs and told that the experiment had to do with how people talk and how they listen. Each subject was instructed to talk to his partner for approximately 2 minutes, telling him about something that had happened during the course of the day, about a movie he had seen, or about any of a number of similar topics. The person listening was asked to respond as he ordinarily would over the telephone, letting his partner know that he was listening by any means he could, asking questions if a point needed clarifying, or in general commenting on the statements made by the speaker. The only restriction was that one person should carry the major portion of the conversation. The subjects were told that at the end of about 2 minutes, they would be asked to reverse roles, the listener becoming the talker, and vice versa.

After receiving instructions the subjects were taken to separate rooms, given hand microphones and sets of headphones, and allowed to familiarize themselves with the equipment. One investigator stayed in one room with one of the subjects, but out of his sight, and designated a time when the conversation was to begin. In an adjoining room with the second subject the other investigator, also out of that subject’s sight, recorded the conversation. The speech of the two subjects was recorded on separate channels of a two-channel tape recorder. The subjects’ headphones were connected in parallel to both channels so that each could hear his own as well as the other’s voice during the session. An oscillograph with chart paper running at .75 inch per second provided a graphic record of the talking sequence which was being recorded simultaneously on the tape. At the conclusion of the conversation the subjects were told the reasoning behind the experiment and the nature of the hypotheses involved.

To control for possible experimenter bias the investigator (ATD) who coded junctures (a procedure described in detail below) neither heard the complete tape recording nor saw the complete typescript of the conversation. He worked with an edited version of the typescript which included only the speech of the primary speaker and listened only to that channel of the tape. The juncture-coded material was then compared with a revised transcript which had been carefully corrected to ensure the inclusion of all nonfluent as well as fluent speech. The revised transcript also contained the exact location of any listener responses during the primary speech segment.

Responses made by the listener during the course of the conversation were classified as either words or brief vocal sounds. The former can be considered a sequence of several words which serve some purpose for the listener like seeking clarification; that is, “How fast do they [machines] go?” The latter, according to Fries (1952), are means whereby the listener.

... in some inconspicuous but conventional way, gives the speaker signals of this continued attention. In telephone conversations these signals consist of brief oral sounds interjected at irregular intervals but not interrupting the speaker’s span of talk. These brief oral sounds are not predictable, but, in a telephone conversation, if such a sequence of utterances occurs without oral signals of attention on the part of the hearer, the speaker usually interrupts with such questions as do you hear me or are you (still) there [p. 49].

Examples of this type of response would include utterances like “yes,” “um-hmm,” “yeah,” “I see,” “good,” “oh,” and so on, in order of their frequency in Fries’ telephone-conversation data.

The oscillograph strip chart was used to locate and measure all pauses during the conversation. Pauses were designated as either juncture pauses (any pause which follows a juncture) or as hesitation pauses (any pause other than a juncture pause).

**Linguistic Variables**

Phonemic clause, as used in this study, is the unit of spoken language set forth by Traugott and Smith (1951). Identifying these clauses involves paying attention to what are called suprasegmental phonemes. Segmental phonemes are the sounds which correspond roughly to the letters of the alphabet as language is written in conventional orthography. The suprasegmentals are those of pitch, stress, and juncture, which are overlaid upon the segmentals to give connected speech its rhythmic characteristic. For the
present purposes we shall discuss only stress and terminal juncture, since they are sufficient to define the phonemic clause; pitch contours are necessary to pronunciation, but vary according to other considerations not germane here. As strings of words are pronounced, certain syllables are stressed because of the words in which they occur, and some are given additional stress because of the syntactic structure of the phrase. Periodically a syllable receives the strongest of the stresses, called the primary stress. Between primary stressed syllables others will be heard to receive various degrees of weaker stress, weaker in comparison to that of the primary stressed word in that clause. Also somewhere between the primary stressed syllables will be heard another phenomenon, the terminal juncture. This marks the end of one phonemic clause and is usually found to follow very closely after the primary stress, sometimes immediately after it, sometimes within one (or less often), two, or more syllables. In the ongoing flow of speech, then, primary stresses and terminal junctures appear alternately and serve to demarcate successive phonemic clauses.

In their original work Trager and Smith (1961, p. 40) defined terminal juncture in terms of slight changes of the pitch phoneme at the very end of the final pitch level of the clause. Accordingly they identified three terminal junctures, one in which the pitch is sustained at the very end of the final pitch of the clause, one in which the pitch rises slightly at the very end, and one in which the pitch falls slightly at the very end. In a later paper Trager (1962) noted that terminal junctures have an additional characteristic in English, that of stretching or lengthening the time required to pronounce the final syllables of the clause. The speaker prepares for the juncture, as it were, by slowing down just slightly as he is coming to it. Trager cites some instrumental work by Martin Joos, referred to in a footnote by Hill (1958, p. 24), which shows the effects of this stretching: over the syllables beginning with the primary stressed syllable and ending with the juncture, the average length of one segmental phoneme is added for the sustained juncture, two phonemes for the rising juncture, and three or more phonemes for the falling juncture. Joos' work was not a formal study based upon a large sample, so the numbers should not be taken literally, but the fact of stretching was demonstrated, and the ordering is a useful guide, when added to the final pitch changes, in coding the three types of terminal junctures.

Phonologically defined in this way by primary stress and terminal juncture, the phonemic clause is a unit of the rhythmical patterning of spoken language. The relationship of the unit to the syntax of the language is undoubtedly an intimate one, but is not well understood at this time. The difficulties of conceptualizing the relationship have to do largely with the distinction between the underlying structure of any language and the performance of individuals as they speak the language. The underlying structure dictates both syntactic and rhythmic patterns. In the ideal form there is probably only one combination of these patterns to convey a given message, but in actual performance speakers do not and cannot always coordinate the two perfectly. Sometimes unit boundaries will coincide and sometimes they will not in conversational speech.

Some evidence bearing on this view comes from data which we have collected on coding junctures from the unpunctuated typescript of a 1,000-word speech sample where the coders did not have the sound of the speaker available to them. There were three coders, one of whom knew about phonemic clauses, but had no direct experience coding them; the other two were completely naive. The coders were told that speech, as they had probably noticed, occurs in rhythmical groups of words, like punctuation in writing, and were asked to mark the typescript wherever they thought such a group ended. They reported that they found it most meaningful to read the typescript silently and imagine how the speaker would have grouped the words. Their coding was compared with that of an experienced coder who worked from the audio tape of the passage. More junctures were coded from the sound version of the material than the typescript version: 143 from sound, 118 by one or more coders from transcript. All three coders agreed on 90 of these, or 63% of the junctures identified from the tape. The three coders agreed on only 6 more junctures which were not noted by the coder who heard the speech. Thus 94% of junctures agreed upon from the typescript were also heard in the speech, while only 63% of those heard were consistently picked up from the typescript alone.

This test is not a complete comparison of the relationship between syntactic and rhythmic structures in spoken language. The coders who worked from the typescript were not instructed to make a syntactic analysis of the material, nor were they professionally trained to do so. On the other hand, the typescript gave many clues about performance characteristics other than syntax. Filled pauses, retraces, anacoloutha, and many articulatory difficulties were included in this carefully prepared transcription and should have enabled the coders to guess accurately what the rhythm patterns were. The results obtained from the test indicate that speech rhythm is more than a matter of the words used and their syntactic combinations.

Evaluating the reliability of juncture coding is beset by a number of problems. An earlier report (Dittmann & Wynne, 1961, p. 203) presented percentages of agreement where coding was done by a group on two occasions 6 months apart. Deciding whether to code a juncture or not yielded 93% agreement, while deciding which juncture to code where one was coded both times was somewhat lower, 87%. The main problem lies in the first of these figures. When a juncture has been coded both times the coders have obviously agreed, but can we say they have agreed when no juncture has been coded? If we assume that there is an opportunity
for a juncture after every word, we can set up a contingency table and test obtained frequencies against those predicted by the marginal totals. We did this in the earlier study, and the resulting $x^2$ was 304.7 ($\phi = .83$). But this figure is spuriously high, first because we allowed "word" to be defined by the typist as she pressed the space bar, and second because not all words could be followed by a juncture in normal speech, an article which precedes a noun being the most obvious example. If we consider the number of juncures agreed upon on both occasions compared with the total number coded either time, the agreement is 79%, but no statistical test is possible with these figures. In the absence of more precise knowledge about the probability of juncures following different words or groups of words, our estimate of reliability will have to rest here. The other reliability figure, of which juncture to code when there is agreement to code a juncture, can be evaluated more easily: $x^2 = 89.5$ ($\phi' = .63$).

Our method of coding juncures in the present study, we feel, is an improvement over the old method; now we play the tape of the material at three-quarters of its normal speed. The recordings are made at 7.5 inches per second; then for playback a sleeve is placed over the capstan of the tape recorder to increase its circumference by a factor of 1.5, and the machine is run at its "3.75" inches per second speed. In this way the material can be heard much more clearly, especially the stretching phenomenon so important to accurate coding of terminal juncures.

**RESULTS**

The main hypothesis, that listeners will respond after juncures which terminate phonemic clauses, but rarely respond within clauses, was clearly upheld by the data. A total of 379 phonemic clauses made up the output of the six speakers, and there were 124 listener responses. Of these responses, 110 followed juncures, and only 14 occurred within clauses, a ratio closely resembling that obtained in the preliminary data which led up to the study. The pattern of responding after juncures was followed consistently by all subjects, with one partial exception: 8 of the 14 nonjuncture responses were contributed by one of the subjects, making his ratio of juncture to nonjuncture responses a little over 3:1.

While the main hypothesis seemed firm so far, the results did not completely rule out the alternative hypothesis that listeners respond to pauses as such and not to juncures alone. To test this idea further, frequency of responses at juncture pauses was compared with that of responses at hesitation pauses. The results are presented in Table 1, which shows that responses are located almost exclusively at juncture pauses. Separate analyses for the six individual subjects yielded similar results: chi-squares were significant at the .05 level or beyond for all subjects and consistently in the direction of more responses at juncture pauses.

<table>
<thead>
<tr>
<th></th>
<th>Listening response</th>
<th>No listening response</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juncture pauses</td>
<td>97</td>
<td>98</td>
<td>195</td>
</tr>
<tr>
<td>Hesitation pauses</td>
<td>4</td>
<td>119</td>
<td>123</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>217</td>
<td>318</td>
</tr>
</tbody>
</table>

Boomer's finding that juncture pauses were longer than hesitation pauses was considered next as a possible contributing factor to the overall significance of the results. Based solely on the ratio of juncture pause to hesitation pause lengths which Boomer found, approximately 10 responses to juncures would be expected for every 75 responses to hesitations. In our data, however, no significant difference was found between juncture and hesitation pause lengths, even taking into account the correlation between the two across subjects. It must be noted that our results cannot be compared directly with Boomer's. The differences can probably be attributed to differences in experimental procedure, which will be discussed below, but pause length, in any case, does not seem to be a major contributor to our results.

Earlier we mentioned that a selective factor might be operating with regard to listener response to juncures of different types: those which sounded the most final seemed more often to elicit some form of response. To examine this idea, we compared frequencies of response to the different juncture types. We found that rising and falling types of juncture were responded to significantly more often than sustained juncures ($x^2 = 68.06$, $df = 2$; $\phi' = .424$). No significant difference was found between the response frequencies to the rising and falling types of juncture. Thus the sustained juncture which is the least final sounding of the three types of terminal juncture is contrasted with the other two in rate of listener response.

Next we compared the three types of juncture for the length of pause likely to follow. For this analysis only those juncures which were not responded to could be included, since the listener's response interrupts the pause, and there is no way of knowing how long it might have been if
BRIEF ARTICLES

TABLE 2
RESPONSES TO DIFFERENT PAUSE LENGTHS OF THREE JUNCTURE TYPES

<table>
<thead>
<tr>
<th>Pause length</th>
<th>Sustained juncture</th>
<th>Rising juncture</th>
<th>Falling juncture</th>
<th>Total</th>
<th>Grand total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Response</td>
<td>No response</td>
<td>Response</td>
<td>No response</td>
<td>Response</td>
</tr>
<tr>
<td>None</td>
<td>2</td>
<td>119</td>
<td>1</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>Short</td>
<td>3</td>
<td>13</td>
<td>10</td>
<td>7</td>
<td>39</td>
</tr>
<tr>
<td>Long</td>
<td>2</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>138</td>
<td>20</td>
<td>33</td>
<td>83</td>
</tr>
</tbody>
</table>

the response had not intervened. Because of the combination of markedly different sample sizes and skewed distributions, the distribution of pause length was dichotomized to yield similar marginal totals (no pause, short and long pauses) and frequency statistics were used. The three juncture types were significantly different on pause length \((x^2 = 73.03, df = 4; \phi = .368)\). Table 2 shows the distributions: only 14% of the sustained junctures were followed by a measurable pause, as compared with 45% of the rising and 65% of the falling types; where pauses did occur, there was a positive relationship between pause length and juncture type, the sustained juncture being associated with short pauses, and the falling juncture with long pauses.

The results of the data thus far indicate that a possible interaction might have occurred between juncture type, pause length, and listener response, since the most final-sounding juncture was responded to more often, was followed by a pause more often, and was associated with longer pauses. But for the reason noted above, namely, that measurement of pause length is interfered with by responses, it is impossible to include all three variables in the same analysis. Since we recognized that this problem made interpretation of the results difficult, however, we did perform the analysis for whatever trends it might show. The data are presented in Table 2. Information transmission analysis (McGill, 1954) showed that both pause length and juncture type were significantly related to listener response, accounting for 18% and 21% of the total uncertainty, respectively, and corroborating our earlier findings. The interaction term was negative, a possibility in this type of analysis. Using this interaction to hold juncture type constant, the relationship between pause length and listener response is not so pronounced, accounting for only 10% of the total uncertainty, but it is still statistically significant. The combination of difficulty in measuring pause length where there is a response and the low frequencies in some of the cells of our table makes it impossible to be more precise in our conclusions.

Our attention next focused on mode of response: would the listener in a conversation be likely to react with brief vocal sounds or more extended comments and questions, depending upon pause length? Our data do not indicate any such trend. A t test for correlated measures using mean pause length before each mode of response did not reach significance. The other relationship to be examined was that between mode of listener response and type of terminal juncture used. A chi-square test was not significant, indicating that simple responses and longer strings of words were as likely to follow sustained junctures as either of the other types. A factor which might temper the obtained negative results on mode of response was a scoring convention: subjects often uttered a response then, after a pause, followed it with a further comment or question. These instances were counted as brief vocal sounds, since that was the immediate response. A much larger body of data would be required to take this scoring problem into account.

DISCUSSION

The results of this investigation strongly support the hypothesis that listeners make use of groups of words as units in decoding speech, groups of the size of the phonemic clause. In addition, the results tell us something more than this: people not only respond to the rhythm of alternating primary stresses and terminal junctures in the speech they are listening to, they also discriminate among different types of junctures, and we know this because they respond differentially to them. Even knowing this, however, we have still not accounted for all the data we have collected. Somewhat fewer than half of the more final-sounding junctures were followed by responses. Why not all of them, or a large majority of them? The subjects were instructed, after all, to respond often. The answer is, of course, that listeners have much more information about
when to intervene than simply the rhythm of the speech they are listening to. The content provides many cues, but there are too many possibilities that simple cues may not be enough. Syntactic sequences provide many other cues. As we noted earlier, syntax and rhythm patterns are basic structural members underlying the rest of a language and may follow each other exactly in ideally fluent speech. It is our impression, however, based on the close listening required for coding junctures, that there is a good deal of slippage between syntax and performance during conversational speech. Talking is a commonplace activity, but tremendously complicated, and the decision points are many: “Should I put it this way or that way?” “Should I continue with this line of thought at all?” “After this long, drawn-out subject phrase, should the verb be singular or plural?” “Maybe I had better qualify what I just said.” “How did I get mixed up with this conditional mode?” “Do I really believe what I am about to say?” These and all the hundreds of other internal questions one has while talking make fluent, “grammatical” speech very difficult. They also make syntactic analysis of ongoing speech just as difficult and lead listeners to wait until they “have really understood” before they respond.

Earlier we mentioned that 8 of the 14 non-juncture responses in our results were contributed by one individual. It is interesting to note that at least 4 of these responses could be identified as a form of verbal onephasmanship commonly played by adolescents. This game can best be illustrated by the following sequence: the speaker would say something like, “Do you know . . . .” and then immediately be interrupted by this listener with, “No, I don’t.” The effect of this playfulness, of course, is to dilute the obtained association between terminal junctures and listener responses.

One difference was found between our results and those reported by Boomer (1965), a difference which calls for explanation. He reports (p. 155) that juncture pauses were significantly longer in his subjects’ speech than hesitation pauses and, further, that this direction of difference obtained in 15 of his 16 subjects. We, on the other hand, found no significant difference in length between hesitation and juncture pauses, and the direction was evenly divided among our six subjects. In view of the consistency of Boomer’s finding, this difference in results does not seem to be a matter of chance variation between the two samples. We believe that the explanation may be found in the comparison of the situations which the two experiments used to obtain the speech samples. In Boomer’s study each subject was asked to speak extemporaneously on any topic for 3 minutes, and he was told that the experimenter would not intervene during that time (see Boomer & Dittmann, 1964, p. 325). The experimenter was seated in view of the subject, and while he did not speak during the 3 minutes, the subject could see that he was attending, and he may well have given some visible cues of his interest. In the present experiment the subjects were in separate rooms, and the listeners were instructed to let the speakers know that they were listening, as in a telephone conversation. As a result there were many interventions, and these, as we have seen, came mainly at junctures. Juncture pauses were thus often terminated before they might have been in a soliloquy, and those which were could not be included in our comparisons of the lengths of the two types of pauses. Thus juncture pause lengths appear shorter in our data than in Boomer’s, being about the same length as the hesitation pauses in his data.

REFERENCES


(Received June 30, 1966)