COGNITIVE PSYCHOLOGY **15**, 411–458 (1983)

Performance Structures: A Psycholinguistic and Linguistic Appraisal

JAMES PAUL GEE

Boston University

AND

FRANÇOIS GROSJEAN

Northeastern University

Two lines of research—one in psycholinguistics and one in linguistics—are combined to deal with a long-standing problem in both fields: why the "performance structures" of sentences (structures based on experimental data, such as pausing and parsing values) are not fully accountable for by linguistic theories of phrase structure. Two psycholinguistic algorithms that have been used to predict these structures are described and their limitations are examined. A third algorithm, based on the prosodic structures of sentences is then proposed and shown to be a far better predictor of performance structures. It is argued that the experimental data reflect aspects of the linguistic cognitive capacity, and that, in turn, linguistic theory can offer an illuminating account of the data. The prosodic model is shown to have a wider domain of application than temporal organization per se, accounting for parsing judgments as well as pausing performance, and reflecting aspects of syntactic and semantic structure as well as purely prosodic structure. Finally, the algorithm is discussed in light of language processing.

Two lines of research—one in psycholinguistics and one in linguistics—have recently been examining the temporal organization of sentences. On the one hand, psycholinguists have studied the "performance structures" obtained from experimental data such as pause durations, transitional error probabilities, and parsing values, and have attempted to predict

The authors are co-equal and their names are simply listed in alphabetical order. This study was conducted while James Paul Gee was a visiting research associate in the Department of Psychology, Northeastern University. The study was supported in part by grants from the Department of Health and Human Services (RR 07143 and NS 14923). We are most grateful to Judy Shepard-Kegl, not only for her comments and suggestions, but for introducing the authors to one another and hence for being at the "source" of this paper. We also thank Bill Cooper, Harlan Lane, Leah Larkey, and Joanne Miller for their useful comments on an earlier draft of this paper, and Peter Eimas and two anonymous reviewers for *Cognitive Psychology* for helpful suggestions on a later draft. Requests for reprints should be sent to James Paul Gee, Applied Psycholinguistics, School of Education, Boston University, 605 Commonwealth Avenue, Boston, MA 02215.

0010-0285/83 $7.50

these structures by means of algorithms which take into account the surface structure of the sentence and the need of the speaker to bisect the output (Grosjean, Grosjean, & Lane, 1979; Grosjean, Lane, Battison, & Teuber, 1981; Dommergues & Grosjean, 1981 as well as Cooper & Paccia-Cooper, 1980, for a different approach to the same problem). On the other hand, linguists have studied the prosodic structure of sentences based on standard methods of argumentation in the development of competence theories. The prosodic theories they have developed are based on hierarchical structures of weak and strong relations defined on pairs of units at each level of linguistic structure from the syllable, through the word and phrase, to the utterance as a whole (Liberman, Note 1; Liberman & Prince, 1977; Selkirk, 1980, Note 2, forthcoming).

In the following discussion we wish to combine these two lines of research. We will first describe "performance structures" and point out their main properties. We will then examine two psycholinguistic algorithms that can be used to predict these structures and point out their limitations. And we will finally propose an algorithm based on prosodic structures relevant to linguistic competence, and show how it is a far better predictor of these performance structures. In addition, this algorithm has the quality of containing certain characteristics that are important for language processing.

1. What are Performance Structures?

With the advent of generative grammar (Chomsky 1957, 1965), psychologists attempted to show the psychological reality of such linguistic notions as the surface structure of a sentence, its deep structure, and the transformations that link the two. To study the reality of surface structures, for instance, they used such tasks as recall (N. Johnson, 1965, Suci, Ammon, & Gamlin, 1967), relatedness judgments (Levelt, Note 3), and pausing (Brown & Miron, 1971; Goldman-Eisler, 1972). Most of these studies authenticated the role of major syntactic units in the processing of language and concluded that the speaker and listener's segmentation of speech is related—to some extent at least—to the structural description of that stream.

More recently, other researchers have been interested not so much in showing the psychological reality of the surface structure of the sentence but in examining more closely the exact relationship that exists between the structures obtained from experimental data and those proposed by linguists. They noted that earlier studies did not always find a perfect correspondence between the experimental data and the structures of linguistic theory, that these studies had used very simple sentences (usually monoclausal) and that the sentences had been balanced (the major constituents, such as the NP and VP, had been of about equal length). Thus

Martin (1970), for example, investigated the mismatches that occurred between parsing data and linguistic theory. He asked subjects to parse sentences by arranging the words of the sentences into "natural groups." The data thus obtained were then hierarchically structured by means of S. Johnson's (1967) clustering program. The results showed that subjects did not automatically group the verb with the NP objects, as linguistic models would predict, but that in many cases (SV)O clusterings were obtained. (See also studies by Suci, 1967; Hillinger, James, Zell, & Prato, 1976.)

In the last few years psycholinguists have investigated the sentence structures obtained from experimental data; they term these structures "performance structures." In a first study, Grosjean et al. (1979) asked subjects to read 14 sentences at five different rates. The pausing values they obtained were averaged and the means, expressed as a percent of the total pause duration in each sentence, were used to make hierarchical representations of the sentences. To do this, the following iterative procedure was employed. First, find the shortest pause in the sentence. Second, cluster the two elements (words or clusters) separated by that pause by linking them to a common node, and delete the pause. (If three or more adjacent words are separated from each other by the same value, make one cluster of these words: trinary, quaternary, etc.). Finally, repeat the process until all pauses have been deleted. Such a structure, taken from Grosjean et al. (1979), is presented in Fig. 1 (top tree).

Other experimental paradigms have been used to obtain performance structures, and generally these structures have been found to be relatively invariant across tasks. For example, Grosjean et al. (1979) asked a different group of subjects to parse the same 14 sentences and found a mean coefficient of correlation of .92 between the pausing and parsing values. Dommergues and Grosjean (1981) obtained performance structures by means of transitional error probabilities (TEPs) in a recall task (N. Johnson, 1965) and found a high correlation between TEPs and parsing values: .87 for balanced sentences and .83 for unbalanced sentences. It appears therefore that the subjective sentence organization that the speaker–hearer imposes on the sentence is relatively invariant across experimental tasks. Of course, there are also specificities linked to each task and these must not be overlooked. In Fig. 1, we can compare the performance structures of a sentence obtained with two different tasks, pausing and parsing. Although the similarity between the two is striking, we should note that the break between the function words such as *our, her, since, the,* and the following content words have much higher relative values in parsing than in pausing. This is because these words are cliticized in speech (leading to very short pauses) but are treated as independent elements in the parsing task. Further, effects on pausing of the com-

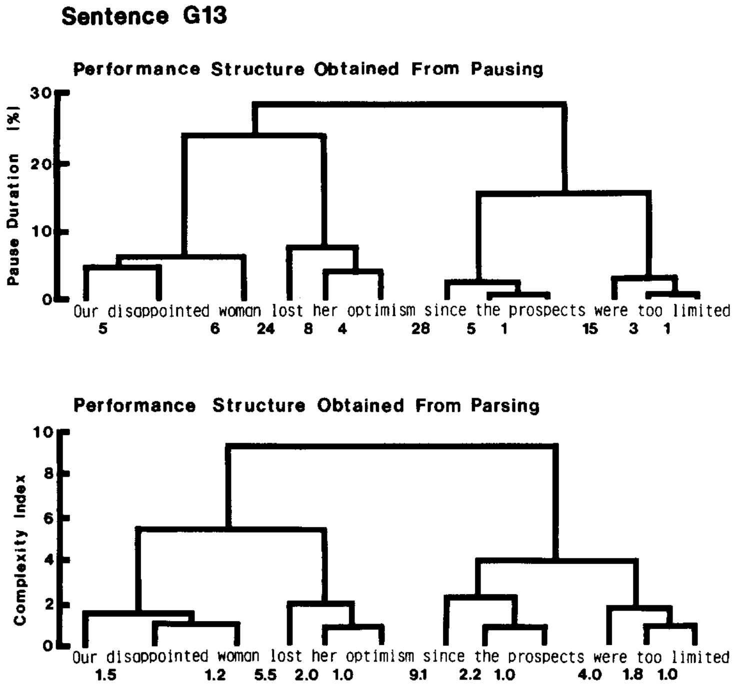


FIG. 1. Performance structures for sentence G13 obtained from pausing and parsing. The values obtained from each task (percentage pause duration and mean complexity index (CI)) are used to give height to the nodes of the performance structures along a ratio scale. (Adapted from Grosjean, Grosjean, & Lane, 1979.)

plexity of individual words and of stress probably do not show up in parsing judgments. To avoid such task specificities and because we are more interested in this paper in the prosodic aspects of speech, we will only examine the pause data published by Grosjean et al. (1979). (The 14 sentences used by Grosjean et al. as well as the percentage pause durations found at each word boundary in each sentence are presented in the Appendix. They are numbered from G1 to G14 and will be referred to in this way in the text.) While we will test out theory solely on the pause data, we will point out throughout the paper how the theory can account for other data as well.

If we examine performance structures, for example those presented in both Figs. 1 and 2, they appear to have a number of properties the significance of which we cannot know in the absence of a theory that explains them. First, the data looks to be broken into "basic units" that are rather small. An examination of the frequency distribution of the 154 pauses obtained from the 14 sentences shows a significant dip at the 7%

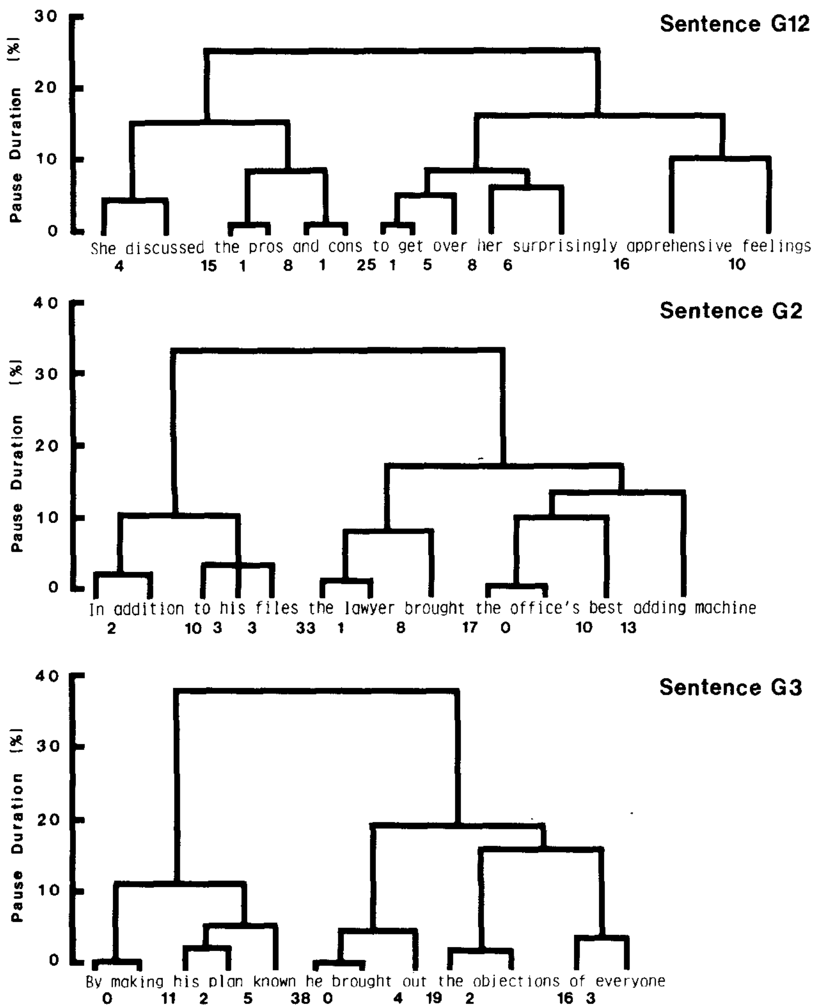


FIG. 2. The performance structures of three sentences based on the percentage pause durations by Grosjean et al. (1979). The pause durations are used to give height to the nodes of the structures along a ratio scale.

pause duration (%PD). In fact, the pauses less than 8% almost always separate function words ("little" stressless grammatical words like "*the*") from words they are phonologically connected with (a phrase like "*the apple*" is pronounced almost as one word). Later, we will show that these boundaries, between function words and the words they are connected with, are the most resistant to pausing. If we look at units separated by pauses ≥8% we find, however, a motley array indeed: single words (e.g., "apprehensive" in sentence G12), syntactic constituents (e.g., the NP "our disappointed woman" in G13 Fig. 1, the VP "lost her optimism" in G13 also, the PP "to his files" in G2), and groups of words that are not syntactic constituents (e.g., "she discussed" in G12, "he brought out" in G3, "by making" in G3, "since the prospects" in G13). Later, however, we will demonstrate that a large majority of these units do fall under a simple linguistic characterization.

Second, performance structures appear to have a pretty rich hierarchical structure. The structures are not flat; they do not have relatively uniform pausing throughout. Rather, there is quite a range in pause durations, for example, percent pause durations in sentence G13 Fig. 1 go all the way from 1 to 28% with intermediate steps of 6, 15, 24%, indicating thereby the clear hierarchical nature of these structures.

A third property of performance structures is that they are more or less symmetrical (or balanced). That is, the main pause break is located close to the middle of the sentence; then, each segment on either side of the break is itself broken up into more or less equal parts and so on. It is this property of symmetry that has led several investigators to incorporate something like a "bisection parser" into their models of pausing behavior (see, e.g., Grosjean et al, 1979; Cooper & Paccia-Cooper, 1980, and discussion below). We will eventually argue that this seeming symmetry of performance structures is an "epiphenomenon" and is the result of the syntactic and prosodic properties of the language. We will need to make no appeal to any purely performance property of "bisection" in our final account of performance structures.

In sum, performance structures can be obtained by various experimental paradigms such as rote and probed recall, slow reading, parsing, and marking relatedness judgments. These structures are relatively invariant across tasks although each paradigm does produce certain characteristic results. In addition, performance structures have three main properties: relatively small basic units, hierarchy, and symmetry. It is interesting to note that nonlinguistic sequential patterns obtained from tapping, or giving oral responses to auditory stimuli have also been shown to share these same three properties (Handel & Todd, 1981).

2. *Predicting Performance Structures: Whole Sentence Algorithms*

Below we present two algorithms that can be used to predict performance structures: The first is an algorithm proposed by Grosjean et al. (1979) to account precisely for these structures and the second is an algorithm proposed by Cooper and Paccia-Cooper (1980) to account for pausing, segmental lengthening, and blocking of cross-word conditioning of phonological rules. Several points need to be made before describing each algorithm and determining how well each accounts for the pause data presented in the Appendix. First, these algorithms are not models of actual performance, in that they do not attempt to account for how speakers actually produce performance structures. Rather, they isolate and combine variables that appear to be important in explaining performance structures. Second, because these are algorithms and not models, the prediction procedure they use is often ad hoc. Their most important trait is that they attempt to isolate the appropriate factors and not so much that they determine precisely the values to give to these factors. Third, both algorithms need to have the whole sentence before they begin their prediction process. In this sense they are very unlike the basically left to right process of spontaneous language production (nor were they intended to be). Fourth, both algorithms have a component which determines an index of the syntactic strength of each word boundary in the sentence. Hence, both algorithms need a surface structure representation of the sentence. In our application of the algorithms we have used the $\bar{X}$ theory of phrase structure (Chomsky, 1970a; Jackendoff, 1977),[1] but we should note that up to now the two algorithms were used with a more traditional phrase structure notation (Chomsky, 1965). Finally, to determine how well each algorithm predicts the pause data, global and mean correlations are computed between the values output by the algorithms and the percentage pause durations of the 14 sentences. Global correlations are based on the total number of word boundaries across all sentences (N = 154) and mean correlations are based on individual correlations obtained for each sentence. The two values are usually very similar but both will be reported here. To test the difference between global correlation coefficients, we will use a two-tailed test of the equality of two correlations coefficients for related samples (Weinberg & Goldberg, 1979).

a. The Grosjean, Grosjean, and Lane (GGL) algorithm. This algorithm has been used to predict performance structures obtained from pausing, parsing, and TEPs in speech, and pausing, parsing, memory probe data, and relatedness judgments in American Sign Language (Grosjean et al., 1979; Dommergues & Grosjean, 1981; Grosjean et al., 1981). It is based on the premise that the importance of a particular break (as evidenced

[1] We used X-bar theory because it is currently the most sophisticated theory of phrase structure. It is, however, not crucial to the understanding of this paper. For those not familiar with X-bar theory, $\bar{\bar{X}}$ can be taken to mean a syntactic phrase of any type, and $\bar{\bar{N}}$ to mean NP, and $\bar{\bar{V}}$ to mean VP. One piece of notation that is relevant is "$\bar{S}$" (S single bar). $\bar{S}$ is a node that dominates a node COMP and a node S. The COMP node is a "catch-all" position for complementizers ("that," "for," "who," "when," etc.), or preposed (fronted) phrases and clauses ("After the game"). S just stands for the basic sentence after the complementizer or preposed material. COMP, of course, can be empty (there may be nothing in front of the basic sentence). See Section 3 below for some discussion.

by the experimental data) is affected both by the relative importance of that break as defined by the surface structure of the sentence, and by the position of that break within the constituent being analyzed: the nearer the middle of the constituent, the more important the break. When constituents are of unequal length, subjects will attempt to displace the pause to a point midway between the beginning of the first constituent (for example, an NP) and the end of the second constituent (for example, a VP), if at that point there occurs a syntactic boundary important enough. It would seem that a compromise takes place between this bisection tendency and the linguistic structure of the sentence. Thus the algorithm takes into account the product of two (sometimes conflicting) demands on the speaker: the need to respect the linguistic structure of the sentence and the need to balance the length of the constituents in the output. It is a simple cyclical algorithm, combining for each word boundary in the sentence, an index of linguistic complexity and a measure of the distance to the midpoint of the segment.

Since the Grosjean et al. algorithm (1979; GGL) is published, we will not repeat it here. We will use certain abbreviations in the discussion below: "CI" is an index of the syntactic complexity at a word boundary; it is based on the number of nodes dominated by the word boundary node (CI = the number of nodes dominated by the node dominating the word boundary, including in the count the word boundary node itself). "%RP" is an index of the relative proximity of the word boundary in question to the bisection point of the constituent. "GGL values" are the values obtained from the GGL algorithm; they are the product of the CI and the %RP and are obtained in a cyclical manner.[2] And "%PD" are the pause durations (expressed as a percentage) presented in the Appendix.

The predicted performance structure for sentence G11 is shown in Fig. 3 (bottom tree). As can be seen, when it is compared to the top tree (the performance structure itself), it is a fairly good match of the pause data. The main break in both structures is after *book,* and the second main breaks are after *agent* and *in which*. And, indeed, the correlation between the predicted values and the %PD is .89. Had the CI values been used

[2] The GGL algorithm, briefly, has the following steps: (1) Starting with the largest constituent that has not been analyzed, compute CI for every word boundary; (2) compute also for each word boundary %RP (= the number of words from the start (or end) of the constituent to the boundary (whichever is less) divided by half the number of words in the constituent, expressed as a %); (3) multiply the two values assigned to each word boundary: The boundary with the largest product is the constituent break and retains its product. No other product is retained; (4) take each of the constituents just created, calculate CIs and %RPs for each word boundary, multiply these values, find the largest product, and ignore the others (i.e., repeat steps 1 to 3 for each constituent). Continue until every word boundary has a value.

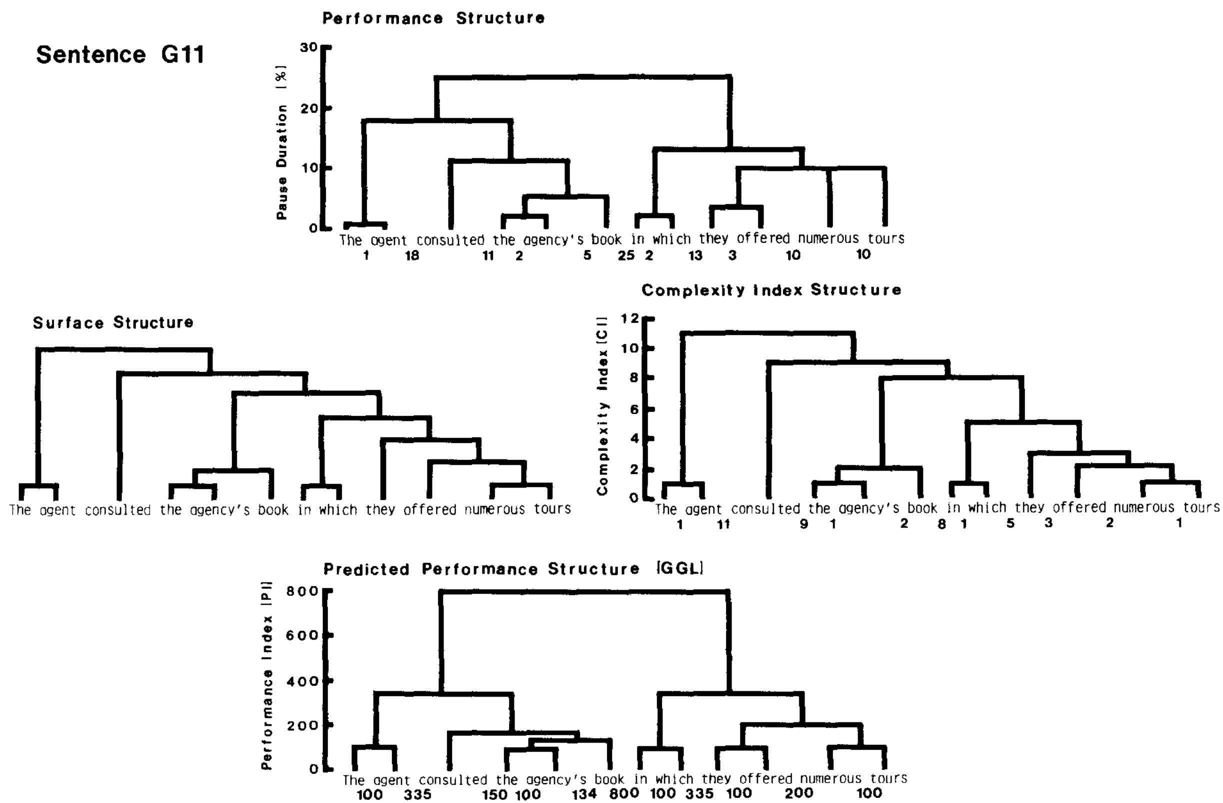


FIG. 3. Four hierarchical structures for sentence G11: the performance structure based on percentage pause durations, the syntactic surface, the structure based on the Grosjean et al. (1979) complexity index (CI), and the performance structure predicted by the Grosjean et al. algorithm (GGL).

by themselves, without correcting for the bisection component (see right hand middle tree), the match would have been much less good. The main break would have been predicted after *agent* and not after *book,* the break after *consulted* would have had too high a value and so on (the correlation between the CI and the pause data would then only be .77).

Overall, the GGL algorithm—which combines each CI multiplicatively with a measure of proximity to the midpoint of the constituent—is a good predictor of the performance data. The global correlation is .83 and the mean of the 14 individual correlations is .86. As expected, the CI by itself does much less well: The global correlation between CI and %PD is .76 and the mean of the 14 correlations is .79. Table 1 presents the individual CI and GGL correlations.

Although the GGL algorithm accounts for 69% of the variance of the pause data, it does leave some factors unaccounted for. These not only keep the global correlation down to .83 but are also responsible for the rather wide range of individual correlations: .63 to .96. One major problem with the algorithm is the complexity index: It takes into account

TABLE 1
Individual Sentence Correlations between Percent Pause Durations and Various Predictors

Sentence	No. words	GGL algorithm CI	GGL algorithm GGL	CPC algorithm BS	CPC algorithm CPC	PHI algorithm
1	11	.69	.77	.87	.81	.97
2	11	.81	.92	.93	.77	.99
3	11	.75	.96	.70	.68	.99
4	10	.84	.93	.91	.80	.99
5	10	.75	.82	.96	.86	.93
6	12	.84	.92	.80	.78	.98
7	11	.64	.70	.79	.59	.97
8	11	.92	.92	.84	.80	.97
9	11	.87	.85	.87	.82	.99
10	11	.88	.96	.79	.84	.96
11	11	.77	.89	.80	.71	.98
12	12	.65	.63	.83	.76	.96
13	11	.85	.92	.94	.88	.94
14	11	.82	.88	.84	.83	.96
Mean		.79	.86	.85	.78	.97
Global r (N = 154)		.76	.83	.81	.75	.96

Note. Grosjean, Grosjean, and Lane's syntactic complexity index (CI) and their algorithm (GCL); Cooper and Paccia-Cooper's boundary strength index (BS) and their algorithm (CPC), and the phonological phrase algorithm (PHI). Global correlations (based on all 154 pauses across all 14 sentences) are also given.

the number of nodes dominated by the node immediately dominating the word boundary (and in this sense gets at the richness of the structure on either side of the boundary) but it does not pay attention to the labels of these nodes. This can have quite serious consequences as can be seen in Fig. 4. The top tree is the performance structure of the sentence *John asked the strange young man to be quick on the task.* The main break is situated between the main clause (*John asked the strange young man*) and the infinitival complement (*to be quick on the task*). Each of these two segments are then broken down into smaller groups: *John asked/the strange young man/to be quick/on the task.* The first two groups are then separated into *John/asked* and *the strange/young man.* The left-hand middle tree in Fig. 4 shows the $\bar{X}$ phrase structure of the sentence (minus node labels). We note that the linguistic structure is right branching with a very short NP subject and a very long VP. We also note that the NP "the strange young man" is right branching and that the infinitival complement has a trinary structure. The tree based on the CI count is presented to the right of the syntactic tree. It is very similar to the $\bar{X}$ tree except that the count attaches the verb (*asked*) to the following NP and transforms the trinary infinitival complement into a binary tree. The correlation between the CI and the %PD is only .64. This is because the CI value between *John* and *asked* is much too high (syntactically it is the main break of the sentence but the main pause break is after *man*), the CI value between the verb (*asked*) and the following NP is too low, and the NP is right branching instead of being binary. Also the CI value between *quick* and *on* is too low. Can the bisection component rectify this? We note first (see bottom tree in Fig. 4) that the algorithm does place the main break between the main clause and the infinitival complement (that is between *man* and *to be quick*). We also note that in the infinitival complement, it puts the main break between the VP and the PP (although it does not give this break a high enough value). The real problem occurs in the main clause. When it is no longer linked to the infinitival complement after step 4 of the algorithm, this clause is totally right branching with the largest CI value between *John* and *asked* (CI of 5) and the lowest value between *young* and *man* (CI of 1). Thus when the %RP is combined with the CI, the largest product (300) occurs between the determiner (*the*) and the adjective (*strange*). Each subpart is then reanalyzed and the final values are such that the highest performance break (between *asked* and *the*) is given the lowest PI value (100), the lowest performance break (between *the* and *strange*) is given the highest PI value, and important pause breaks such as between *John* and *asked* and between *strange* and *young* are given close to minimum GGL values (134). These mismatches lead the GGL algorithm to predict the pause data of this sentence rather poorly (r = .70).

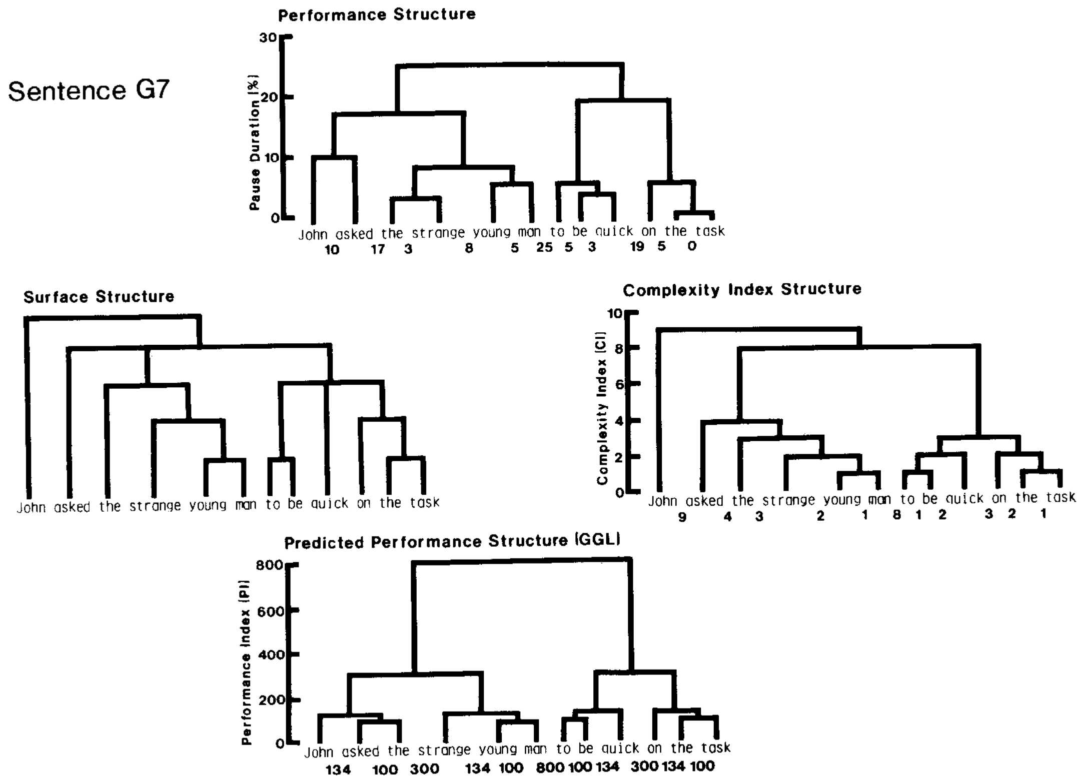


FIG. 4. Four hierarchical structures for sentence G7: the performance structure based on percentage pause durations, the syntactic surface structure, the structure based on the Grosjean et al. (1979) complexity index (CI), and the performance structure predicted by the Grosjean et al. algorithm (GGL).

A small but important change that can be made to the syntactic structure of the sentence (and hence to the CI account) is to link the function words to their content words (determiners to nouns, particles to verbs, pronouns to verbs, and so forth), before drawing the surface structure tree of the sentence. This would ensure that the breaks between function and content words would usually have CIs of 1 and would thus be in line with the %PDs which are very low at these locations (often 0, 1, or 2%). In Fig. 5 we have done just this. As can be seen, in the left-hand middle tree, the function words have been grouped with their content words (thick lines) and then the rest of the syntactic structure has been drawn. This small modification of the surface structure tree does not, admittedly, put the main CI value after the main clause but it does change the configuration of the NP in the main clause (*the strange young man* is now a binary structure) and the CI is now more important at the break between *to be quick* and *on the task*. This results in an increase in the correlation between the CI and the %PD: from .64 it goes to .73. Applying the bisection component, we obtain a tree (bottom structure) that is now quite similar to the performance structure (top tree). Not only is the main break where it should be (as it was in Fig. 4), but now the heights of the node between *quick* and *on,* and *strange* and *young* are appropriate. Although the nodes between *John* and *asked* and *asked* and *the strange young man* are still low, the overall correlation (.88) is now far superior to the one obtained in Fig. 4 (.70).

When the correction for function words is applied to the syntactic structure of each of the 14 sentences, the global correlation between the GGL values and the %PD goes up from .83 to .89, a significant difference ($t = 5.14$, $p < .01$), and the mean of the 14 individual correlations goes from .86 to .90. What is especially important is that the range of the 14 correlations is now much smaller: it goes from .83 to .98 instead of going from .63 to .98. We should also note that now the CI measure by itself is as good a predictor as the GGL values (global mean of .87 as compared to .89) but there is still an indication that GGL does better at predicting the %PD of unbalanced sentences. The global correlation between %PD and the CI corrected for function words is .79 for the three unbalanced sentences ($N = 34$) and this correlation increases to .85 when the bisection component is brought in. However, the difference between the two coefficients does not reach significance, probably because of the small number of data points.

Thus, attaching function words to the appropriate content words before computing the CI allows the algorithm to account for 79% of the variance instead of only 69%, but it does not solve the labeling problem completely. The measure still remains blind to the actual nodes in question. An example of this can be seen in Fig. 5 (bottom structure) where in the last

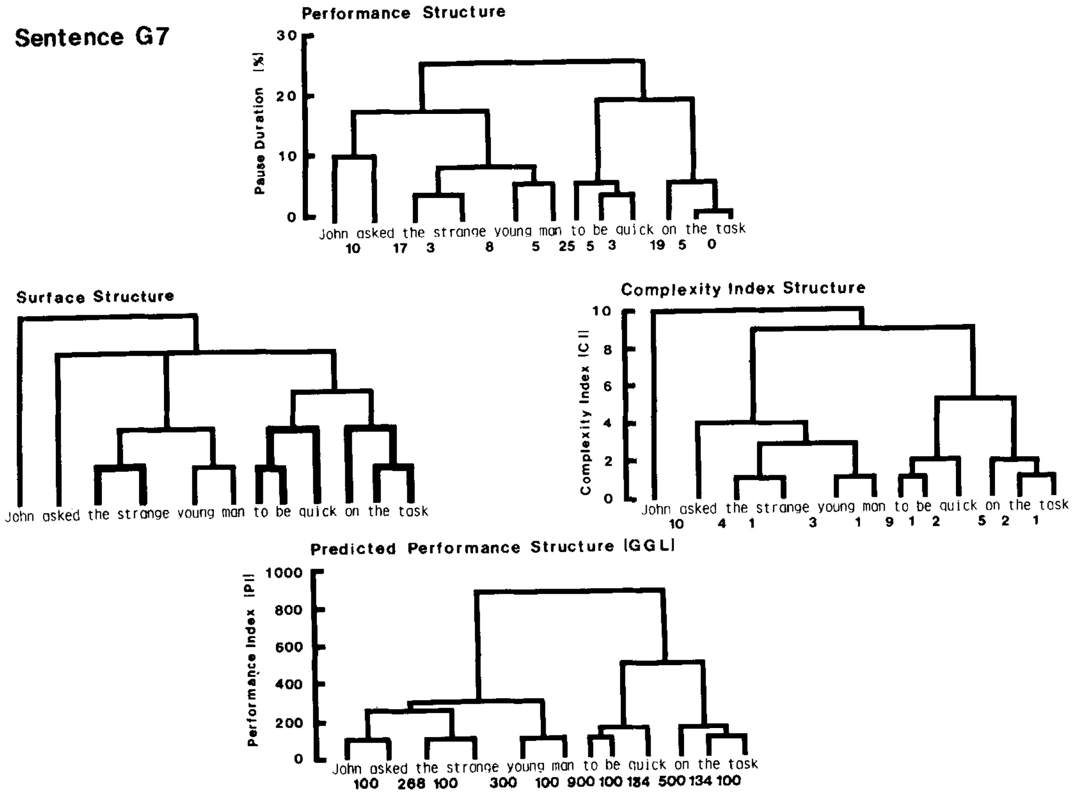


FIG. 5. Four hierarchical structures for sentence G7: the performance structure based on percentage pause durations, the syntactic surface structure after having grouped function words with their appropriate content words, the structure based on the Grosjean et al (1979) complexity index (CI), and the performance structure predicted by the Grosjean et al. algorithm (GGL).

stages of the iterative cycle, the algorithm is faced with two segments: *John asked* and *to be*. The word boundary in each segment receives a CI value of 1 and a %RP of 100, giving a PI of 100. And yet, we are dealing with two very different breaks: an NP–VP break on the one hand and an infinitive marker–copula break on the other. This distinction is made by the speaker (10% PD as opposed to 5%) but not by the algorithm (also note the difference between the %PD at the *John/asked* break in G7, i.e., 10%, and that at the break between *cold* and *winter* in G10, i.e., 5%). Another clear example is seen in Fig. 3 where the break between *the* and *agent* (Det–N) is given the same GGL value (100) as that between *numerous* and *tours* (adj–N). In the data, however, one finds a large break in the latter case (10%) but only a small break in the former case (1%).

Other factors may need to be accounted for by the GGL algorithm. One of these seems to concern the impact that the length of the words has on the preceding and following pauses. Grosjean et al. (1979) found, for example, that the %PD of the break preceding a 1-, 2-, and 3-syllable word (at a constant CI value of 1) remains constant (about 4%) but then rises to 9% before a 4-syllable word and to 11% before a 5-syllable word. A second factor concerns focus stress. Words or phrases that carry focus stress will usually be preceded by rather high %PD. Thus in sentence G6 (*That the matter was dealt with so fast was a shock to him*) the focus stress is on *so fast* and therefore it is preceded by a %PD of 18, the second highest %PD in the sentence. And in sentence G1 (*When the new lawyer called up Reynolds the plan was discussed thoroughly*) the word carrying the main focus stress (*thoroughly*) is preceded by a pause of 21%, also the second highest pause in the sentence. In neither case does the GGL algorithm give values that are high enough.

To conclude, the GGL algorithm is simple to compute and is a relatively good predictor of the pause data (see Grosjean et al., 1979; Grosjean et al., 1981; and Dommergues & Grosjean, 1981 for its ability to predict other types of data). The predictive power of the algorithm could be greatly strengthened if it took into account the label of the nodes as well as word length and focus stress. It is interesting to note that whatever changes are brought to the algorithm, it will not help it model any better the actual production process followed by the speaker. Admittedly, Grosjean et al. (1981) have proposed that the speaker–listener initiates any sentence processing task with a base line structural expectation which is an unmarked binary tree. If the surface structure tree is close to the unmarked tree, that is, the sentence is balanced, then to that extent the processing can proceed in terms of such a structure. When it comes to an unbalanced sentence, however, the unmarked tree and the surface structure assign different hierarchical structures. This will lead subjects to weigh the surface tree by the unmarked tree in performing the hierarchical analysis. One problem with this explanation, however, is that the subjects must have the whole sentence before being able to compute the surface structure and the unmarked structure and then combine the two, whereas there is little evidence that this is the case in real time output. This theory can thus explain in part the data obtained from paradigms in which the sentences are known in advance, but cannot account for the performance structures of totally new sentences. In addition, few sentences are exactly 2, 4, 16, or 32 words long and yet these are the only structures that have a totally binary structure. And, lastly, the bisection component in the algorithm is based on the distance from the midpoint of the constituent at hand (the further away the word boundary from the middle of the sentence, the smaller the %RP) whereas a binary tree does not have a strictly decreasing set of nodes on either side of the midpoint. In short, the GGL algorithm may be an adequate predictor of pause data in read out speech, but it has little to say about sentence production.

b. The Cooper and Paccia-Cooper (CPC) algorithm. In their book, *Syntax and Speech* (1980), Cooper and Paccia-Cooper propose a general algorithm to account for segmental lengthening, pausing, and blocking of cross-word conditioning of phonological rules. This algorithm, derived from their research on the relationship between syntax and temporal variables in read out speech, is characterized in the following way by the authors:

> In developing this algorithm, the primary aim was to provide a quantifiable method of predicting these prosodic effects in individual utterances without regard to whether steps included in the algorithm or their order of application might be analogous to human processing operations involved in programming such attributes. (P. 182)

This ambitious algorithm is composed of numerous steps (up to 14) and predicts at various points along the way the probability of pausing at each of the word boundaries in a sentence, the probability of blocking, the probability of segmental lengthening, as well as predicting the duration of segments and the duration of pauses (we should note that the authors (personal communication) claim that their algorithm is a better prediction of phonological blocking and segmental lengthening than it is of pausing). In order to determine how well the CPC algorithm predicts our 14 performance structures, we followed the first five steps outlined by Cooper and Paccia-Cooper. The remaining steps were not followed as they do not change the relationship among the values obtained at step 5. Hence the correlation coefficients obtained with the values predicted at that step will not be any different from those obtained at later steps.

The CPC has the following general characteristics. First, like the GGL algorithm, it is based on the whole sentence. That is, the surface structure of the whole sentence is needed in order to predict values at the individual word boundaries. Second, the CPC algorithm puts much weight on its syntactic complexity index (what it calls the boundary strength). It is based on branching depth rather than node height, it considers only nodes that flank the word boundary in question (unlike the CI which takes into account the full richness of the structures on either side of the boundary), it differentiates between the types of nodes (brancing S nodes are given more weight and minor category or function word nodes are not counted), and it gives added weight to the left flanks in the node count. Thus, instead of being a mere reflection of the structural tree as is the CI, the boundary strength index (= BS) is a performance algorithm in itself and the tree obtained from it is often quite different from its surface structure counterpart (see middle trees in Fig. 6). Third, the CPC algorithm has a bisection component, but the proximity index is calcualted only on content words and bisection applies only when the constituent in question contains seven major category words or more. Thus, for most sentences, bisection only applies once, whereas in the GGL algorithm it applies iteratively until every word boundary has a "largest product." Finally, a correction factor is included for constituents that are either very long or very short. This is based on the general finding that the longer the constituent, the longer the pause (or the segmental lengthening) at the end, and the shorter the constituent the less important these temporal variables.

Due to the length and complexity of the CPC algorithm, we cannot present it here, and thus, we recommend the reader interested in the details to the full description of the 14-step algorithm on pages 180–193 in Cooper and Paccia-Cooper (1980). In our discussion we will use the following abbreviations: "BS" is an index of the syntactic complexity at a word boundary; it is based, among other things, on the number of flanking nodes on the immediate left and right of the word boundary. "CPC values" are the values obtained from the CPC algorithm; they are the product of the BS index and a bisection index that is itself based on major lexical words. Bisection only applies to constituents that contain more than seven major category words.

A comparison of the performance pause structure of sentence G11 (top tree in Fig. 6) and of its predicted structure by the CPC algorithm (bottom tree) shows a number of similarities but also some mismatches. The algorithm predicts the main pause break accurately (it is between *book* and *in which*) but it has problems with the other values. The second main performance break is after *agent* (18%) but the CPC algorithm gives this break a value that is far too low (1.16) and the third main break (after *in which,* 13%) is also given too low a value (.86). Conversely, the breaks

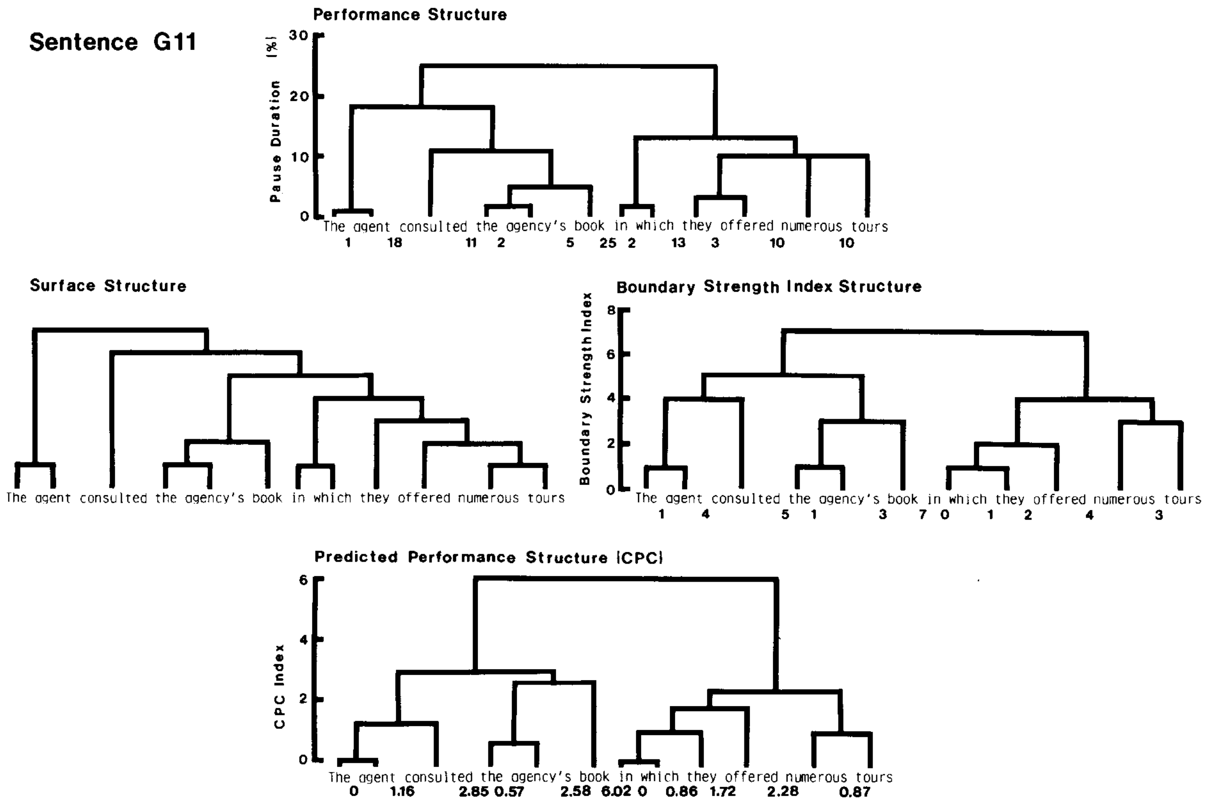


FIG. 6. Four hierarchical structures for sentence G11: the performance structure based on percentage pause durations, the syntactic surface structure, the structure based on Cooper and Paccia-Cooper's boundary strength index (BS), and the performance structure predicted by the Cooper and Paccia-Cooper (1980) algorithm (CPC).

after *agency's* (5%) and *they* (3%) are given values that are too high (2.58 and 1.72). This explains why the correlation between the %PD and the CPC values is only .71. The global correlation for all 14 sentences is .75 which is significantly different from GGL's .83 ($t = 3.01$, $p < .01$). The mean correlation is .78 and the range extends from .68 to .88.

To explain why the CPC algorithm only accounts for 56% of the variance as compared to GGL's 69%, one needs to examine Cooper and Paccia-Cooper's BS index, their bisection index, and the way they combine the two in their algorithm. A comparison of the performance structure in Fig. 6 (top tree) and of the corresponding BS structure (middle right tree) shows the two to be quite similar. The flanking node count of the BS, its greater weighting of left flanks and of branching S nodes, and its elimination of the minor category nodes allow it to predict the main break correctly (between *book* and *in which*) whereas the CI index (Fig. 3) does not do so. And although *consulted* is incorrectly paired with *the agent* by the BS algorithm, the value at the break between the two is appropriately quite high (4). Also the NP object of the main clause (*the agency's book*) is parsed appropriately. The subordinate clause (*in which they offered numerous tours*) is not structured correctly by the BS, but the values of certain breaks such as those between *offered* and *numerous* and *numerous* and *tours* correspond quite well to the pause data. Thus, the BS and the %PD are correlated quite highly (.80) as compared to a correlation of .77 with the CI in this sentence. Overall, the BS index appears to be a better predictor of the performance data than the CI: .81 as compared to .76. (The means of the 14 individual correlations are .85 and .79, respectively. However, this difference is not significant ($t = 1.69$, NS).)

Unfortunately, the predictive power of the BS index is reduced instead of increased when the bisection component is introduced in steps 2, 3, and 4. As can be seen in Fig. 6, the performance structure predicted by the CPC algorithm (bottom tree) resembles less the actual performance structure (top tree) than does the BS tree (middle structure on the right). Although the main break is predicted correctly, the values of the other important breaks are not what they should be. The values of important breaks on either ends of the sentence (after *agent* and after *numerous*, for example) are far too low whereas the values of unimportant breaks toward the middle of the sentence (after *agency's* and after *they*) are too high. This is because the bisection cycle is only applied once in the algorithm. Thus breaks near the middle of the sentence receive high bisection indices, and hence high BS $\times$ Bisection products, whereas breaks at the extreme ends of the sentence receive low bisection indices, and hence low products. This fact may be helpful for minor breaks at either ends of the sentence, but not for important syntactic breaks whose values are thereby considerably lowered. Thus, for instance, the break between *agent* and *consulted* receives a BS of 4 which represents 57% of the BS value of the main break in the sentence (7). This corresponds quite well to the %PD at that break (18%) which represents 72% of the value of the highest performance break (25%). However, after adding the one cycle bisection component in steps 2 and 3, the final product at that break is 1.16, a value that is only 19% of the value of the predicted main break (6.02). A similar situation occurs at the break between *numerous* and *tours:* the %PD is 10 (40% of the highest %PD), the BS index is 3 (43% of the highest BS), but the final CPC algorithm value is .87 (a mere 15% of the highest CPC value). We should note that Dommergues (Note 4), who has done a careful comparison of the GGL and CPC algorithms, has found that if the CPC algorithm contained a cyclical bisection component (as does the GGL algorithm), and if it applied to all words and not just to content words, its predictive power would be increased but that it would not reach a level higher than that reached by the BS itself. This would seem to show that the BS index already accounts for the speaker's need to bisect the output and that applying a bisection component can either hinder the predictive power of the algorithm (as in the case of the one-cycle bisection) or have no effect on it (as when the GGL iterative bisection is applied).

To conclude this section on the GGL and CPC algorithms, we should note that each has its problems and each its advantages. The GGL algorithm could be greatly improved if the label of the nodes were taken into account in the CI, as well as if the length of the words and the focus stress were accounted for. On the other hand, the GGL algorithm has the advantage of being straightforward and easy to work through. It is also a good predictor of the %PD as well as of other experimental data such as relatedness judgments, parsing values, transitional error probabilities, and probe latencies in both speech and American Sign Language.

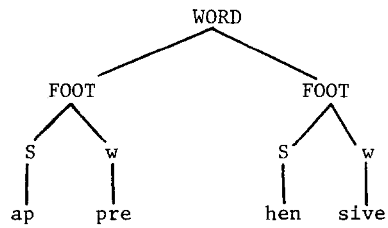
As for the CPC algorithm, it too has problems and advantages. On the negative side we can note that the algorithm has not been previously tested for its predictive power. Had this been done, it would have been discovered that adding a bisection component actually hinders, instead of improves, the algorithm. In addition, the BS index is somewhat difficult to compute: Some of the instructions are vague and certain terms are perhaps not well defined, so that different BS values may be computed by different people at the same word boundary. However, BS is an interesting index: it takes into account node labels (and hence lowers the values between function and content words) and only counts flanking nodes, with the effect of balancing unbalanced trees. The BS index by itself is therefore a good predictor of the pause data.

Both algorithms are considered by their authors as first steps toward accounting for prosodic variables in the production of language (however, as we have pointed out, they also account for the other data, such as parsing values and relatedness judgments). They are both "whole sentence algorithms," in that they take as input the entire surface structure of the sentence. Furthermore, both compute values in ways that do not directly offer very fruitful hypotheses for a real processing model, and they are not explicated in terms that are naturally and directly relevant to an overall theory of linguistic structure, from either the point of view of competence or performance. An algorithm that accounts for the performance data and that, at the same time, has the property of working left to right, building a new representation that integrates all levels of linguistic structure in a disciplined way, would seem to offer us a better chance of understanding the coding of language into speech. We will now turn to such an algorithm, an algorithm that treats the issue of prosody directly and on its own terms.

3. *Predicting Performance Structures: A Prosodic Structure Algorithm*

What we would like to know now is why the correlation between the GGL and CPC algorithms and the data, though quite good, is not better? What further factors need to enter the picture? There is also a deeper and more interesting question. What account can we give that will unify the various determinants of pausing into a linguistically and psychologically relevant system? Such a system, if it exists, would have to be relevant to the theory of linguistic competence, since we have already shown that pausing is determined in large part by syntactic structure, and to theories of performance, since we are, after all, dealing with a performance task.

a. Prosodic structure. There is, in fact, a likely candidate for such a unifying system, a system with the right sort of properties, namely the level of the grammar that specifies or assigns prosodic (rhythmical) structure to sentences. This part of the grammar is not as widely known to psycholinguists as the part that assigns segmental phonological representations to sentences or the part(s) that assigns syntactic representation to sentences. Recent phonological theories (and prosody falls into the phonological component of the grammer) have greatly modified the classical view of phonological theory in Chomsky and Halle (1968). These theories see the utterance as having a suprasegmental, hierarchical organization in terms of a "metrical" (or rhythmical) tree which is binary branching and between whose nodes the prominence relation "strong/weak" is defined. The details of these theories are not relevant to our claims (and in fact, as is the way with theories, are subject to relatively rapid change). We will seek, however, to use (and ultimately confirm) some of the basic concepts behind such theories. What is important in such theories for us is that they see the sentence as made up of a number of *prosodic units,* units which partly, but not wholly, match the phonological and syntactic units of the sentence.[3] Each unit specifies the rhythmical properties of some level of the sentence, whether it be the syllable, the word, the phrase, or the sentence as a whole. For example, the word at the prosodic level is seen as made up of "feet," which are patterns of strong and weak syllables, much like in the meter of a poem.

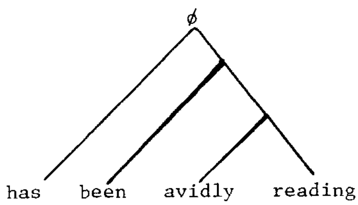


While it is true that even the rhythmical properties of words influence pausing behavior, we are obviously interested in matters above and beyond the level of the word. And what is interesting for us is that words at the prosodic level are seen as grouping together into units smaller than a syntactic phrase, that is, they group into units called "phonological phrases," hereafter referred to as "ϕ-phrases." This unit is intermediate in size between the word and the syntactic phrase. And it turns out to be just the right unit we need to predict the distribution of pauses in our data.

The idea behind a "ϕ-phrase" is extremely simple. Consider, then, for a moment a syntactic phrase, say the VP in the top part of Fig. 7: "has been avidly reading about the latest rumors in Argentina." Any syntactic phrase has a *head.* The head is the main word around which the phrase is organized (the category from which the phrase is projected, in linguistic terminology). While the notion "head" can be rigorously specified in linguistic theory (see Jackendoff, 1977), intuitively the head of a VP is its main Verb, the head of an NP is its main Noun, the head of an Adjective Phrase is its main Adjective, and so forth. In our VP ("has been

[3] The version of prosodic theory we have borrowed most from is Selkirk (1980, forthcoming, Note 2, Note 5). In work in progress both A. Prince and E. Selkirk are developing an alternative phonological theory, based on "metrical grids" (see Liberman, Note 1), as opposed to the hierarchical metrical theory we have delineated here. Again, however, it is not our interest to support the details of a particular phonological theory, but rather to demonstrate how certain abstract properties, shared presumably by any adequate prosodic theory, give an illuminating account of long-standing issues stemming from performance data (see Chomsky, 1976, on the metatheoretical issue involved here).

avidly reading about the rumors in Argentina") the head is the Verb *reading*. In the NP "the small boy from New Jersey" the head is the Noun *boy*. In a phrase, all other words beside the head either modify, specify, or complement the head in some way. A ϕ-phrase is defined simply as follows: In any syntactic phrase, all the material up to and including the head is a ϕ-phrase. Thus, in our example VP, the material up to (i.e., "has been avidly") and including the head (i.e., "reading") is a ϕ-phrase: "has been avidly reading." And in a NP like "small boy from New Jersey," where the head of the phrase is *boy*, the material "the small boy" makes up a ϕ-phrase (we defer for just a moment what happens to the material in the syntactic phrase that is left over). To construct a tree structure for a ϕ-phrase we just join all the elements preceding the head to the head in a simple right branching structure (in actuality the nodes would be labeled as "strong" or "weak" in prosodic theory, a detail that is not relevant here):

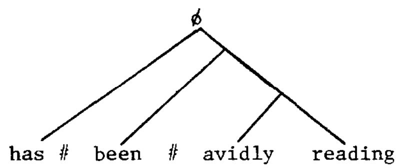


ϕ-phrases are basically made up of one major stress (on the head) preceded by weaker or zero stresses. They have actually been around in one form or another, usually under the name of "stress groups," for some time, though current theory fits them into an articulated theory of prosody that is part of the grammar as a whole.

We must stop a moment to make a special proviso about "function words." The matter seems trivial, but it is not. Function words are the "little" words that belong to nonlexical categories such as determiner, auxiliary verb, or conjunction. Such words bear little or no stress and in fact lose their word status, so to speak, in prosodic theory. They behave as if they were single weak syllables attaching themselves parasitically to a sister constituent next to them (they act like enclitics, they ride prosodically on another word).[4] They are subject to various phonological

[4] Bisyllabic function words (like "about") carry some significant stress to distinguish the two syllables, though not significant compared to the nonfunction words in the sentence. We use the word "clitic" loosely by analogy to the enclitics in a language like French (where there is a yet tighter bond between the enclitic and the word it is phonologically adjointed to). Two nodes A and B are sisters if they are immediately dominated by the same node (note that the direct object of a verb is a sister to the verb since both are immediately dominated by VP and that an object pronoun is, thus, attached to the verb by "#").

processes, such as vowel reduction, and of course they are resistant to any pause between themselves and the word they have adjoined to. Of course, there are various complications in how to formally handle function words in a prosodic theory, some of which are dealt with in detail by Selkirk (1980, Note 2, Note 5, and forthcoming). We will place a "#" in any boundary between a stressless function word and the word next to it. This # notates that the two words have been almost joined to each other for purposes of rhythm or prosody (our # symbol is just a shorthand here for the operation of various phonological rules, such as Selkirk's Monosyllable Rule, operating on function words):



Psychologists, as we have seen, have ignored the fact that function words do not behave prosodically as full words to their peril. This treatment of function words has an interesting consequence for the view we take of prepositional phrases (PPs), an ever-present category in English sentences. Prepositions are the heads of PPs (they are the category around which the phrase is organized), but they are also (most of them) function words. As stressless function words they lose their word status prosodically, and so cannot serve as heads any longer. In particular, they cannot terminate ϕ-phrases (since heads terminate ϕ-phrases). Rather, they make up ϕ-phrases with the material up to and including the next head "down the line," namely the head of the NP that serves as a complement to (that follows) the preposition. In a PP like "about the latest rumors in Argentina" (the material left over in our VP after making "has been avidly reading" a ϕ-phrase), the preposition "about" is prosodically weak and so we make up a ϕ-phrase out of all the material up to and including the head of the NP following this preposition ("the latest rumors in Argentina," with head *rumors*):

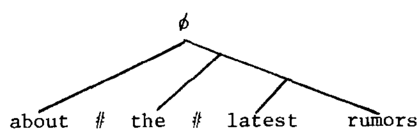


Figure 7 shows how, at the prosodic level, the syntactic VP "has been avidly reading about the latest rumors in Argentina" is split up into ϕ-phrases. On top of the figure we show the syntactic constituent structure of the VP, in terms of X-bar notation, at the bottom we show its prosodic

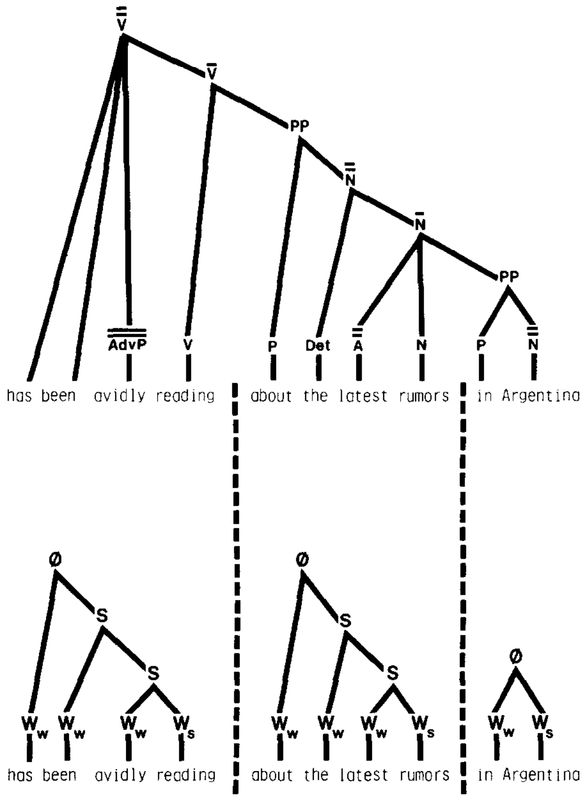


FIG. 7. A VP split into its ϕ-phrases. The top part shows the syntactic constituent structure of the VP (in bar-notation); the bottom part shows its prosodic structure. In the prosodic diagram, ϕ = a ϕ-phrase, W = a prosodic word, w = prosodically weak element, S = prosodically strong element.

structure (with markings for prosodic word (W), and weak (w) and strong (S) prominence relations for those familiar with Selkirk's theory). Notice that we run up to the head of the VP (called $\bar{V}$ in the diagram)—"reading"—and make this material a ϕ-phrase. Then, ignoring "about," since it is a preposition, we run up to the next head of a phrase, the head of the NP complement to (following) the preposition "about" (called $\bar{\bar{N}}$ in the diagram, i.e., "the latest rumors in Argentina" with head "rumors") and make this material a ϕ-phrase. This leaves the PP "in Argentina," which is made a ϕ-phrase, since "Argentina" is the head of the NP following "in" (in this case the NP just is "Argentina").

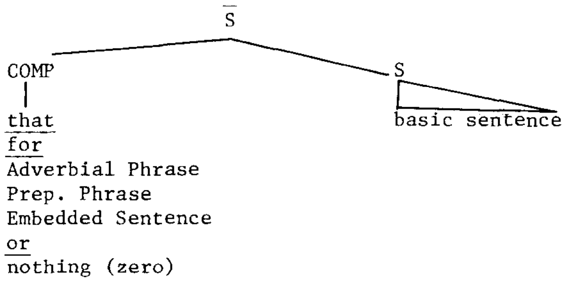
It is important to note that the ϕ-phrase need not be isomorphic to any constituent of syntactic structure. For example, in the prosodic structure in Fig. 7 (bottom), *about the latest rumors* is not a syntactic constituent of the VP at the top of Fig. 7, nor is *has been avidly reading*. But, as Selkirk points out, there is one important syntactic generalization about ϕ-phrases which may be important to a theory of speech processing. Every ϕ-phrase ends in the head of a syntactic phrase. The structuring of an utterance into ϕ-phrases thus produces information about "headhood," information that is crucial in understanding the syntax and semantics of a sentence.

Φ-phrases are, of course, not the end of the line. They have to be assembled into larger units. These units are called "intonational phrases," hereafter referred to as "I-phrases." Selkirk assumes that the choice of just which or how many ϕ-phrases go into an intonational phrase is free. In order to account for real data, however, we obviously have to give some procedures for combining ϕ-phrases. In our prosodic algorithm for pausing, which we give below, we will do so.

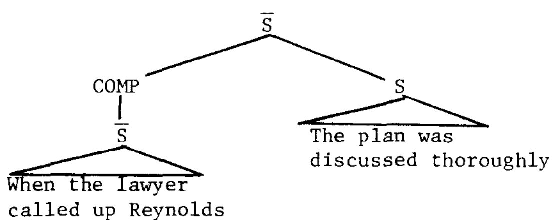
It is important to step back a minute and point out that the various units of prosodic theory (foot, word, ϕ-phrase, I-phrase) are put forth as part of a competence theory of linguistic structure. They are posited on grounds internal to the structure of the language. For example, Selkirk claims that each unit is the domain of application of various phonological rules. We have taken these units from current linguistic theory precisely because we want to test whether they can account for actual empirical data. We hope to suggest a convergence of theoretical and experimental concerns. The theory is motivated in the first case by technical considerations from linguistic theory (which may in the end turn out to be valid or not). But we hope to give it independent motivation by showing it to be a successful predictor of actual empirical data, as well as suggestive for further work in the theory of speech processing.

b. The PHI algorithm. In constructing our algorithm we used devices from current linguistic theory, some of which turn out to have little or no impact on the predictions we make. This is not surprising as they often involve rather minute considerations that do not turn up too often in any small sample of data. For example, once one "corrects" for function words, whether one uses traditional phrase structure notation (as in Chomsky, 1965) or bar-notation (as in Jackendoff, 1977) does not make a lot of difference over our 14 sentences. Though in cases where the two theories make different predictions, the bar-notation theory turns out usually to be the better. But we wanted by and large to use the "best guesses" from linguistic theory we could, if only to ensure that if we failed, our failure would not be attributable to antiquated views of sentence structure. Furthermore, our algorithm was constructed by a linguist in advance of knowing what would have much empirical impact or not. Thus, we here give our algorithm in its base form, leaving some technical "niceties" of linguistic theory to the footnotes. They may, of course, eventually turn out to be quite relevant over a larger span of data.

To leave our algorithm in a relatively simple form, we will take a moment here to define some basic notions. A sentence is made up of a basic sentence (= S), which is sometimes preceded by certain sorts of material. This preceding material might be, for instance, a complementizer ("that," "for," "who," "while," etc., if the sentence is embedded in another one), as in the case of the word "that" in front of "John told a lie" in the utterance "that John told a lie is amazing." Or it might be an adverbial phrase or a prepositional phrase, as, for example, the prepositional phrase "in addition to his files" in the utterance "In addition to his files, the lawyer brought a bookcase." There are other possibilities as well. Let us call a basic sentence together with any preceding material (of course, there may be none) an $\bar{S}$ ("S-bar"), where $\bar{S}$ is made up of a node "COMP" that dominates the preceding material and a node "S" which dominates the basic sentence:



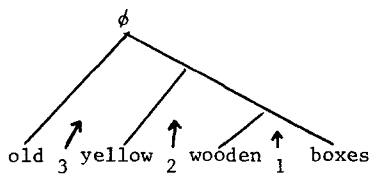
This $\bar{S}$ can be an utterance on its own (as in "In addition to his files, the lawyer brought a bookcase") or embedded in the COMP of another $\bar{S}$, as in:



Of course, an $\bar{S}$ can be embedded in another utterance as either its subject or object, as in "*That John told a lie* is amazing" or "I think *that John told a lie*" (and they can be "extraposed," as in "It's amazing *that John told a lie*" or "It is clear *that John told a lie*").

The algorithm we will give below works on one $\bar{S}$ at a time (if COMP is empty, i.e., nothing precedes the basic sentence (S), then we just have the structure $\bar{S}$ directly over S). It first works on material in COMP (material like preceding adverbial phrases or subordinate clauses; of course, complementizers, like "that," are function words and thus end up "glued" to the item following them), if there is any, and when this material is bundled into ϕ-phrases and I-phrases, it moves on to S. It does not move to the next $\bar{S}$ in the sentence, if there is one, until it has finished with S and then bundled COMP and S together under the $\bar{S}$ it started with. The algorithm works left to right, essentially outputting a word at a time until it has a ϕ-phrase. It then stops and builds a tree for this ϕ-phrase. It moves on, outputting more words and making them into ϕ-phrases, all the while bundling the ϕ-phrases into larger prosodic units (I-phrases) when it can. When it reaches the end of COMP it stops and bundles all its material together, and then moves on to S. Only after it has finished S and bundled the material in S together with that in COMP, does it move to the next $\bar{S}$. That is, it stops at clause (S or $\bar{S}$) boundaries. We end up with a whole prosodic representation, in terms of ϕ-phrases and I-phrases, of the utterance.

We need a method to count the complexity of a boundary, so that we can compare the predictions our algorithm makes with those studied earlier (the more complex a boundary, the longer the pause we predict). At the level of the ϕ-phrase we use essentially the method "CI" of Grosjean et al. (1979): the numerical value of a boundary between two words is the number of branching nodes dominated by the node dominating this boundary, including in the count the word boundary node itself (boundaries containing # do not count as branching for CI values). For example, in the ϕ-phrase below we draw arrows pointing to the word boundary nodes and give the numerical value of the boundaries as determined by CI:



Our algorithm could be construed in two ways. It could either be seen as going through a fully constructed syntactic tree word by word. Or it could be seen as going through a string of words one by one left to right constructing structure (a prosodic tree) by making guesses about which words are heads and where syntactic boundaries are. Such guesses are in fact not difficult (though of course not fail-proof either). For example, the head of a NP is always a noun, and a noun in the singular or plural form; any other noun in a NP is marked with possessive " '*s*" or is preceded by a preposition (and so is the head of a smaller NP inside the bigger one). Knowing what other phrases a noun or a verb subcategorizes for in the lexicon allows good guesses as to the boundaries of NPs and VPs, and of course, after we have processed an initial NP it is a good guess that it is the subject of S and preceding a VP. However, for our purposes here, it is easier to talk in terms of the first construal. Nonetheless, it is the fact that our algorithm, which already works left to right, could be rephrased in terms of the second construal (by building in the proper heuristics) that we hope will give it the future potential to be used in actual models of speech processing. We postpone further discussion of the left to right nature of our algorithm until after we have presented it.

Since the algorithm starts by building or outputting ϕ-phrases, we call it the "PHI algorithm" (of course, a more sophisticated algorithm would also build foot and word prosodic structure). We present the algorithm below.[5]

THE PHI ALGORITHM

We start with a string of words that has its syntactic structure (heads and phrase boundaries) marked. In addition, we assume that a "#" has been placed between any function word and its sister constituent (i.e., between a preposition and its following NP, between a subject pronoun and the auxiliary or verb following it, between an auxiliary verb and the next auxiliary or verb following it, between an object pronoun and the verb preceding it, between a complementizer and the S following it (it attaches to the first word of the S), and so forth). For an example we give below a string of words (our G2) with some of its phrase boundaries marked, and with its nonfunction word heads marked with "H". Function words lose their word status and so cannot count as heads for the purposes of prosody. We should note that it is perfectly easy to add a step to the algorithm that would cause the algorithm to place "#" boundaries itself. But to avoid clutter we will not do so.

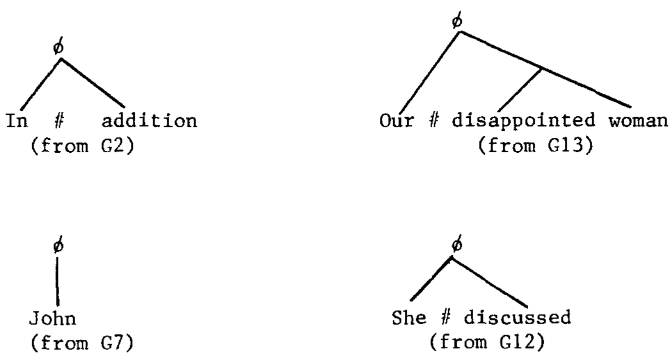
```
     H (noun)                H (noun)
     |                       |
[In # addition [to # his # files]PP]PP


     H (noun)       H (verb)                          H (noun)
     |              |                                 |
[the # lawyer]NP   [brought [the # office's best adding-machine]NP]VP
```

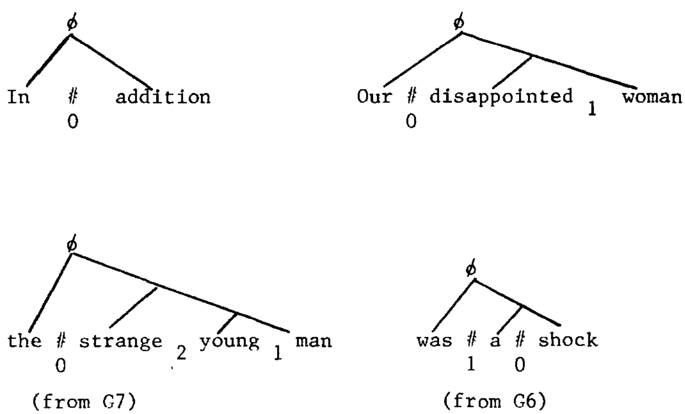
The algorithm has the following steps:

[5] As is true of much linguistic work in prosodic theories, some of the insights around which PHI is based were anticipated by Martin's (1972) groundbreaking paper.

STEP ONE: Start with the initial $\overline{S}$ in the utterance (it may be the only $\overline{S}$ in the utterance, or it may be embedded inside another $\overline{S}$). Output the words of the utterance one by one left to right. Stop at any head of a phrase (marked in our example with H, but one could build in heuristics at this point) and make all the material output thus far a φ-phrase. After the construction of each φ-phrase, check the following steps. Examples:

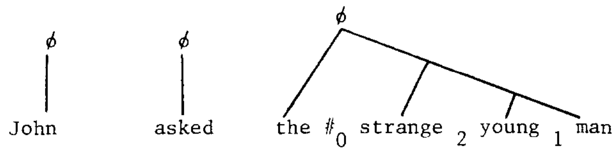


STEP TWO: Assign the value 0 to any boundary containing a # (unless this boundary dominates another such boundary, in which case assign it the value 1). To compute the values of all other boundaries in the φ-phrase use CI. Examples:



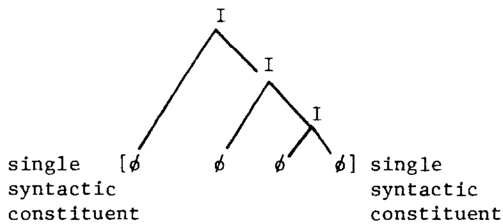
STEP THREE: Continue outputting words, stopping at each head of a phrase, making all the material since the last φ-phrase, together with this head, a φ-phrase. That is, continue outputting φ-phrases. As φ-phrases are produced or output, repeat step 2 above, which assigns values to the boundaries in the φ-phrase. As step 3 operates it produces φ-phrases. These must be bundled together into larger units, namely into

"I-phrases." This is done according to the procedures laid out below in the three substeps 3A, 3B, and 3C. Each substep represents a special principle (3A represents the integrity of the syntactic phrase, 3B the pivotal role of the verb, and 3C bundles what is left over). Use these substeps whenever their conditions of application are met. Example:



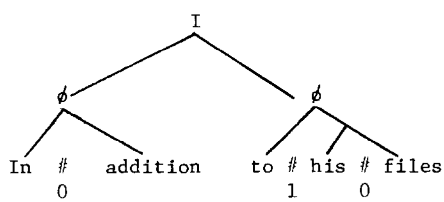
(Output ϕ-phrases, repeating steps 1–3, each time checking steps 3A, 3B, and 3C below to see if and how ϕ-phrases can be bundled together.)

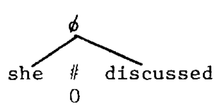
STEP 3A (The Syntactic Constituent Rule): All ϕ-phrases that are part of the same syntactic constituent, excluding the VP (see 3B), must end up bundled together as one larger unit. Do so by right branching and adjoining all the ϕ-phrases in a syntactic constituent under successive I nodes, as:

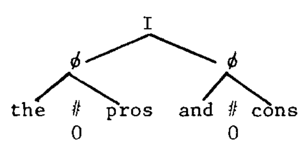


(The motivation for this right branching is simply that it follows closely the branching in the syntax itself—and all things being equal we want our prosodic theory to match up in a simple way with our syntactic theory. Furthermore, material preceding a head is more tightly bound, syntactically and semantically, to the head the closer it is to the head, and this is also represented in our right branching structures here.)

Examples of the application of 3A:

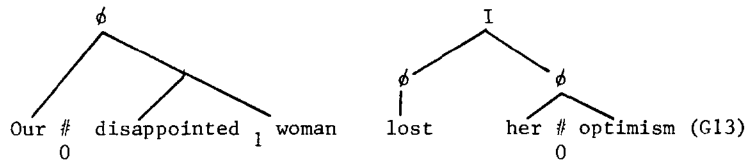




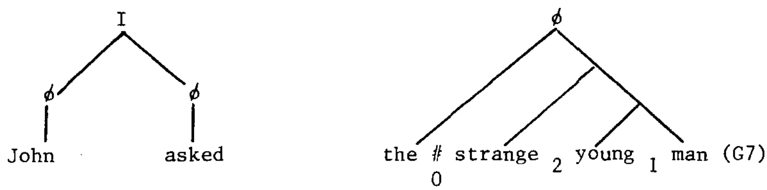


(from G12)

STEP 3B (The Verb Rule): In the typical case the ϕ-phrase headed (terminated) by the verb (e.g., "has been avidly reading") is bundled with the following prosodic unit (ϕ- or I-phrase) in the VP to make up a prosodic unit (I-phrase) that respects the integrity of the VP (for us, only subcategorized complements of the verb make up a VP with the verb; adverbs like "thoroughly" in "discuss the matter thoroughly" are outside modifiers of the basic VP "discuss the matter"). There is only one case where the verb does not pattern with what follows it. If the combination of the verb ϕ-phrase and the preceding prosodic unit, usually the subject, is a simpler prosodic unit (has less branches) than the prosodic unit following the verb, then the verb and this preceding prosodic unit together make up a larger prosodic unit (I-phrase). Thus, if the subject and verb together is a simpler prosodic unit than the unit following the verb (e.g., the direct object phrase), then the verb patterns with the subject—otherwise it patterns with what follows and respects the VP. The motivation for this principle is given below (after the presentation of the algorithm). Examples:

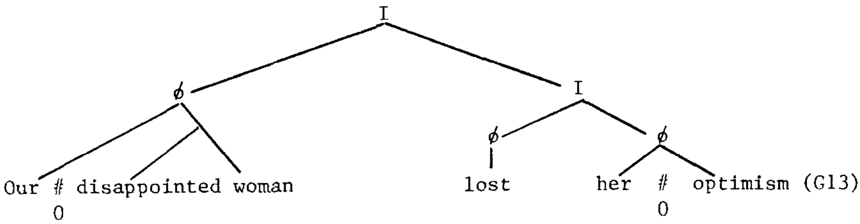


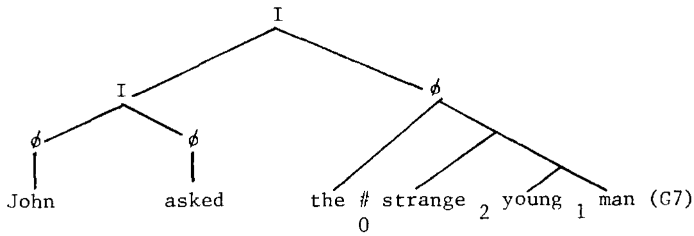
(Verb patterns with object)



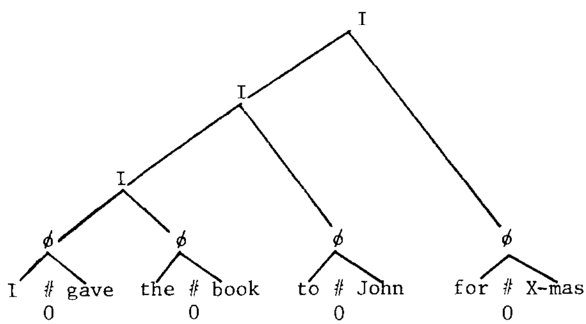
(Verb patterns with the subject)

STEP 3C (The General Bundling Principle): For any string of ϕ- or I-phrases not already bundled by principles 3A and 3B, bundle these together under I. (We will assume that when there are more than two units to be bundled together, the bundling is done by *left branching* and adjunction under I. The reason for this is simply that this allows us to say that if a ϕ-phrase is output and does not fall under 3A or 3B, then it is simply added as a right branch under I to the previous prosodic unit. In fact, however, there are never more than two unbundled units left over to be bundled in our data.) Examples:





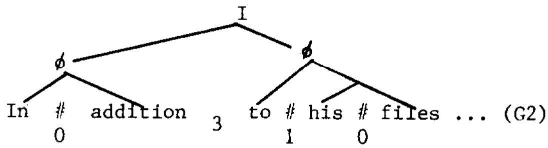
(In fact, both these examples follow also from 3A,
since S is a syntactic constituent.)



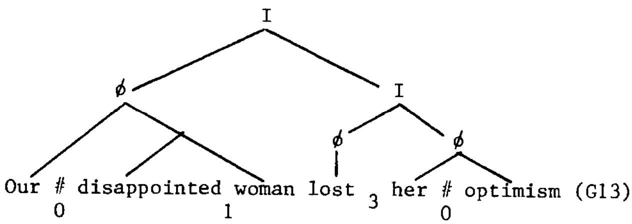
(Predicted, but not tested)

STEP FOUR: To compute the numerical value of a boundary immediately dominated by I, count the number of ϕ's and I's this I dominates, including in the count this I node itself (this procedure, of course, operates any time it can, as ϕ-phrases are output left to right and bundled by the preceding steps).[6] Examples:

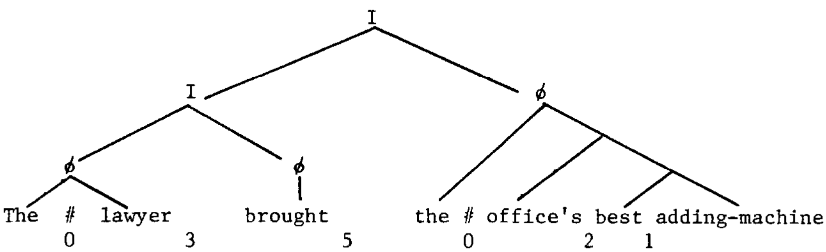
[6] Notice that PHI limits the application of CI (to step 2) which is all to the good since CI values peak very fast as one moves further up the tree. However, the correlations between PHI and the data which we present below are little changed if one uses CI all the way up the tree.



(Must stop, since an S boundary has been reached, i.e., this is the end of COMP in G2.)

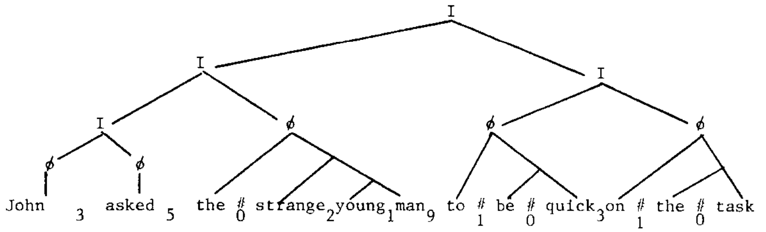


STEP FIVE: Go on to the following S (having finished a COMP) or $\overline{S}$ (having finished a preceding $\overline{S}$). Thus, for example, after completing "In addition to his files" for sentence G2, we go on to S and ultimately derive:[7]



STEP SIX: Bundle an I-phrase in COMP with the I-phrase in S (this follows from 3A in fact). Bundle all remaining I's of the utterance by left branching and adjunction under I (this follows from 3C in fact). Use step 5 above to compute the numerical value of a boundary immediately dominated by I. Thus, in the example above, "In addition to his files" will bundle with "the lawyer brought the office's best adding machine" under I. The boundary value immediately dominated by this I would be 9. Another example of a prosodic tree for a whole utterance follows:

[7] "The office's best adding-machine" should, perhaps, have been assigned the structure: [[the office's]ϕ[best adding-machine]ϕ]$_I$. This would in fact have made our predictions come out even better. However, we do not now know how full NP determiners like "the office's" behave prosodically.



(G7)

The six steps above form the basic part of PHI. Originally, when we constructed it we wanted also to consider the effects of sentence stress and complex, multifooted words on pausing. It has repeatedly been suggested that boundaries before sentence stress and before and after complex words have added complexity and that this is reflected in temporal data (see for example, Selkirk, 1980; Cooper & Paccia-Cooper, 1980). We hypothesized two rules that would reflect such factors. The rules are as follows:

STEP SEVEN (Complex Word Rule): Add 1 to the value of any boundary before or after a complex word (provided this boundary is flanked by two items in the same ϕ-phrase). A complex word is a word made up of two or more feet (obviously this is only a rough approximation). For example, in the sentence "In addition to his files, the lawyer brought the office's best adding-machine," the PHI value between *best* and *adding-machine* is raised from 1 to 2. And in the sentence, "She discussed the pros and cons to get over her surprisingly apprehensive feelings," the PHI value between *surprisingly* and *apprehensive* is raised from 2 to 4 (1 for ech of these two complex words) and the value between *apprehensive* and *feelings* is raised from 1 to 2 (multifooted words have a primary stressed syllable and at least one secondary stressed syllable).

STEP EIGHT (Sentence Stress Addition): Add 1 to any boundary before a sentence final word and its associated function words (attached by #) provided this word bears unmarked sentence stress.[8] For example, the value before *adding-machine* in G2, before *feelings* in G13, or before *thoroughly* in G1 is upped by 1. The values in the prosodic tree given as an example in step 6, then, are final ones, save for the 3 before "on the

[8] In the Appendix we mark with "´" any item in our 14 sentences which is considered to have unmarked (basic, nonemphatic) sentence stress, see Ladd (1980) for a full discussion. Unmarked sentence stress was marked on any full lexical item (which was not a locative or temporal modifier) that terminated both a main clause and the utterance as a whole.

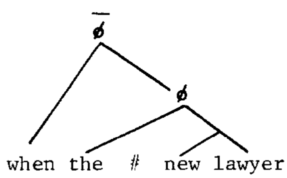
task." This is raised by Sentence Stress Addition to 4, since "task" carries unmarked sentence stress in this sentence.

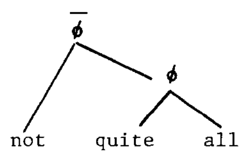
Steps 7 and 9 can make small differences in individual sentences, and whenever we have tested them, whether on the data below or elsewhere, they have usually made the right prediction. But they apply in our data here so few times that across all 154 boundaries in the data, whether or not we use them makes no statistical difference. Thus, for our purposes here, it is the construction of ϕ-phrases and the bundling of these into I-phrases that is really crucial.

The basic shape of our algorithm is derived from prosodic theory, with the essential insight being that there is a unit between the level of the word and the syntactic phrase. In some cases, however, prosodic theory does not determine any particular choice as to how we should proceed, and here we use additional sources of motivation which have invariably come from discourse theory. The left branching in steps 3C and 6 was chosen because work on discourse structure has repeatedly argued that material in a sentence gets more important communicatively (is newer and more foregrounded) as we progress toward the end of the sentence (see, e.g., Daneš, 1974; Firbas, 1962). Thus, we place such material higher in the tree (dominating the material to the left) and thereby place a larger break in front of it.

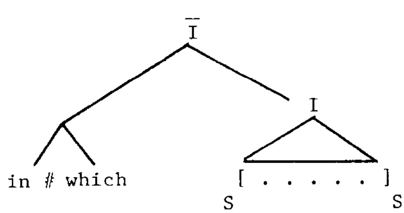
The bundling principle in 3B, The Verb Rule, is perhaps the most interesting case. First, the previously mentioned work on discourse has argued that the verb can either pattern with the older information which is typically in front of it or with the newer information which is typically after it. It is the pivot around which the information structure of the sentence is organized. Second, data from child language indicates that in some cases the verb is a cohesive unit with the object (VO) and in other cases a cohesive unit with the subject (SV), depending upon the semantic and syntactic properties of the verb (see Lempert & Kinsbourne, 1979). Third, the verb has a somewhat ambivalent role syntactically. It is usually considered the head of the VP, but some linguists have also argued that it is the "head" of the sentence as a whole (see Jackendoff, 1977). Finally, work in psycholinguistics on a wide variety of tasks has also suggested relative freedom in the patterning of the verb (e.g., Suci, 1967; Martin, 1970; Grosjean & Collins, 1979; Levelt, Note 3). We thus hypothesize that, given its pivotal role, the verb is fairly free to reflect rhythmical factors, and 3B is our best guess as to how it does so. But our motivation has already suggested that which way it patterns may well also be influenced by semantic and discourse factors as well, especially in reading in context or in spontaneous speech.

Finally, in this paragraph we will append a brief note of only technical linguistic interest (but a matter which does affect the values at some of our boundaries), before turning to a point of much more than technical interest. Our technical point here does have implications both for linguistic theory and for psychologists who would like to design prosodic models. We initially developed a system to treat material in COMP based on the fact that COMP can contain three rather different sources of material: full phrases, which often have their own intonational contour ("In addition to his files, the lawyer"), simple complementizers (e.g., *that, for, since, when, who*), and, finally, wh-phrases that have carried along ("pied–piped") other material (e.g., "in which"). The first sort of COMP (full phrase or clause) is always treated by our algorithm as a unit on its own and is bundled according to the principles outlined above. Sentence initial complementizers and other function words are counted as "upbeats" to the ϕ-phrase adjacent to them and which they are adjoined to (we use the notation $\bar{\phi}$ for this). For example:





Both $\bar{\phi}$ and ϕ in the above structures are counted as ϕ's in our algorithm. Finally, the full interrogative phrases in COMP were treated as upbeats to I's (because they are intermediate in complexity between the first two cases), e.g.:



(e.g., they offered numerous tours)

Such decisions, we must admit, turn out to make little difference for our algorithm, though in another context we hope to motivate their linguistic interest (for example, in predicting where simple complementizers can and cannot be deleted).[9]

Finally, though, there is a quite important point to be made about the prosodic structures our algorithm produces. Such structures, leaving aside any overt making of syntactic structure, encode and integrate much more information than one traditionally associates with rhythmical structures. For example, since ϕ-phrases always terminate on the heads of syntactic phrases, our structures encode the location of heads of phrases. Such information is obviously of semantic import as well, since the semantics of a phrase is determined by and organized around its head. Further, Selkirk argues that each prosodic unit is the domain of operation of various phonological rules, and thus relevant to the workings of the phonological component. We have already indicated the possibility that various properties of our prosodic structures encode or reflect discourse-structure properties. And the placement of I-phrases are relevant to what can and what cannot bear a separate intonational contour, a property that is obviously relevant to semantics and discourse (we have not, however, given the rules for determining which I-phrases in spontaneous speech would actually carry a separate intonation, though in many cases it is just the top one that does so). Finally, we suggest below that our prosodic structures also reflect properties of the logical form of sentences. This suggests the possibility that such structures are not just prosodic structures, but really a basic linguistic structure, perhaps the only one, or at least the critical one, in processing. At any rate, the fact the PHI inte-

[9] It should be noted that correlations with the data are virtually the same if one does not follow our treatment of COMP, but instead just counts complementizers as function words, perhaps immune to vowel reduction, output as a single ϕ-phrase with the following material, and computes by CI all the way up the tree. Some additional assumptions we make impinge on particular analyses in current work in syntax and semantics. A good deal of recent work on syntax has argued that certain $[\mathrm{V\ PP}]_{\mathrm{VP}}$ structures restructure as $[[\mathrm{V\ P}]_{\mathrm{V}}\ \mathrm{NP}]_{\mathrm{VP}}$ (Hornstein & Weinberg, 1981; Chomsky, 1980a, 1981). Thus, in such cases, we apply both analyses and average their values (e.g., *wondered about this extraordinary story* = *wondered* (*about this extraordinary story*) and *wondered about* (*this extraordinary story*)). In the case of one sentence, sentence G14, we average two different analyses for two parts of the sentence: *since she# was indecisive* vs *since she#was#indecisive*, and *her#friend asked#her to#wait* vs *her#friend asked#her#to#wait*. The first case is transparent. Either *she* cliticizes to *was* and comes out as a ϕ-phrase with it, or *was* cliticizes to *indecisive* and therefore the subject pronoun has to output with this unit as a ϕ-phrase. The latter case is a bit more interesting. It is probably the case that we must build into steps 2 and 3 a device to lower the value at the boundary between two words which have undergone vowel reduction (Cooper and Paccia-Cooper's BS builds in a similar device). The averaging approach in fact looks plausible, but we do not have enough evidence to make any concrete proposals.

grates phonological (e.g., information about #-marked boundaries), syntactic, and semantic/discourse information, as well as rhythmical information, in one structure turns out, we believe, to be a key reason for its success, which we survey below.

Algorithms are not in themselves interesting unless they are steps on the way to a model of the human mind. PHI is based on a prosodic theory that is meant to be a model of the native speaker/hearer's linguistic competence. Furthermore, it incorporates certain constraints that make it begin to approximate a true performance model. PHI works largely left to right, but not wholly so. We do not want a model to be wholly left to right as this would be tantamount to a claim the people do not have any expectations or make any predictions about what is to come (and, after all, they do sometimes go down garden paths). But ideally, we want these expectations to be locally and quickly resolved.

PHI does not need a whole phrase structure tree for a sentence in order to begin bundling its words into prosodic units. It needs, for the most part, only lexical and morphological information to bundle words into ϕ-phrases. To bundle these into I-phrases by Rule 3A (the Syntactic Constituent Rule) it needs to know only whether the *lowest level* constituent it is working on at the time is terminated or not; anything above and beyond this is irrelevant. And often the lexical specification for a noun or verb gives us a good idea of what sort of structure will or will not follow it. For step 3B (the Verb Rule) we do not need even this much top-down information, but need to know only whether a sentence has a "short subject," so to speak, and whether the verb is followed by a subcategorized (required) complement (information that comes from our lexical knowledge). 3C (the General Bundling Principle) requires no top-down information, as it simply adds units incrementally to what has already come before and been built up.

Finally, PHI needs to know where clause boundaries are. This information is often quite locally specified, signaled, for example, by the presence of a complementizer, infinitive marker, or conjunction. In addition, the presence of an embedded clause is often signaled by the lexical requirements of the verb. For example, in a sentence like G7, "John asked the strange young man to be quick on the task," our lexical knowledge that the verb "ask" requires (or at least heavily favors) an infinitival clause after its object tells us that there is a clause boundary after "man" and before "to" (and this is also signaled by the occurrence of "to" as well). This clause boundary will stop the algorithm and predict the presence of a large break after "man" (the biggest in the sentence if "ask" has only its minimal lexical specifications), since the algorithm requires all preceding material to be a finished higher order unit, and starts anew on the following material.[10]

Thus, for PHI to be turned into a more realistic model, we would have to add quite local, usually lexically determined, devices to guess boundaries. Its left to right operation needs no more supplementation than this. And, in fact, just such devices follow naturally from the competence and processing model ("lexical–functional grammar") developed recently by Bresnan and Kaplan (see Bresnan, 1982).

How well does PHI predict the pause data in our 14 sentences? In Fig. 8 we present the performance structure of sentence G11, as well as the structures predicated by the GGL, the CPC, and the PHI algorithms. The structure for the PHI algorithm is obtained simply by taking the numerical values assigned by PHI and branching by the iterative procedure we discussed in Section 1 (i.e., find the smallest value and cluster the two elements separated by this value by linking them to a common node: cluster the two elements (words or clusters) separated by the next smallest value, and so on, until a tree is built for the whole sentence). This procedure simply has the effect of lowering the nodes dominating function words, i.e., dominating boundaries containing a #, and assigned 0 in the PHI algorithm, thus making PHI comparable to CPC and GGL, as well as to performance structures. If we had liked, we could have produced such trees directly in our algorithm, by having boundaries assigned a # lowered at step 1 of the algorithm. At that point we chose to leave uniform right branching in the ϕ-phrase because this is in fact how it is done in the prosodic theories we have borrowed from.

As we saw earlier, the GGL algorithm (see the left-hand tree in Fig. 8) is a good predictor of the pause data ($r = .89$). The main break is located correctly (after *book*) as are the second main breaks (after *agent* and after *in which*). But the height of the various nodes is not always what it should be. Values between *agent* and *consulted,* between *consulted* and *the agency's book,* between *in which* and *they,* and between *offered, nu-*

[10] The numerical values assigned to the largest break in each of our 14 sentences are actually somewhat arbitrary. It turns out (importantly) that the actual pause duration of the longest pause in each sentence does not correlate all that well (is not a factor of) the overall length of the sentence (for example, it is possible for a short, less complex sentence to have a longer main break that a longer, more complex sentence). In fact, our predictions would have been even better had we simply assigned an arbitrary high value to the main break in each sentence. All the numerical values in the sentences that are less than the highest one turn out to be nearly perfect predictors of the actual weight of a boundary vis-a-vis the other boundaries in the sentence and across all the sentences. Thus, it is only at the main break that our algorithm needs to know how much material follows in the sentence as a whole (because it is only there that the lowest constituent it is working on is the sentence as a whole), and it is only here that the actual numerical values appear somewhat arbitrary.

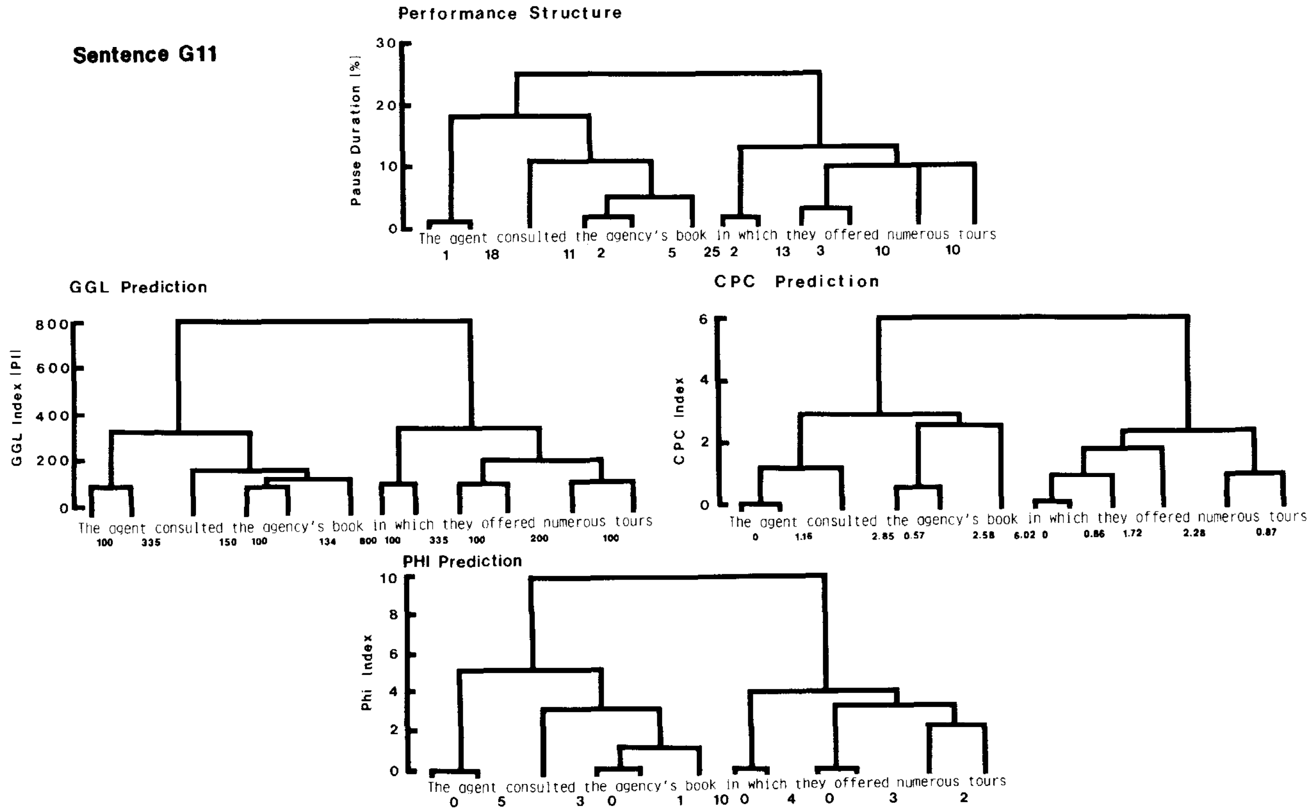


FIG. 8. Four hierarchical structures for sentence G11: the performance structure based on percentage pause durations and the performance structures predicted by the Grosjean et al. (1979) algorithm (GGL), the Cooper and Paccia-Cooper (1980) algorithm (CPC), and the PHI algorithm.

merous, and *tours* are all too low. Hence the good but not perfect correlation between the data structure and the predicted structure. Turning to the prediction made by the CPC algorithm (right-hand tree in Fig. 8), we note that it does less well ($r = .71$). Not only does it have problems predicting the second main breaks (after *agent* and after *in which*) which are given values that are much too low, but it also gives too much importance to minor breaks such as after *agency's* and after *they*. This, as we saw earlier, is due to the one-cycle bisection component, which puts too much weight on the middle of the sentence and not enough on its outer parts.

In comparision to these two algorithms, the PHI algorithm (bottom structure in Fig. 8) produces an almost perfect copy of the performance tree. Not only are all branches correct (except for the last one), but above all the nodes are of the right height. Note for instance the height of the node dominating *consulted* and *the agency's book* or that dominating *they offered* and *numerous tours*. The only minor mismatch is the height of the node dominating *the agent* and *consulted:* It is slightly too low in the PHI tree. This nearly perfect match between the actual performance structure and the predicted structure is reflected in the correlation coefficient: .98.

As can be seen in Table 1 (see Section I), the PHI algorithm does extremely well on all 14 sentences. The mean correlation is .97 (GGL obtains a mean of .86 and CPC a mean of .78) and the global correlation is .96, as compared to .83 for GGL and .75 for CPC. Both differences are highly significant: $t = 10.8$, $p < .01$ and $t = 14.66$, $p < .01$, respectively. What is especially striking about the PHI algorithm is that it predicts every sentence equally well: The range for the 14 sentences goes from .93 to .99, a .06 difference, whereas the other two algorithms have much larger differences: .33 for GGL and .29 for CPC.

The PHI algorithm is not only the best predictor of the 14 sentences, it is also a very fine predictor of other pause data. Using a magnitude production technique, Grosjean (Note 6) obtained pause values for the Pop Fan passage. This is a connected piece of discourse that has two sentences, eight clauses, and 44 word boundaries. The values obtained with the PHI algorithm are correlated .94 with the 44 pause durations, whereas the GGL values and the CPC values are only correlated .80 and .73, respectively. The difference between the PHI correlation and the other two correlations is highly significant: $t = 4.09$, $p < .01$ and $t = 5.61$, $p < .01$, respectively.

Although primarily a prosodic algorithm, PHI is also a good predictor of the parsing data obtained by Grosjean et al. (1979). The global correlation is .93, the mean of the 14 correlations is .94, and the range extends from .86 to .98. But such good prediction of the parsing data should not come as a total surprise as Grosjean et al. (1979) found a .92 correlation between pausing and parsing values. In fact, the difference between pausing and parsing seems to be due almost entirely to the most motor-level and stress-related aspects of PHI, i.e., the operation of the procedure placing "#" in a boundary and addition to boundary values due to the Complex Word Rule (step 7) and Sentence Stress Addition (step 8). Abstracting from these, parsing and pausing appear to give virtually identical results, suggesting that our "prosodic model" is really just a linguistic model per se, one which captures in a single system information related to each level of the grammar. In fact, the main characteristic of the PHI algorithm that explains why it accounts for 92% of the pause variances of the 14 sentences (as opposed to 69% for GGL and 56% for CPC) may well be the fact that it integrates phonological, rhythmic, syntactic, and possibly semantic information, in a constrained way and in terms of one representation.

PHI does extremely well against the data. Nonetheless, it is clear that there are many other factors that may well influence pausing, though in much more minor ways than the factors we have indicated. The phonological segments on either side of a boundary can make a difference. There may be tendencies to isochrony over and above the somewhat balanced units our prosodic model already gives. Particularly rhythmical runs of syllables, alliteration and other sound effects, and two or more weak syllables coming together may all affect the strength of certain boundaries (as they are known to do in poetry). Consider *closing / his cli/ent's book* where the break between *closing* and *his client's* is a bit shorter (6%) and that between *his client's* and *book* a bit longer (6% also) than we might expect. A break between a verb particle and a following NP in the same (lowest) I-phrase should be lowered (e.g., *call up Reynolds*). Clearly, there is a need for an articulated theory of nonlexical words and how they function, for example, how they function in chains (*that she was having a party*) and how they are influenced by different phonological effects and different syntactic and prosodic positions. Gaps (trace and PRO) may influence pausing, as well, perhaps, as the control properties of the verb. The semantics of certain words (e.g., superlatives) undoubtedly affects pausing, as do factors to do with how expected or unexpected a piece of information is. Finally, it is possible that subordinate clauses and presupposed material work differently in regard to pausing at their internal breaks than main clauses in some respects.

The list could be continued, but the point is clear: With the multiplicity of diverse factors that could affect pausing, it is remarkable indeed how well our prosodic algorithm correlates with the data. We take this as an indication that prosodic units at and above the level of the ϕ-phrase, and the bundling principles we have defined, are the key higher order determinates of pausing, and that a theory of prosody is the crucial unifying system in regard to pausing and parsing. In a sense, what we have done is to show that performance structures are quite simply a reflection of prosodic structures.

We can also reach an important conclusion about our data base. Pause data of the sort we have used (from readings at various rates), are well behaved and do reflect underlying systems of linguistic competence. Such data can be used as one sort of inductive base for the development of prosodic theories. Surely, there is no necessity for this to have been the case. Such data could have turned out to be the product of other cognitive systems largely irrelevant to the linguistic faculty itself.

c. PHI and language processing. The ϕ-phrase, and the prosodic structures it is embedded in, may turn out to play a role in language processing. If we look at our data, and ignore for the moment boundaries with a # in them (which invariably have quite small pauses), it turns out that the large majority of the boundaries in the data are ϕ-phrase boundaries. If we can continue to abstract away from #-marked boundaries, and call those units separated by PDs $\geq 8\%$, "basic (small) units" in the data (see Section 1), it turns out that 74% of these units are ϕ-phrases. The ones that are not ϕ-phrases invariably turn out to be cases where a ϕ-phrase has gotten "too long" and has been broken down into smaller units made up of complex, multifooted prosodic words. For example, in "She discussed the pros and cons to get over her surprisingly apprehensive feelings" (G12), the first four "basic units" in the data are ϕ-phrases: (she#discussed), (the#pros), (and#cons), (to#get over). The last ϕ-phrase (her#surprisingly apprehensive feelings) is a long one, with two multifooted words, and is broken down in the data into its prosodic multifooted words: $(\text{her\#surprisingly})_W$ $(\text{apprehensive})_W$ feelings. The speaker can, then, output ϕ-phrases or break complex ϕ-phrases into foot structures and create a rhythmical string out of small units which are still bigger than single lexical items in the vast majority of cases (note that even the prosodic word "her#apprehensive" is bigger than a single lexical item). This may make the work of speech production easier and more efficient. In any event, we believe that the hypothesis that the production system "wants" smallish, rhythmical chunks, bigger than a single lexical item, but smaller than the typical phrase, to be a fruitful path for future research (for an indication that this is true in spontaneous speech, see Chafe, 1982 and Clancy, 1982).

Φ-phrases encode what is probably the smallest bit or bundle of coherent semantic information above the level of the single lexical item. In fact, the information they encode is realized as a single, often morphologically complex, lexical item in many other, less analytic languages than English. Φ-phrases are of three basic sorts: (a) determiners, articles, and such grammatical words, plus modifiers, plus a phrasal head (e.g., "the#young woman")—a tightly bound package of conceptual information that might well be lexicalized as a single lexical item in another language; (b) subject and/or object pronouns plus auxiliary verbs, plus the main verb (e.g., "will#have#been#reading#it")—material that many languages express as a single morphologically complex word via the use of enclitics and/or affixes; (c) and, finally, prepositions plus their complements (e.g., "in#the#house)—material that in case languages is expressed as a single word, i.e., noun + case marker, as in Latin "puellae," "to the girl." Thus, the ϕ-phrase may be a prosodic analog of material that functions on-line much like "big words" (perhaps, the ϕ-phrase is necessitated by the highly analytic character of English—research on morphologically complex languages is crucially needed in psycholinguistics, here and elsewhere).

We have already pointed out that PHI integrates phonological, syntactic, and perhaps even semantic/discourse information in a single representation. Its sensitivity to preposed material (in COMP) and its left branching for nonsubcategorized material (i.e., phrases that are not specified in the lexicon as optional or obligatory complements of the verb, for instance), serve in fact to ensure that topicalized (preposed to the front of S) and focused (at the end of a clause) phrases will be high in the tree. Such phrases will thus have scope over the material beneath them, technically speaking, the material they command (a node A commands a node B if the first branching node above A dominates B). And this is the position that they occupy in all likelihood in a representation of logical form (Jackendoff, 1972; Chomsky, 1970b, 1976, 1980b, 1981; Gueron, 1980). Thus, for example, in the prosodic representation for G6, "That the matter was dealt with so fast was a shock to him," "that the matter was dealt with so fast" is an I-phrase under another I node which also dominates "so fast" (which cannot bundle with the verb since it is not a complement to the verb). This gives "so fast" scope over (it commands) "that the matter was dealt with," a position it has in logical form (for some degree x of fastness, the matter was dealt with x). Thus, prosodic representations may also reflect properties of logical form as well.

Finally, then, the fact that PHI is constructional and left to right, that it could be extended by heuristics to be even less top-down than it is, as well as the fact that it integrates a variety of linguistic factors in a disciplined way, utilizing units that are part of the theory of competence, but clearly relevant to the theory of performance, recommends it as a future candidate to be an integral part of a general model of language processing.

APPENDIX: THE 14 SENTENCES USED BY GROSJEAN, GROSJEAN, AND LANE (1979)

The values at each word boundary correspond to the percent pause duration (%PD) obtained at that boundary.

(G1) When 5 the 0 new 2 lawyer 20 called 3 up 5 Reynolds 30 the 3 plan 10 was 1 discussed 21 thóroughly.
(G2) In 2 addition 10 to 3 his 3 files 33 the 1 lawyer 8 brought 17 the 0 offices 10 best 13 ádding-machine.
(G3) By 0 making 11 his 2 plan 5 known 38 he 0 brought 4 out 19 the 2 objections 16 of 3 éveryone.
(G4) That 5 a 5 solution 15 couldn't 7 be 3 found 30 seemed 9 quite 6 clear 17 to 3 them.
(G5) Not 11 quite 10 all 17 of 4 the 3 recent 9 files 23 were 7 examined 10 that 6 day.
(G6) That 7 the 2 matter 13 was 2 dealt 2 with 18 so 5 fast 28 was 5 a 3 shock 13 to 2 him.
(G7) John 10 asked 17 the 3 strange 8 young 5 man 25 to 5 be 3 quick 19 on 5 the 0 tásk.
(G8) Closing 6 his 5 client's 6 book 26 the 2 young 4 expert 15 wondered 8 about 9 this 11 extraordinary 8 stóry.
(G9) The 3 expert 19 who 3 couldn't 5 see 9 what 7 to 4 criticize 29 sat 5 back 14 in 2 despáir.
(G10) After 5 the 5 cold 5 winter 14 of 3 that 5 year 32 most 5 people 10 were 5 totally 11 fed-úp.
(G110 The 1 agent 18 consulted 11 the 2 agency's 5 book 25 in 2 which 13 they 3 offered 10 numerous 10 tóurs.
(G12) She 4 discussed 15 the 1 pros 8 and 1 cons 25 to 1 get 5 over 8 her 6 surprisingly 16 apprehensive 10 féelings.
(G13) Our 5 disappointed 6 woman 24 lost 8 her 4 optimism 28 since 5 the 1 prospects 15 were 3 too 1 limited.
(G14) Since 11 she 2 was 13 indecisive 13 that 2 day 34 her 2 friend 12 asked 2 her 9 to 6 wáit.

REFERENCES

Bresnan, J. (Ed.) *The mental representation of grammatrical relations*. Cambridge, MA: MIT Press, 1982.
Bresnan, J. On complementizers: Toward a syntactic theory of complement types. *Foundations of Language*, 1970, **6**, 297–321.
Brown, E., & Miron, M. Lexical and syntactic predictors of the distribution of pause time in reading. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 658–667.
Chafe, W. L. Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy*. Norwood, NJ: Ablex, 1982.
Clancey, P. M. Written and spoken style in Japanese narratives. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy*. Norwood, NJ: Ablex, 1982.
Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.
Chomsky, N. *Aspects of a theory of syntax*. Cambridge, MA: MIT Press, 1965.
Chomsky, N. Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English transformational grammar*. Waltham, MA: Ginn, 1970. (a)
Chomsky, N. Deep structure, surface structure and semantic interpretation. In N. Chomsky, *Studies on semantics in generative grammar*. The Hague: Mouton, 1970. (b)

Chomsky, N. *Reflections on language*. New York: Pantheon, 1975.

Chomsky, N. Conditions on rules of grammar. *Linguistic Analysis*, 1976, **2**, 303–351.

Chomsky, N. On binding. *Linguistic Inquiry*, 1980, **11**, 1–46. (a)

Chomsky, N. *Rules and representations*. New York: Columbia Univ. Press, 1980. (b)

Chomsky, N. *On governance and binding*. Dordrecht, the Netherlands: Foris Press, 1981.

Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.

Cooper, W., & Paccia-Cooper, J. *Syntax and speech*. Cambridge, MA: Harvard Univ. Press, 1980.

Daneš, F. (Ed.), *Papers on functional sentence perspective*. The Hague: Mouton, 1974.

Dommergues, J. Y., & Grosjean, F. Performance structures in the recall of sentences. *Memory and Cognition*, 1981, **9**, 478–486.

Firbas, J. Notes on the function of the sentence in the act of communication. *Sbornik praci filosoficke fakulty brenske university 11*, 1962, **A10**, 133–147.

Givon, T. *On understanding language*. New York: Academic Press, 1979.

Goldman-Eisler, F. Pauses, clauses, sentences. *Language and Speech*, 1972, **15**, 103–147.

Grosjean, F., & Collins, M. Breathing, pausing and reading. *Phonetica*, 1979, **36**, 98–114.

Grosjean, F., Grosjean, L., & Lane, H. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 1979, **11**, 58–81.

Grosjean, F., Lane, H., Battison, R., & Teuber, H. The invariance of sentence performance structures across language modality. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, **7**, 216–230.

Guéron, J. On the syntax and semantics of PP extraposition. *Linguistic Inquiry*, 1980, **11**, 637–678.

Handel, S., & Todd, P. Segmentation of sequential patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 1981.

Hillinger, M., James, C., Zell, D., & Prato, L. The influence of prescriptive and subjective phrase markers on retrieval latencies. *Bulletin of the Psychonomic Society*, 1976, **8**, 353–355.

Jackendoff, R. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press, 1972.

Jackendoff, R. *$\overline{X}$ Syntax: A study of phrase structure. Linguistic Inquiry Monograph Two*, Cambridge, MA: MIT Press, 1977.

Johnson, N. The psychological reality of phrase structure rules. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 469–475.

Johnson, S. Hierarchical clustering schemes. *Psychometrika*, 1967, **32**, 241–254.

Ladd, R. *The structure of intonational meaning: Evidence from English*. Bloomington, IN: Indiana: Indiana Univ. Press, 1980.

Lempert, H., & Kinsbourne, M. Action as substrate for early syntax. In P. L. French (Ed.), *The development of meaning*. Hiroshima, Japan: Bunka Hyoron, 1979.

Liberman, M., & Prince, A. On stress and linguistic rhythm. *Linguistic Inquiry* 1977, **8**, 249–336.

Martin, E. Toward an analysis of subjective phrase structure. *Psychological Bulletin*, 1970, **74**, 153–166.

Martin, J. G. Rhythmic (Hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 1972, **79**, 487–509.

Prince, E. Toward a taxonomy of a given—New information. In P. Cole (Ed.), *Radical pragmatics*. New York: Academic Press, 1981.

Selkirk, E. O. The role of prosodic categories in English word stress. *Linguistic Inquiry*, 1980, **11**, 563–605.

Selkirk, E. O. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press, forthcoming.

Suci, G. The validity of pause as an index of units in language. *Journal of Verbal Learning and Verbal Behavior,* 1967, **6,** 26–32.

Suci, G., Ammon, P., & Gamlin, P. The validity of the probe-latency technique for assessing structure in language. *Language and Speech,* 1967, **10,** 69–80.

Weinberg, S., & Goldberg, K. *Basic statistics for education and the behavioral sciences.* Boston, MA: Houghton Mifflin, 1979.

REFERENCE NOTES

1. Liberman, M. *The intonational system of English.* Unpublished Doctoral Dissertation, Cambridge, MA: MIT (Distributed by the Indiana University Linguistics Club), 1975.
2. Selkirk, E. O. *On Prosodic structure and its relation to syntactic structure.* Paper presented to the Conference on Mental Representation in Phonology (Nov. 1978), sponsored by the Sloan Foundation (manuscript, Dept. of Linguistics, University of Massachusetts at Amherst).
3. Levelt, W. *The perception of syntactic structure.* Heymans Bulletin No. HB-69-31B, Groningen, The Netherlands: Groningen University, 1969.
4. Dommergues, J. Y. *A comparison of two algorithmic predictors of performance structures.* St. Denis, France: working paper, University of Paris VIII, 1980.
5. Selkirk, E. O. *The phrase phonology of English and French.* Unpublished Doctoral Dissertation, Cambridge, MA: MIT, 1972.
6. Grosjean, F. *Le rôle pové par trois variables temporelles dans le comprehension orale de langlais.* Unpublished Doctoral Dissertation, University of Paris VIII, 1972.

(Accepted April 20, 1983)