A Computational Grammar of Discourse-Neutral Prosodic Phrasing in English

J. Bachenko and E. Fitzpatrick

AT & T Bell Laboratories
Murray Hill, NJ 07974

We describe an experimental text-to-speech system that uses information about syntactic constituency, adjacency to a verb, and constituent length to determine prosodic phrasing for synthetic speech. A central goal of our work has been to characterize "discourse neutral" phrasing, i.e. sentence-level phrasing patterns that are independent of discourse semantics. Our account builds on Bachenko et al. (1986), but differs in its treatment of clausal structure and predicate-argument relations. Results so far indicate that the current system performs well when measured against a corpus of judgments of prosodic phrasing.

1 Introduction

In previous work (Bachenko et al. 1986), we described an experimental text-to-speech system that determined prosodic phrasing for the Olive–Liberman synthesizer (Olive and Liberman 1985). The system generated phrase boundaries using information derived from the syntactic structure of a sentence. While we saw significant improvements in the resulting synthesized speech, we also observed problems with the system. Often these stemmed from our assumptions that both clausal structure and predicate-argument relations were important in determining prosodic phrasing. This paper reconsiders those assumptions and describes an analysis of phrasing that we believe corrects many of the problems of the earlier version. Like the earlier version, it has been implemented in a text-to-speech system that uses a natural language parser and prosody rules to generate information about the location and relative strength of prosodic phrase boundaries.

Our current analysis rests on two ideas. First, it is possible to describe a level of prosodic phrasing that is independent of discourse semantics. Second, this discourse-neutral phrasing depends on a mix of syntactic and nonsyntactic factors; chiefly, syntactic constituency, left-to-right word order, and constituent length. There is no necessary fit between syntactic structure and phrasing, since prosodic phrasing may ignore major syntactic boundaries in order to satisfy the constraints on phrase length. Our approach thus follows that of Grosjean et al. (1979), namely, that phrasing reflects ". . . 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" (p. 75).

Section 2 will outline our analysis, focusing on the relationship between syntactic and prosodic structure. The analysis is developed within the framework of generative grammar, but we believe it is consistent with other approaches to syntactic description.[1] Our main point will be that the syntax plays a necessary but not sufficient role in determining phrasing, its effects being filtered by separate conditions on prosodic well-formedness (e.g. length). Section 3 describes the implementation of our analysis in an experimental text-to-speech system, and Section 4 summarizes our main conclusions. Unless otherwise noted, the corpus we used as a source of observations on phrasing in human speech consisted of a taped professional dramatization of the Sherlock Holmes story *The Speckled Band* and a documentary about Mount Everest that includes both professional "prepared" narration and the spontaneous speech of interviews. The Holmes story involved two male speakers and one female speaker; the Everest documentary involved a male narrator and several male interviewees. Both of us independently transcribed the tapes according to our perceptions of prosodic phrasing. Other examples come from transcriptions of speech that we recorded at Bell Laboratories. In the transcriptions, we both distinguished three types of prosodic event: a primary phrase boundary, a secondary phrase boundary, and the absence of a boundary. The most salient characteristic of the primary phrase boundary was a pause, while that of the secondary boundary involved a change in pitch. In comparing the two transcriptions, we discarded cases in which there was a discrepancy between the two markings, which left us with a corpus of approximately 500 sentences against which the prosodic phrasing rules were tested.

2 An Analysis of Phrasing

2.1 Background: The Factors That Contribute to Prosodic Phrasing

2.1.1 Syntax and Prosodic Phrasing

What is the exact contribution of syntax to sound? There are well-known local syntactic phenomena that affect both phonetic segment quality and the stress pattern of the phrase. A difference in syntactic category affects phonetic quality in examples 1–3 below:

1. a. They live in Canada. (*live* = verb)
   b. He ate live lobster. (*live* = adjective)
2. a. Can you estimate the damage? (*estimate* = verb)
   b. Give us an estimate. (*estimate* = noun)
3. a. As the water grew colder, their hands grew number. (*number* = adjective)
   b. Do you have his phone number? (*number* = noun)

The type of object that a verb takes correlates with the pronunciation of the verb in 4 and 5:

4. He resided in Holland. (*reside* pronounced [rizaid])
5. He resided the house with aluminum. (*reside* pronounced [risaid])

Syntactic category information also influences word prosody, as in 6 and 7, where knowledge of category membership is necessary to determine the correct stress pattern.

6. a. Both content and style are important. (*content* = noun)
   b. They are content to remain here. (*content* = adjective)
7. a. This ticket is invalid. (*invalid* = adjective)
   b. He is an invalid. (*invalid* = noun)

Syntactic category may also affect phrasal stress. For example, the sequence *power units* has stress on *units* in the verb-noun sequence in 8, but it has stress on *power* in the noun-noun sequence in 9:

8. If house current fails, power units from battery.
9. The power units failed.

Finally, syntactic gaps affect segment quality. For example, the vowel in the preposition *to* is normally weak, as in 10. But if a gap like the one that is associated with the question word *who* in 11 follows the preposition, the vowel of *to* is strong.

10 We spoke to John. (*to* pronounced/tə/)
11. Who did you speak to? (*to* pronounced /tu/)

When it comes to sentence-level prosody, especially phrasing, it is often true, as we will see below, that a sequence of words dominated by the same syntactic node cohere more closely than a sequence of words dominated by two different nodes. This observation has led some researchers, e.g., Cooper and Paccia-Cooper (1980), to claim a direct mapping between the syntactic phrase and the prosodic phrase. However, this claim is controversial because of the misalignments that occur between the two levels of phrasing. For example, in considering the connection between syntax and phrasing, the linguistic literature most often refers to examples of embedded sentences. Sentences like 12, from Chomsky (1965), are frequently cited. (Square brackets mark off the NP constituents that contain embedded sentences.)

12. This is [$_{NP}$ the cat that caught [$_{NP}$ the rat that stole [$_{NP}$ the cheese]]]

In such cases, the syntactic constituency indicated by bracketing is not in alignment with the prosodic phrasing. Instead, 12 has the prosodic phrasing in 13a. The phrasing in 13b, which most closely matches constituency, is strange at best. (In these and other examples, the most prominent prosodic boundaries are marked by vertical bars.)

13. a. This is the cat ‖ that caught the rat ‖ that stole the cheese,
    b. ??This is ‖ the cat that caught ‖ the rat that stole ‖ the cheese.

To account for such mismatches, "readjustment rules" that change constituent structure by adjoining each embedded sentence to the node dominating it have been posited. The result is a flattened structure that more accurately reflects the prosodic phrasing. In Chomsky and Halle (1968), this flattening process is not part of the grammar. Rather, it is viewed as "... a performance factor, related to the difficulty of producing right branching structures such as [12]" (p. 372). Thus phrasing, in their approach, is only indirectly related to syntax, since readjustment is done by special rules outside the grammar proper.

Langendoen (1975) proposes readjustment rules similar to those of Chomsky and Halle, but he claims that the readjustment of structure is part of the grammar, not part of the performance model. He thus makes explicit what is often a tacit assumption in both the linguistic and psycholinguistic literature[2]—that there is a direct connection between syntactic constituency and prosodic phrasing, with apparent misalignments readjusted before syntax interfaces with prosodic phonology.

Langendoen's proposal works well for sentences such as 12 because it predicts that important prosodic phrase boundaries will coincide with sentence boundaries. But this does not always fit the prosodic facts—sentences that lack overt complementizers or relative pronouns often resist the insertion of a break to set them off. For example, when applied to *They believe California sales are still off 75 percent,* readjustment rules cause the embedded sentence to be set off prosodically, as in 14a. This seems quite unnatural compared with 14b (an observed example), where a boundary has been inserted not before the sentence, but after the embedded subject.

14 a. ??They believe ‖ California sales are still off 75%.
b. They believe California sales ‖ are still off 75%.

Similarly, flattening the relative clause *it saw* in *It was ready to bite the first person it saw* has the questionable effect of inserting a prosodic phrase boundary before the relative clause, as in 15a. But in our data, this sentence actually has the phrasing in 15b, where the relative clause is not set off. (The relative clause in these examples is italicized.)

15. a. ??It was ready to bite the first person ‖ *it saw.*
b. It was ready to bite ‖ the first person *it saw.*

Moreover, there are certain distinctions among clause types, for example the difference between restrictive and appositive relatives, that are captured only by the presence or absence of a separate prosodic phrase for the clause. However, Langendoen's claim that embedded clauses are flattened would nullify this difference. The flattening and consequent setting off of restrictive relatives would render sentences such as 16a unintelligible because the associated appositive reading forces a contradiction between *who came from Plymouth* and *who came from Falmouth.* We believe the only intelligible version of this sentence is 16b, where there has been no readjustment of the *come from* clauses.

16. a. ??The pilgrims ‖ who came from Plymouth ‖ were a lusty bunch ‖ while the pilgrims ‖ who came from Falmouth ‖ were not.
b. The pilgrims who came from Plymouth ‖ were a lusty bunch ‖ while the pilgrims who came from Falmouth ‖ were not.

In sum, the contribution of syntax to sound is borne out by several phenomena. Even at the level of prosodic phrasing, syntactic constituents often cohere. Where misalignments between the syntactic and the prosodic phrasing occur, however, the notion of readjusting the syntax to fit the prosody is problematic and, we believe, compares unfavorably with an approach that views the semantic and phonological components as contributing to prosodic phrase boundary determination.

2.1.2 SEMANTICS AND PROSODIC PHRASING

The syntax/prosody misalignment may be viewed as resulting in part from semantic considerations. Both predicate-argument relations and discourse factors have been examined for their possible input to prosodic phrasing.

Crystal (1969) claims that prosodic phrase boundaries will co-occur with grammatical functions such as subject, predicate, modifier, and adjunct. Selkirk (1984) and Nespor and Vogel (1986) take a similar approach, but within a different theoretical framework. Previous versions of our work, as described in Bachenko et al. (1986) also assume that phrasing is dependent on predicate-argument structure. The problem here is that the phrasing in observed data often ignores the argument status of constituents. In 17a–f, for example, the phrasing makes no distinction between arguments and adjuncts. All of the sentences have the same $X(VY)$ pattern even though Y is a complement in the first case (*the first serious attempt*) and an adjunct in the others. (The complement in 17a and the adjuncts in 17b–f are italicized.)

17. a. A British expedition ‖ launched *the first serious attempt.*
b. A single bright light ‖ shone out *from the darkness.*
c. There were several little changes ‖ carried out *about that time.*
d. Were there any gypsies ‖ camping *in the plantation. . . .*
e. . . .like the claws of a crab ‖ thrown out *on each side.*
f. Two years ‖ have passed *since then.*

The relation between discourse and prosodic phrasing has been examined in some detail by Bing (1985), who argues that each noun phrase in an utterance constitutes a separate prosodic phrase unless it is destressed because of reference to previous discourse. Bing also observes that constituents that refer to items newly introduced into a discourse tend to be longer. This may be the reason that word count and syllable count play a prominent role in prosodic phrasing (see Section 2.1.3.). To our knowledge, no work has explicitly explored the relation between the length of a constituent and its status in the discourse.

Hirschberg and Litman (1987) and Litman and Hirschberg (1990) also examine the relation between discourse and prosodic phrasing. Their work succeeds in distinguishing the use of items like *now, so,* and *well* as discourse cues from their denotative lexical use on the basis of a complex combination of pitch accent type and phrasing.

The Hirschberg and Litman studies identify a specific discourse distinction that relates to phrasing. These studies are not intended to give a picture of the extent to which discourse relates to phrasing. On the other hand, Bing's work gives a broader picture of the relation between discourse and phrasing, but it deals only with noun phrases. Thus both of these efforts leave open the question as to whether discourse features completely determine prosodic phrasing or are a complement to some more basic set of determinants, syntactic and/or phonological. In other words, when prosodic features that reflect facts of the discourse are removed, is there a residual, neutral phrasing?

Our work on the prosodic phrase status of clause final prepositional phrases, which we discuss below, suggests the existence of a discourse-neutral phrasing that depends on syntactic constituency mediated by string adjacency and length of a potential prosodic phrase.[3] Such phrasing provides us with a typical phrasing pattern analogous to the typical phrasal stress patterns examined in Liberman and Prince (1977), which "are often overwhelmed by the chiaroscuro of highlight and background in discourse, but retain the status of null-hypothesis patterns that emerge when

there is no good reason to take some other option" (p. 251). This approach to prosodic phrase boundary determination brings us closer to a framework in which phonological, syntactic, and discourse features all contribute to prosodic phrasing.

The possibility of a discourse-neutral prosodic phrasing is also of import to the prosodic quality of synthetic speech, since it allows us to "get by" without a complete description of the discourse features of a given text, many of which have yet to be characterized. Interestingly, in the data we examined we found only 14 percent of the phrases to be discourse-determined.

The identification of a preferred phrasing that is independent of discourse also aids us in identifying and characterizing the discourse features that impinge on prosodic phrasing. Several well-known discourse phenomena—coreference, contrast, and parallelism—affected the phrasing of the clause final prepositional phrases in our corpus. We are left with three or four unexplained cases that are suggestive of a discourse explanation.

2.1.3 PHONOLOGICAL LENGTH AND PROSODIC PHRASING

The psycholinguistic studies of Martin (1970), Allen (1975), Hillinger et al. (1976), Grosjean et al. (1979), Dommergues and Grosjean (1983), and Gee and Grosjean (1983), responding to the idea of readjusted syntax as the source of prosodic phrasing, show that grammatical structure, even if readjusted, is not in itself a reliable predictor of prosodic phrasing: mismatches between syntax and prosody occur often and systematically, and can be related to specific nonsyntactic factors such as length and word frequency. For example, although prosodic boundaries between subject and verb do occur, there also exist prosodic patterns in which the boundary comes between the verb and object, i.e., the data reveal both $X(VY)$ and $(XV)Y$ groupings. Grosjean et al. (1979) claims that such mismatches are due for the most part to constituent length, which interacts with grammatical structure and, in some cases, overrides it. Thus syntactic and prosodic structure match when the major constituents of a sentence are roughly equal in length; for example, the main prosodic phrase break corresponds to the subject-predicate boundary in *Waiters who remember well* ‖ *serve orders correctly.* Discrepancies in length throw constituents off balance, and so prosodic phrasing will cross constituent boundaries in order to give the phrases similar lengths; this is the case in *Chickens were eating* ‖ *the remaining green vegetables,* where the subject-predicate boundary finds no prosodic correspondent.[4]

The most explicit version of this approach is the analysis presented in Gee and Grosjean (1983) (henceforth G&G). Drawing on the psycholinguistic studies mentioned above and on aspects of the grammar of prosody outlined in Selkirk (1984), G&G propose an algorithm for mapping syntactic structure onto a hierarchical representation of phrasing; the rules they present accomplish this by integrating syntactic information (e.g. constituent structure, left-to-right ordering) with information about constituent length. We have found that their rules, which are described in detail, provide a productive model for investigations of phrasing, and in what follows we shall frequently refer to their analysis. But, as we will show, G&G fall short of providing a comprehensive theory. Their rules are too limited and their syntax too underspecified to achieve moderate coverage for an unrestricted collection of sentences or to provide an adequate description for implementation.[5]

2.2 CURRENT ANALYSIS

Our goal has been to develop a theory of syntax/prosody relations that we could test in an experimental text-to-speech system. We approached the problem with the assumption that there is a level of prosodic phrase determination that does not include discourse factors, and that aiming for this level would yield an appropriate phrasing for a sentence. Both the output of the system and our preliminary findings, which show that discourse factors influence just a small part of the phrasings that follow a verb, indicate that this approach is feasible.

The analysis that we arrived at takes G&G and, to some extent, Selkirk (1984) as its starting point. Hence we are assuming that there is no necessary match between syntactic structure and prosodic phrasing. Prosody rules refer to syntactic structure, but they are not obliged to preserve it; independent principles of prosodic well-formedness, in particular length calculations, may create entirely different structures that appear at odds with the syntax. Here we shall describe the main features of our analysis and then go on to a description of the implementation.

Our prosody rules are intended to account for two aspects of phrasing: boundary location and boundary salience. In 18, for example, the rules need to stipulate that a phrase boundary comes between the subject and predicate.

18. The light among the trees ‖ was extinguished.

But when there is more than one important phrase boundary in a sentence, the rules will also specify a relative salience, or perceptibility, for each boundary. Thus in the observed sentence 19, where an adjunct has been prefixed to the sentence, the boundary between subject and predicate diminishes in deference to the stronger boundary between adjunct and core sentence. A single vertical bar marks the diminished boundary.

19. About nine o'clock ‖ the light among the trees | was extinguished.

After deciding which boundaries will be diminished and which highlighted, the rules assign each boundary an acoustic value that reflects its relative strength. Our current

system uses three values made available by the synthesizer. A pause and its concomitant prosodic effects mark the strongest boundaries, a pitch change signifies intermediate boundaries, and the weakest boundaries are assigned a phrase accent.

Our work so far has focused solely on the issues of location and relative salience; the rules for associating the different boundaries with specific intonation contours and acoustic values are still quite rudimentary. Consequently our discussion will center on the location and salience rules, and we will mention the third rule class only in passing.

2.2.1 BOUNDARY LOCATION RULES

The location rules identify possible boundary sites by first deriving phonological words from the lexical items in a parse tree, and then grouping the phonological words into larger phonological phrases. The boundaries that separate phonological phrases are the candidates for prosodic phrase boundaries, since the prosodic phrases of speech consist of one or more phonological phrases.

Our rules for phonological word formation are adopted, for the most part, from G&G, Grosjean and Gee (1987), and the account of monosyllabic destressing in Selkirk (1984). Thus in our analysis, rules of **phonological word formation** apply to the non-null terminal nodes in a syntax tree. If the terminal is a content word, i.e. noun, verb, adjective, or adverb, then this terminal may have the status of a phonological word on its own. Otherwise the word combines with one or more orthographically distinct words to form a single phonological word that has no internal word or phrase boundaries. This is accomplished by adjoining a word to its left or right neighbor depending on its lexical category and its position in the tree. Function words, e.g. auxiliary verbs, articles, prepositions, pronouns, and conjunctions, are all eligible for adjunction in certain syntactic contexts. Content words, copular verbs, demonstratives, quantifiers and elements in the complementizer node can serve as hosts for the adjoined material or stand alone.

Figure 1 illustrates the effects of phonological word formation; the "+" indicates that adjunction has taken place.

Article adjunction, for example, attaches *a, an,* and *the* to a following word, so that *a sudden* and *the trees* in Figure 1 each becomes a single phonological word in which the article acts as an unstressed syllable. The rule of preposition adjunction, which has applied twice in Figure 1, attaches a preposition to the material on its right only if it is the head of a PP, otherwise the preposition attaches leftward. Thus in 20a, where the preposition is a syntactic head, *in* + *the* + *dimly* forms a single phonological word after article and preposition adjunction. The phrase boundary in this case will precede the preposition. But in 20b, where there is no PP (rather, the preposition is a sister of the verb in the syntax tree), *filled* + *in* is a phonological word. Hence the boundary will follow the preposition.

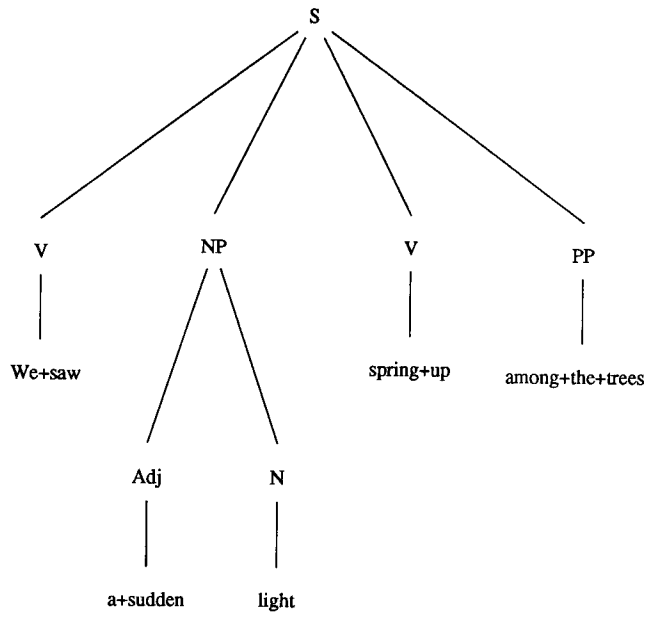


Figure 1 Phonological Word Formation.

20. a. Holmes waited | *in* + *the* + *dimly* lit room.
    b. . . .and filled + *in* | a few of the gaps.[6]

Rules of **phonological phrase formation** now build the next level of prosodic constituents.[7] Each phonological phrase consists of a syntactic head and the material that intervenes between it and a preceding head (usually, the pre-head modifiers, e.g. pre-nominal adjectives, pre-verbal adverbs). Following Selkirk (1984), we have limited the eligible head categories to noun, verb, adjective, and adverb (although adjectives and adverbs do not count if they directly precede and modify another head). Examples 21a–b illustrate the results of phonological phrase formation. In each case, the phonological phrase is created by a left-to-right process that collects material up to and including the head of a syntactic constituent. Every phonological phrase boundary thus marks a syntactic head as well as the site of a possible prosodic boundary in speech. (The sequences with + are words formed by adjunction; | stands for a phonological phrase boundary.)

21. a. A + British expedition | launched | the + first serious attempt.
    b. We + saw | a + sudden light | spring + up | among + the + trees.

Which boundaries become the prominent ones is determined by the salience rules described below.

The elements of phonological phrases cohere strongly in speech—they cannot be separated into smaller phrases without a dramatic effect on the semantic content of the sentence. In 22a, for instance, the italicized phrase must be treated as a minimal element with respect to phrasing; the

phrasing of 22b, where an important break comes before the syntactic head, is rare in our data.

22. a. I shall never forget | *that April morning.*
    b. ??I shall never forget *that* | *April morning.*

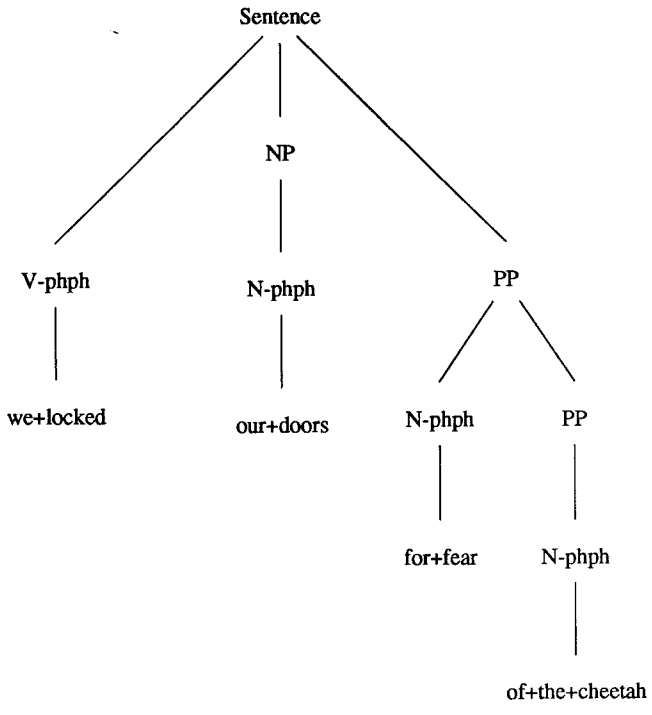
Example 23a, with a phonological phrase boundary before the preposition *of,* follows the pattern that is found in nearly all of our data. The observed pattern in 23b, where the boundary follows the *of,* creates a noticeably theatrical effect with emphatic stress on *strange.*

23. a. He's a collector | *of strange animals.*
    b. He's a collector of | *strange animals.*

2.2.2 BOUNDARY SALIENCE RULES

The rules for salience apply to a combination of phonological phrases and syntactic constituents. Their input is a structure like that in Figure 2, where syntactic constituents may contain one or more phonological phrases. The NP in Figure 2 consists of a single phonological phrase, and the top-level PP contains two. The absence of VP in this figure will be explained below.

When they apply, the salience rules merge phonological phrases to create larger prosodic phrases, which are also merged into a final phrase hierarchy. Boundaries between the phrases are thus diminished or emphasized, finally giving the impression of a balanced, rhythmic pattern whose components appear to be equal in length. The salience rules apply on the basis of (i) adjacency to a verb, (ii) length, and (iii) constituent type.



We locked our doors for fear of the cheetah

Figure 2 Input to Salience Rules
(phph = phonological phrase).

All three factors interact in the initial balancing of material around a verb. In this process, the verb groups to the left to form a $(XV)Y$ pattern or it groups to the right to produce a $X(VY)$ pattern. Our analysis generally follows G&G, who propose the following rule, where, in their formulation, X is a prosodic constituent or null, V is a verb, Y is a nonsentential complement, and C refers to phonological word count.

24. **Verb Balancing Rule**
    in $[X\ V\ Y]$
    if $C(X) + C(V) < C(Y)$
    then $[(XV)Y]$
    otherwise $[X(VY)]$

The rule works through a sentence from left to right. It says that if combining the verb with the constituent to its left yields a word count less than that of the complement, the verb forms a prosodic phrase with its left neighbor. Consequently, the phrase boundary following the verb is strengthened. For all other cases, the verb groups to the right so that the boundary preceding the verb becomes reinforced (G&G 442). In 24, X and Y contain prosodic constituents, either phonological phrases or prosodic phrases formed by other salience rules (e.g. the constituent rules that build NP and PP into separate prosodic phrases; see G&G, p. 441). Word count (C) is determined solely by the number of phonological words.[8]

Example 25a gives the phonological phrasing (indicated by |) for *This little incident gives a new zest to our investigation.* Applied to this string, the verb rule will group the verb to the right and derive the $X(VY)$ pattern since *This little incident* plus *gives* adds up to four words, while *a new zest* is only two words; the *to*-phrase is not adjacent to the verb and so is not considered by the rule. The final phrasing is given in 25b, where the verb rule accounts for the break after the subject, and length rules that are discussed below account for the second boundary.

25. a. This little incident | gives | a + new zest | to + our + investigation.
    b. This little incident ‖ gives a + new zest | to + our + investigation.

Similarly in 26, *Holmes' eyes traveled* adds up to three, but the post-verb conjoined phrase *round and round* only counts for two. Hence the verb groups to the right.

26. Holmes' eyes ‖ traveled round and + round.

The sentences in 27 follow the $(XV)Y$ option. In 27, *as + the + lamp* plus *was + lit* add up to two words versus the three words *in + one of + the + sitting rooms.* In 27b, *Chickens* plus *were + eating* also add up to two words, *the + remaining green vegetables* adds up to three. In 27c, *and + his + eyes* plus *were + fixed* adds up to two words, while *in + a + dreadful rigid stare* adds up to three.

27. a. as + the + lamp was + lit ‖ in + one of + the + sitting rooms. . .
    b. Chickens were + eating ‖ the + remaining green vegetables.
    c. and + his + eyes were + fixed ‖ in + a + dreadful rigid stare.

In their formulation of the verb rule, G&G impose two conditions on its application. First, the rule may examine only arguments of the verb, i.e., complements are candidates for *Y* in 24, but modifiers and adjuncts are not. Hence the rule must have access to VP constituency since verb complements, in the generative grammar framework assumed by G&G, are represented as sisters of V in VP, while modifiers and adjuncts are outside of the VP. Second, the rule cannot cross S boundaries—embedded clauses form separate prosodic units in G&G's analysis.

Our studies indicate that these conditions are too strong: balancing around a verb often crosses both VP and S boundaries in our taped data. For example, in 28a–b, where the verb and its complement occur in a single prosodic phrase, the phrasing may appear to be influenced by the presence of a VP (the complement is italicized.).

28. a. A + British expedition ‖ launched *the + first serious attempt.*
    b. The + 48 channel module ‖ can + have *only two di-groups.*

Yet the verb also forms a single prosodic phrase with sentence adjuncts. This is the case in 29a–c, where the verb and adjunct are separated by an important boundary in the syntax (VP), but not in the prosody. In these sentences, while a secondary break may set off the adjunct, the main prosodic phrase boundary comes before a verb + adjunct sequence (the adjunct is italicized).

29. a. Seven of + our + porters ‖ were + killed *in + the + fall*
    b. a + crack ‖ opened *in + the + snow.*
    c. the + elements of + personal interest ‖ must + be + introduced *at + all costs*

If the verb-balancing rule is restricted to subcategorized complements of the verb, as G&G assume, then the phrasing in 29a–c has no explanation since, with the restriction, the main boundary in these sentences has to come between the verb and adjunct, a prediction that contradicts the observed pattern and sounds strange at best. Sentences such as those in 29a–c suggest that *Y* in the verb rule of 24 should not be limited to material within VP but should include anything to the right of V. Hence we are assuming that the key to phrasing in 28a–b, 29a–c is the adjacency relationship between a verb and the constituent on its right, not verb phrase structure, or, equivalently, the complement versus noncomplement status of a constituent. In particular, phrasing around the verb depends on the relative length of constitutents that are adjacent to a verb and, as we will observe, the presence of specific verb-adjacent items (e.g. phrasal *and*).

The phrasing of embedded sentences follows a similar course. Prominent phrase boundaries often co-occur with clause boundaries, e.g. *Customers are asking retailers ‖ whether their watermelons come from California,* but we believe that this is only apparently due to clausal constituency. In our data, components of S can form a prosodic phrase with nearly any adjacent material, regardless of where the S boundary falls. For example, 30 a–c has the most prominent break before the matrix verb, not before the embedded S (which is italicized). The sentences of 31 have the most prominent prosodic break within the embedded sentence.

30. a. Even my + fiance ‖ believes *it's only my + imagination.*
    b. Did + Dr. Roylott ‖ continue *to + practice?*
    c. A + terrible change ‖ began *to + come + over our + stepfather.*
31. a. They + believe *California sales ‖ are + still off 75 percent.*
    b. I've + heard *that + the + crocuses ‖ promise very well.*
    c. Experience has + proved *that + savages ‖ are the + tyrants of + the + female sex.*
    d. I + seem *to + see dimly ‖ what you're driving at.*

Here again it appears that G&G's formulation of verb balancing, which prohibits the rule from crossing sentence boundaries, is too strict. Requiring the prosody rules to preserve the constituent status of embedded clauses predicts that when the verb is followed by an embedded clause, as in 30a–c and 31a–d, balancing is superceded by rules that work on the internal structure of the embedded S; the result in most cases is a prominent prosodic boundary before the embedded clause. According to this approach, the sentence *Did Dr. Roylott continue to practice?* would have the odd phrasing *Did Dr. Roylott continue ‖ to practice?* instead of the observed phrasing of 30b.

The phrasing patterns of 30a–c and 31a–d are easy to explain if we assume that prosody rules ignore clausal constituency so that verb balancing applies across S boundaries. Evidence from sentence adjuncts, e.g. the purposive in *We went out to the Himalaya to climb. . .* and the relative clause in *It was ready to bite ‖ the first person it saw* leads us to believe that prosody rules ignore these clausal constituents as well. Our analysis thus adopts the basic mechanism G&G propose with their verb rule, but without the conditions on verb complements and clausal constituency. While NP, PP, and AdjP constituents "count"in our prosody rules, VP and S constituents do not. Consequently, eligible material to the right of V in the verb rule 24 may be a complement, a modifier, an adjunct, or the initial constituent of an embedded sentence. The exact treatment of the left-adjacent constituent is still a topic for investigation, and there is currently no requirement that the material to the left of a verb be a subject.

G&G intend the balancing rule to make the verb a prosodic center by grouping constituents in such a way as to

create, in most cases, two phrases of approximately equal length, with the verb as a left or right edge. During subsequent processing, this balancing effect is usually lost since neither length nor adjacency to a verb play any further role in G&G's analysis. After verb balancing, the remaining constituents are bundled from left to right into a left-branching binary tree like that in Figure 3. As we will discuss below, the higher the constituent is in the prosody tree, the more prominent will be the boundaries that set it off in speech. Hence, when constituents are bundled, as in Figure 3, material that comes at the end of a sentence will usually be set off by the largest breaks. This is what happens in 32a–b, where a strong boundary before the final (italicized) constituent seems desirable.

32. a. . . .walk away from your helper ‖ *approximately 50 feet.*
    b. I suddenly heard in the silence of the night ‖ *that same low whistle.*

But the application of bundling immediately after the verb rule often leads to problems. The salience of a boundary that occurs toward the end of a sentence tends to be overestimated when the final constituent is relatively short. When this happens, the final constituent may be set off unnaturally from the rest of the sentence. In 33a–c, for example, a strong boundary before the italicized constituent is inappropriate.[9]

33. a. ??The speaker pronounced the names of the characters ‖ *on the left.*
    b. ??We locked our door for fear ‖ *of the cheetah.*
    c. ??She had caught an early morning train ‖ *from London.*

We believe that the problems raised by final bundling can be avoided largely by extending the effects of length and verb adjacency beyond the verb balancing rule. In our analysis, this is accomplished by an adjacency rule and two length rules that, in effect, sustain verb centering and determine the prosodic weight of constituents not adjacent to a verb. The adjacency rule in 34 applies after the balancing rule of 24 and groups the "unclaimed" verb-adjacent constituent with the phrase that was formed by 24 (this will be either (VY) or (XV)). X in 34a and Y in 34b are prosodic constituents.

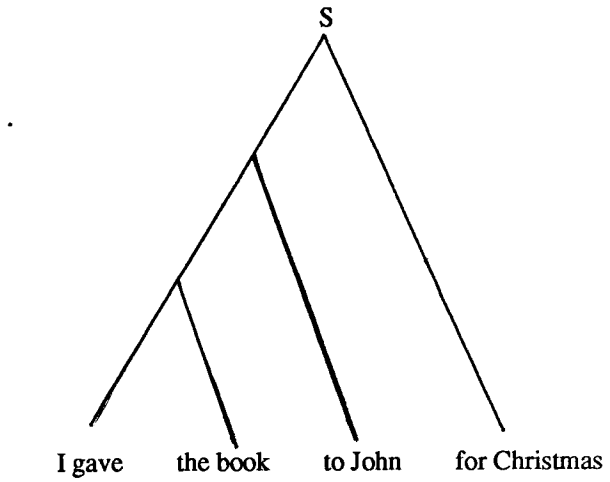


Figure 3 Final Bundling (adapted from G&G, p. 443).

34. **Verb Adjacency Rule**
    a. $[\ldots X(VY)\ldots] \rightarrow [\ldots(X(VY))\ldots]$
    b. $[\ldots(XV)Y\ldots] \rightarrow [\ldots((XV)Y)\ldots]$
    where . . . is a phonological phrase or null and X may or may not be a subject.

Case 34a says that if (VY) is a prosodic constituent, then create a new prosodic constituent composed of (VY) and adjacent material to the left. Likewise 34b creates a new prosodic constituent by combining (XV) with adjacent material on the right. What results is a prosodic verb phase—a cluster of two prosodic phrases with the verb in the middle as a left or right edge. In 35a–b, for example, constituents abutting the verb have been worked into a single prosodic verb phrase by (i) the verb balancing rule, which groups the verb rightward with the complement (giving . . . $X(VY)$. . .), and (ii) the adjacency rule in 34b, which generates a larger phrase containing the constituent on the left (= . . .$(X(XY))$. . .). The prosodic verb phrase is italicized; its internal boundary is marked by a single vertical bar.

35. a. *Everest | is an + enormous pyramid* ‖ with + three wide faces and + three ridges.
    b. *Mrs. Welles | wrote a + weekly sports column* ‖ for + the + Christian Science Monitor.

Constituents that are not adjacent to the verb form the periphery of a prosodic verb phrase. In 35a, the periphery consists of *with three wide faces* and *and three ridges;* in 35b it consists of *for the Christian Science Monitor.* While our understanding of phrasing of the periphery of a sentence is far from satisfactory, we have been conducting a study that so far suggests that, at least with respect to clause-final prepositional phrases, length of a peripheral constituent is an important determinant of its prosodic prominence. The phrases considered all occur to the right of the verb with some material intervening between the verb and a prepositional phrase, PP. The intervening material varies from a full phrase to a single word that functions as a syntactic head to which the PP under consideration is a complement. In a test of 129 clause-final prepositional phrases, we noted the pattern shown in Figure 4. The clear bifurcation of the numbers in this test suggests that length of the PP is determining the degree to which final PPs are set off. Many of the 18 apparent counterexamples to this claim also indicate that length establishes a discourse-neutral phrasing for sentence-final PPs that is contravened only by predictable syntactic and discourse factors. For example, the *of*-PP in the partitive construction of 36a is not in a separate prosodic phrase, even though it contains

	PP=1 stress foot	PP>1 stress foot
Pause precedes	6	63
No perceptible prosodic event precedes	48	12

Figure 4 Phrasing of final PPs.[10]

two stress feet (*west* and *wing*). This contrasts with 36b, where the *of*-PP is not a partitive.

36. a. . . .blue smoke | curling + up from + the + chimneys || showed that part of + the + west wing. . .
b. A + single bright light || shone + out from + the + darkness || of + the + west wing.

Similarly, while the italicized PP in 37a follows the rule of being greater than one phonological word and therefore set off, a similar PP in 37b is not set off because the repetition of *room* causes this item to be destressed and consequently to merge with the relevant PP.

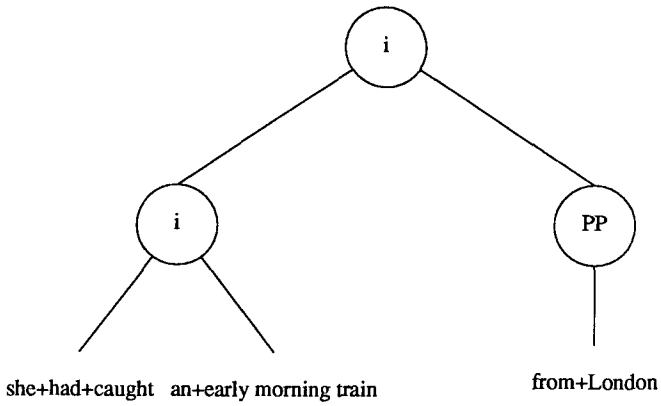
37. a. Holmes passed at + once || into + the + room || *in + which Helen Stoner* || was now sleeping.
b. so + that I've + had to + move out + of + my + own room || into + the + room next door || the + room *in + which my + sister died.*

In the implementation, the length rules extend these results to include post-verb peripheral constituents other than PPs; specifically, interjections and adverbials. Currently, the length rules perform two operations on post-verb peripheral material, depending on word count. In one case, if a phrase, *P1,* consists of a single phonological word, it is adjoined to the most recently created phrase, *P0* (usually a prosodic verb phrase). The result is a new phrase, *P2,* whose boundary salience is equal to that of *P0.* This is illustrated in Figure 5, where *i* is a number representing salience (Section 3 shows how salience indices are derived). In the second case, a longer phrase, *P1,* will be bundled with its preceding material, *P0,* in order to form a new phrase, *P2,* whose salience is the sum of salience value for *P0* and the value of *P1* (=word count + 1). In Figure 6, for example, the 'long' PP is set off by a relatively large index.

Peripheral material at the beginning of a sentence is currently picked up by left-to-right bundling without regard to constituent type or length. The correct treatment of sentence-initial peripheral material remains a topic of investigation.

2.2.3 A PROSODIC LEFT CORNER CONSTRAINT

In some cases, phrasing is influenced by the lexical content of a constituent. Phrasal *and,* i.e. the left corner of a NP, PP, AdjP, or AdvP conjunct, always starts a new phrase, as in *Next to it || he placed a box of matches || and a candle.*[11]



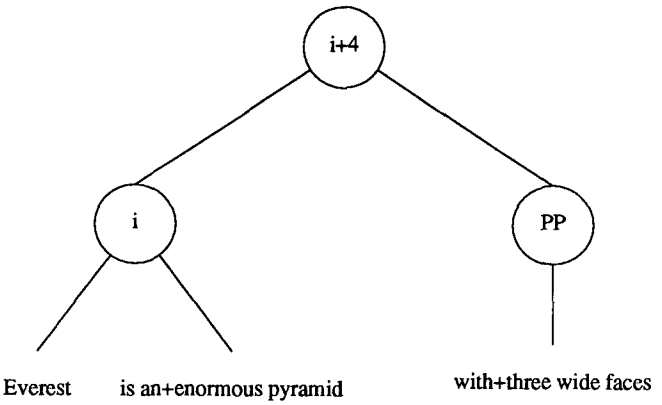
She had caught an early morning train from London

Figure 5 Adjunction of Short Peripheral Phrase (*i* = boundary salience).

In our analysis, this fact is captured by a global constraint on the prosody rules: no boundary location or salience rule may apply to a constituent whose left corner is a phrasal conjunct. Hence when phrasal *and* is adjacent to a verb, the prosodic left corner constraint will block the rules that form prosodic verb phrases. In 38 the verb rule is prevented from merging *was extinguished* with *and all,* as it would do if *and* were treated as any other function word.

38. The light among the trees was extinguished and all was dark.

Although the constraint should probably include other material, such as the subordinate conjunctions (e.g. *because, while, although*), our current analysis acknowledges only phrasal *and, or,* and *nor,* requiring that they always introduce a separate phrase. In our implementation of the



Everest is an enormous pyramid with three wide faces

Figure 6 Long Prepositional Phrase Attachment (*i* = boundary salience).

prosody rules, we have extended the constraint to punctuation. For example, the comma in *After Robert ate, his cat Freddy took a nap* is first adjoined to the constituent on its right, to give *After Robert ate , +his cat Freddy took a nap,* and then, like phrasal *and,* obligatorily starts a new phrase, preventing the prosody rules from merging *ate* and *his cat.*

2.2.4 SUMMARY

In this section, we have described an analysis of prosodic phrasing that incorporates two classes of rules. Boundary location rules form phonological words from the terminal elements of syntactic structure and build phonological phrases (the first tier of prosodic constituency), using information about syntactic heads. Boundary salience rules assign a relative strength, or perceptibility, to each phrase boundary according to syntactic labeling, length, and adjacency; they ignore verb phrase and clausal constituency and predicate-argument relations. The primary salience rules are, in order, syntactic constituency, which converts NP, PP, and AdjP constituents into prosodic phrases (see G&G, p. 441); verb balancing and verb adjacency, which derive prosodic verb phrases; length rules, which apply to the material on the right of a prosodic verb phrase; and final bundling, which collects phrases built up by the previous rules into a binary tree (see Figure 3). Location and salience rules are both subject to a left corner constraint on their application. In the following section, we discuss how the rules work in an experimental text-to-speech system.

3 AN EXPERIMENTAL TEXT-TO-SPEECH SYSTEM

We have built an experimental text-to-speech system that uses our analysis of prosody to generate phrase boundaries for the Olive–Liberman synthesizer (Olive and Liberman 1985). Two concerns motivated our implementation. First, we hoped the system would provide us with a research tool for testing our ideas about syntax and phrasing against a large unrestricted collection of sentences. Second, we wished to investigate how well our approach would work for determining prosodic phrasing in a text-to-speech synthesizer. Existing text-to-speech systems perform well on word pronunciation and short sentences,[12] but when it comes to long sentences and paragraphs, synthetic speech tends to be difficult to listen to and understand. Many investigators (e.g. Allen 1976; Elowitz et al. 1976; Luce et al. 1983; Cahn 1988) have suggested that the poor prosody of synthetic speech, in comparison with natural speech, is the primary factor leading to difficulties in the comprehension of fluent synthetic speech. And while researchers in text-to-speech synthesis have adopted a variety of approaches to prosodic phrase generation—from the simple punctuation-based rules and function word listings of existing commercial systems to the sophisticated prosodic heuristics described in Emorine and Martin (1988) and O'Shaughnessy (1989)—the generation of appropriate prosodic phrasing in unrestricted text has remained a problem.

As we will show, our results so far indicate that our experimental system, which assigns a discourse neutral prosodic phrasing on the level of sentences, provides a significant improvement in the quality of synthesized speech. We believe that one reason for the improvement has to do with the increased pitch range that our system uses. Text-to-speech systems that lack sentence-level phrasing must take a conservative approach to pitch settings in order to avoid misleading and inappropriate pitch modulations. Correct phrase identification makes it possible to adopt an expanded pitch range that greatly enhances the naturalness of the final speech. In constructing the system, we focused on two core questions: (i) what kind of parser is needed for the prosody rules? and (ii) how should prosodic phrasing, i.e. boundary location and strength, be represented?

3.1 PARSING FOR PROSODY

The rules for discourse-neutral phrasing that we propose need examine only a subset of the syntactic information that most parsers provide. That is, the rules require access to lexical category, syntactic heads, NP/PP/AdjP/AdvP constituency, and left-to-right word order, but not to clausal constituency, predicate-argument relations, or modifier attachment. We believe the rules must also recognize the trace of *wh*-movement (e.g. the trace that precedes the phrase break in *The slope on which we were standing [trace] started to move*), although other null terminals such as the trace of passivization are ignored.[13] At the outset of our project, we had available to us a moderate coverage deterministic parser—Fidditch[14]—that we adapted to the syntactic requirements for prosodic phrasing. This modified "speech parser" produces parse trees like that in Figure 7. The tree represents syntactic information necessary for phrasing, but omits nodes (S, VP) and

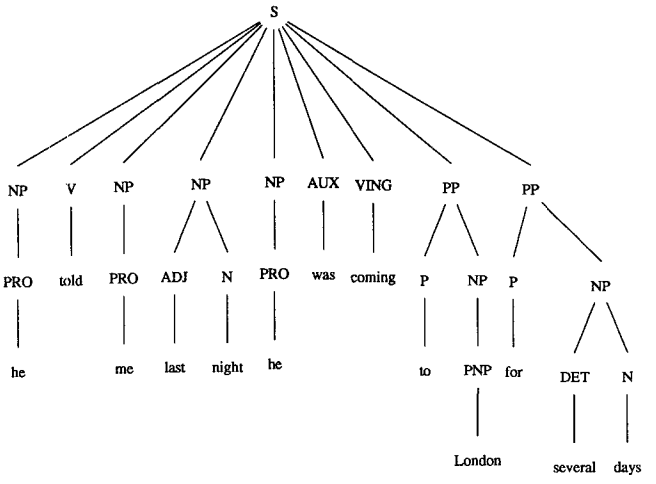


Figure 7 Syntactic Input to Phrasing Rules.

branches that are unnecessary for prosodic phrase determination.

3.2 PROSODIC PHRASE REPRESENTATION

Following G&G, we require that the prosody rules build a binary tree whose terminals are phonological words and whose node labels are indices that mark boundary salience. An alternative representation based on Liberman and Prince (1977) is presented in Selkirk (1984), which contends that prosody, including prosodic phrasing, is more properly represented as a grid instead of a tree. Although a grid may be more descriptively suitable for some aspects of prosody (for example, Sproat and Liberman (1987) use the grid representation for their implementation of stress assignment in compound nominals), we are not aware of any evidence for or against a grid representation of discourse-neutral phrasing.

Figure 8 shows the phonological phrase tree that is built from the syntactic structure of Figure 7. The rules for building this tree apply from left to right, following the analysis we described in the preceding section. Figures 9–11 show the prosodic phrase derivation. Numbered nodes refer to salience values, with higher numbers indicating greater salience. The index is assigned according to phonological word count, with one point added for the node itself. Figure 11 is the final prosodic phrase tree; in the notation we have been using, the phrasing represented by Figure 11 is *He told me | last night || he was coming to London || for several days.*

Figure 9 shows the effect of two applications of verb balancing. Applying from left to right, the rule first looks at the phonological verb *he + told + me.* Since the material to the left of the verb is null, the rule must group this verb with the constituent on its right to form the node labeled ④. On its second application, the rule balances the prosodic phrase it has just formed against the single phonological word *to + London,* and groups the verb rightward to form node ③.

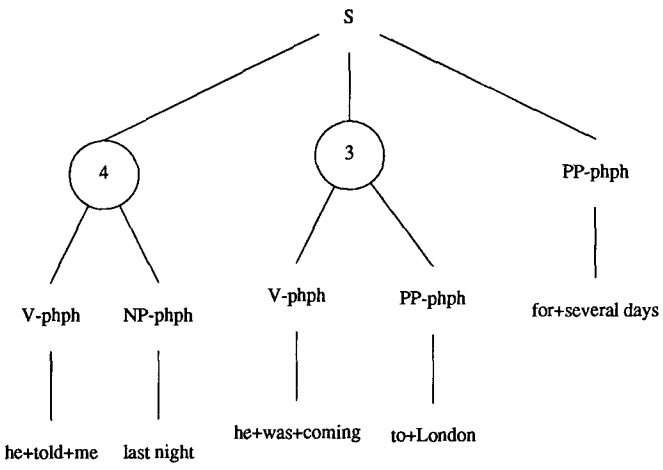


Figure 9 Two Applications of Verb Balancing.

Verb adjacency is now triggered by the contiguity of ④ and the verb *he + was + coming.* In Figure 10, this rule has formed the prosodic verb phrase ⑧.

It remains to collect the "long" peripheral constituent *for + several days.* In Figure 11, the length rule has built the final node of the tree; because the peripheral item consists of two phonological words, the value of the topnode is affected by the word count of the peripheral item. If the peripheral item had consisted of a single phonological word, the value of the top node would have been 8 instead of 11.

Finally, each node index is converted into one of three acoustic values. High indices are marked as a minor phrase boundary; mid-range indices are signified with a downstep on the first phonological word following the boundary

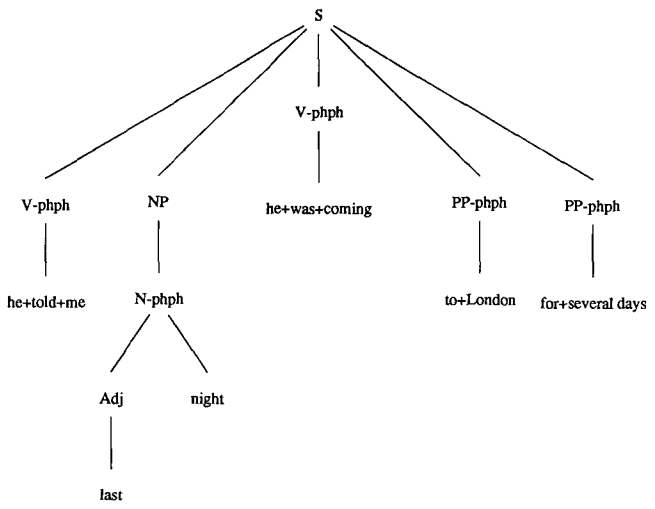


Figure 8 Phonological Phrasing.

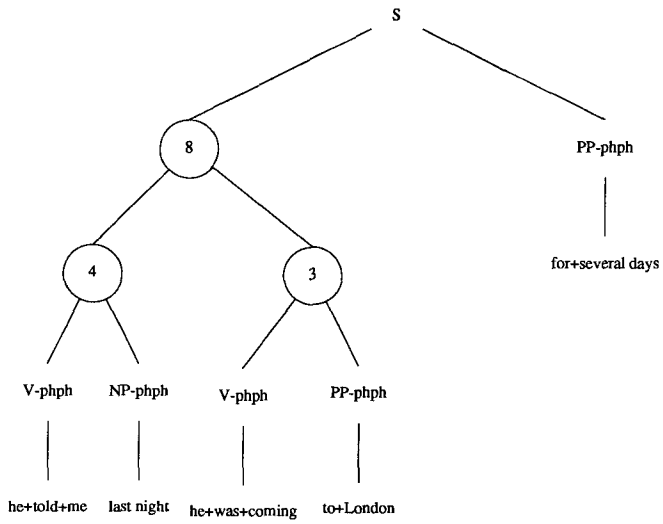


Figure 10 Verb Adjacency.

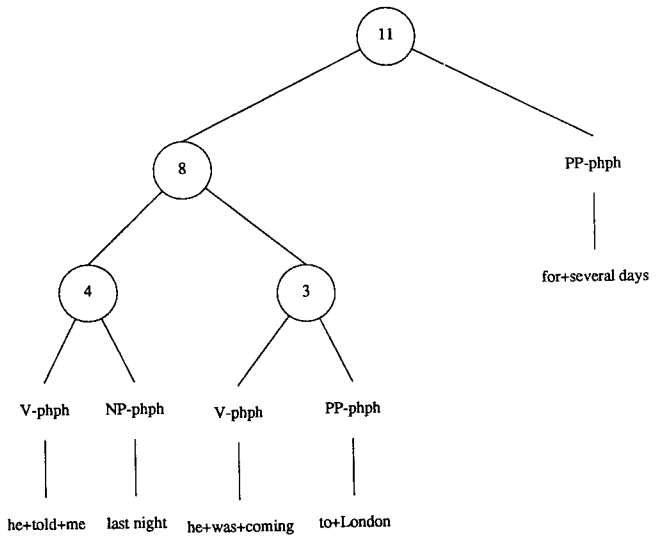


Figure 11 Length Rule—Long Phrase (final rule).

(Pierrehumbert 1980), and those in the lowest range receive a phrase accent. The mapping onto three different values reflects a simple subjective choice. A more complete analysis would consider finer acoustic "tunings" for the indices.

3.3 EVALUATING THE PHRASING SYSTEM

In testing a prosodic phrase system, it is necessary to be clear about the goals of the work. At the level of basic inquiry, such a system should aim to correlate well with human speech production, assuming that we are able to abstract away from speaker variability. However, in working under the assumption that discourse plays a role, however limited, in prosodic phrase determination, the comparison of synthetic to human phrasing becomes less meaningful since factors unknown to our system influence the placement of phrase boundaries.

At the level of system building, a prosodic phrasing system should aim to make a text-to-speech system easier to understand. In general, adding phrasing to a text-to-speech system will enhance synthetic speech both by breaking a run-on delivery into more easily processed chunks of speech and by allowing the pitch range to be increased, enhancing naturalness. The primary requirement for making synthetic speech easier to understand, however, is that the phrasing system should first do no harm, i.e., it should avoid misleading phrasings and phrasings like 38 that are difficult for a listener to process.

38. ??This is || the cat that caught || the rat that stole || the cheese.

To address the issue of avoiding undesirable phrasings, a system like ours must be run against a set of utterances of disparate lengths and syntactic types to see where it breaks. To do this, we accumulated a set of syntactically varied test sentences, had them read by two speakers, marked the prosodic phrase boundaries as we had done with the Holmes and Everest data, and ran the sentences through our prosodic phrase system for comparison with the human speakers' phrasing. Phrasing in the speech of the two human speakers was almost identical; where the speakers failed to match, we depended on the speaker with the more detailed phrasing to obtain a single transcript for comparison. The sentences with the phrase markings produced by the system (S) and by the human speakers (H) are given in Appendix A. In addition, we used the sentences of Grosjean et al. (1979) to test our system; these were the same sentences used by G&G. They were produced by six subjects speaking each sentence at five different speaking rates. We chose these data both because they form an established, though small, corpus and because their phrase markings, derived by oscillographic tracings of pause duration, provide a check against the possibility of error in human judgments of pause location and salience. A comparison of these data with the output of our system is given in Appendix B.

In comparing human and synthetic prosodic phrasing with an eye on synthetic phrasings that are either misleading or unprocessable, errors in the assignment of primary phrase boundaries are the most egregious. In the data in Appendix A, the human speakers produced 31 intrasentential primary phrase boundaries, of which our system matched 16 with its equivalent of a primary phrase boundary. The other primary phrase boundaries in the human productions were matched in 9 cases with a secondary phrase boundary, and in 5 cases with a tertiary boundary. The system thus missed 1 primary boundary, the one after *books* in 17. In the absence of research on the relative significance of different boundary types, we are assuming that, with respect to the comprehensability and acceptability of synthetic speech, the difference between a primary and secondary phrase boundary is minimal. On the other hand, the tertiary boundary produced by our system is almost imperceptible and cannot be considered equivalent to the primary boundary. The system thus matched, either exactly or approximately, 80% of the primary boundaries.

In looking at the primary phrase boundaries that the system failed to duplicate, we want to be sure that we avoid difficult phrasings like that of 38 rather than match any particular phrasing exactly. In sentences 9, 12, 17, and 21, where our system produced a tertiary boundary at the location of the human speakers' primary phrase boundary, comprehensibility has not been diminished. The listener is misled only in sentence 17, where our system generated a boundary at a location different from that of the human speakers. In the system's phrasing of 17, the boundary before *books* corresponds to an interpretation of the utterance in which the books are written by her uncle, which, while it is a possible phrasing, presumes a specific discourse setting.

A comparison of the secondary phrase boundaries in Appendix A shows similar correspondences. The human speakers produced 26 secondary phrase boundaries. The

system matched 11 exactly and 12 with a primary or tertiary boundary. It thus failed to match 3 secondary boundaries, in sentences 9, 18, and 20, but only in sentence 18 is the absence of the boundary misleading.

To sum up, there are 2 boundaries out of the 57 produced by the human speakers whose absence in the system's output is truly problematic. Evaluation of the prosodic phrase system, however, must also consider cases of overgeneration. It is significant here that in the 21 sentences of Appendix A, the system never generated a primary phrase boundary that did not correspond to a boundary in the human productions, although the system overgenerated secondary phrase boundaries at 7 locations. (We ignore the overgeneration of tertiary boundaries as perceptually insignificant.) Three of these secondary boundaries are problematic. Those in sentences 17 and 18 were discussed above as resulting from the misplacement of a boundary. The other unwanted secondary boundary occurs after *was* in sentence 10 and results in an utterance that, like 38, is difficult to process.

A comparison of our system's output with G&G's productions, given in Appendix B, shows similar results. Of the 14 primary phrase boundaries in G&G's sentences, the system matched 12 exactly. There were 2 primary boundaries, however, that the system missed completely. Both of these errors were due to the presence of a sentential subject. For sentence 6, no parse was produced, but otherwise the inability of our system to generate the correct result here stems from the fact that the system discards the syntactic sentence node in the derivation of the prosodic phrasing.[15] With respect to overgeneration of phrase boundaries, the system overgenerated seriously only once, in sentence 4, where the subject is sentential.

We find these results encouraging; with respect to matching, there were only four significant problems in the two corpora, and overgeneration of a primary boundary occurred only once. The suitability of our system for speech applications will depend on future tests to determine whether listeners prefer "prosodized speech" that is imperfect, i.e., speech that will have some phrasing errors, to the relatively "flat" speech of systems that lack our phrasing rules. The weak link in our current system is the parser: most problems with phrasing arise from parsing errors; in particular, incorrect part of speech assignment, incorrect analysis of pre-head modifiers, and failure to recognize idiomatic or semi-idiomatic expressions. Problems with the prosody rules come mainly from phenomena that we have not adequately studied, e.g. the proper treatment of material on the left periphery of a prosodic verb phrase and the status of complements to nouns and adjectives.

4 Conclusions

We have discussed the notion of discourse neutral prosodic phrasing in English and presented an analysis that characterizes this phrasing in terms of constituency, adjacency, and length. In our analysis, the contribution of syntax to discourse-neutral phrasing consists of lexical categorization; NP, PP, and AdjP constituency; and syntactic head identification. Length is an independent phonological factor. Because they refer to both syntactic and phonological information, phrasing rules are free to generate prosodic structures that may or may not resemble syntactic structures. Hence, in speech, it is possible but not expected that phrase boundaries will co-occur with major syntactic boundaries.

Our results suggest that, in an implemented system, the parsing requirements for speech systems are quite different from those for systems providing information retrieval, machine translation, or text generation. In particular, there seems to be no need for a parser to identify VP and S constituents, nor to specify predicate-argument relations.

The distribution of the phrasing of clause-final PPs given in Figure 4 may indicate the extent of the relation between the discourse neutral phrasing and the phrasing imposed by discourse. We assume that discourse phrasing may shift neutral boundaries in order to reflect, for example, emphasis, contrast, parallelism, coreference, and the particular structure of the discourse. The exact connection between the level of phrasing we describe and discourse-dependent phrasing is a question for future research. We need to know what aspects of the discourse are relevant for phrasing, as well as how the discourse information and the phrasing specifications should be related. For example, it is not clear whether the discourse-neutral phrasing represents a set of pre-determined values that are reset when necessary by discourse features or whether this is a true default situation in which the discourse neutral phrasing is inserted when discourse phrasing is underspecified. Any contribution in this area will greatly enhance our understanding of the relation between the various components of the grammar—syntactic, semantic, and phonological.

Acknowledgments

We are grateful to Francois Grosjean and Terry Langendoen for their comments on an earlier version of this paper, and to Jack Lacy for his invaluable assistance with the implementation and testing of the system. Special thanks go to two anonymous *Computational Linguistics* reviewers whose comments have greatly helped us improve the original manuscript. None of those mentioned is responsible for any shortcomings of the work described here.

References

Allen, G. 1975 Speech Rhythm: Its Relation to Performance Universals and Articulatory Timing. *Journal of Phonetics* 3: 75–86.

Allen, J. 1976 Synthesis of Speech from Unrestricted Text. *Proceedings of the IEEE* 4: 433–442.

Bachenko, J.; Hindle, D.; and Fitzpatrick, E. 1983 Constraining a Deterministic Parser. *Proceedings of the National Conference on Artificial Intelligence* (AAAI-83).

Bachenko, J.; Fitzpatrick, E.; and Wright, C. E. 1986 The Contribution of Parsing to Prosodic Phrasing in an Experimental Text-to-Speech System. *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 145–153.

Bierwisch, M. 1966 Regeln fur die Intonation deutscher Satze. In: Bierwisch, M. (ed.), *Studia Grammatica VII: Untersuchungen uber Akzent und Intonation im Deutschen.* Akademie-Verlag, Berlin: 99–201.

Bing, J. 1985 *Aspects of Prosody.* Garland Press, New York, New York.

Cahn, J. 1988 From Sad to Glad: Emotional Computer Voices. *Proceedings of Speech Tech '88:* 35–36.

Chomsky, N. 1965 *Aspects of the Theory of Syntax.* MIT Press, Cambridge, MA.

Chomsky, N. and Halle, M. 1968 *The Sound Pattern of English.* Harper & Row, New York, New York.

Church, K. 1988 A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text. *Proceedings of the Second Conference on Applied Natural Language Processing (ACL):* 136–143.

Cooper, W. E. 1976 Syntactic Control of Timing in Speech Production: A Study of Complement Clauses. *Journal of Phonetics* 4: 151–171.

Cooper, W. and Paccia-Cooper, J. 1980 *Syntax and Speech.* Harvard University Press, Cambridge, MA.

Crystal, D. 1969 *Prosodic Systems and Intonation in English.* Cambridge University Press, Cambridge, U.K.

Dommergues, J.-Y. and Grosjean, F. 1981 Performance Structures in the Recall of Sentences. *Memory and Cognition* 9: 478–486.

Downing, B. 1970 *Syntactic Structure and Phonological Phrasing in English.* Ph.D. Dissertation, University of Texas, Austin, TX.

Elovitz, H.; Johnson, R.; McHugh, A.; and Shore, J. E. 1976 Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 6: 446–459.

Emorine, O. M. and Martin, P. M. 1988 The Multivoc Text-to-Speech System. *Proceedings of the Second Conference on Applied Natural Language Processing (ACL):* 115–120.

Fitzpatrick, E. 1989 The Preferred Prosodic Phrasing of Prepositional Phrases. Unpublished data.

Fitzpatrick, E. and Bachenko, J. 1989 Parsing for Prosody: What a Text-to-Speech System Needs from Syntax. *Proceedings of IEEE Artificial Intelligence Systems in Government (AISIG) '89.*

Gee, J. P. and Grosjean, F. 1983 Performance Structures: A Psycholinguistic and Linguistic Appraisal. *Cognitive Psychology* 15:411–458.

Grosjean, F. and Gee, J. P. 1987 Prosodic Structure and Spoken Word Recognition. *Cognition* 25: 135–155.

Grosjean, F.; Grosjean, L.; and Lane, H. 1979 The Patterns of Silence: Performance Structures in Sentence Production. *Cognitive Psychology* 11: 58–81.

Hillinger, M.; James, C. T.; Zell, D. L.; and Prato, L. M. 1976 The Influence of Prescriptive and Subjective Phrase Markers on Retrieval Latencies. *Bulletin of the Psychonomic Society* 8: 353–355.

Hindle, D. 1983 User Manual for Fidditch, a Deterministic Parser. *NRL Technical Memorandum* #7590-142.

Hirschberg, J. and Litman, D. 1987 Now Let's Talk About Now: Identifying Cue Phrases Intonationally. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics,* 163–171

Klatt, D. H. 1987 Review of Text-to-Speech Conversion for English. *Journal of the Acoustic Society of America* 82: 737–793.

Klatt, D. H. 1975 Vowel Lengthening Is Syntactically Determined in Connected Discourse. *Journal of Phonetics* 3: 129–140.

Koster, Jan. 1978 Why Subject Ss Don't Exist. In: Keyser, Samuel J. *Recent Transformational Studies in European Languages.* Linguistic Inquiry Monograph 3. MIT Press, Cambridge, MA.

Langendoen, D. T. 1975 Finite-State Parsing of Phrase-Structure Languages and the Status of Readjustment Rules in Grammar. *Linguistic Inquiry* 6: 533–554.

Litman, D. and Hirschberg, J. 1990. Disambiguating Cue Phrases in Text and Speech. *Proceedings of the 13th International Conference on Computational Linguistics (COLING).*

Liberman, M. Y. and Prince, A. 1977 On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8: 249–336.

Luce, P. A.; Feustel, T. C.; and Pisoni, D. B. 1983 Capacity Demands in Short-Term Memory for Synthetic and Natural Speech. *Human Factors* 25: 17–32.

Martin, E. 1970 Toward an Analysis of Subjective Phrase Structure. *Psychological Bulletin* 74: 153–166.

Nespor, M. and Vogel, I. 1986 *Prosodic Phonology.* Foris Publications, Dordrecht, The Netherlands.

Olive, J. P. and Liberman, M. Y. 1985 Text-to-Speech—An Overview. *Journal of the Acoustic Society of America,* Supplement 1: 78, S6.

O'Shaughnessy, D. D. 1989 Parsing with a Small Dictionary for Applications such as Text-to-Speech. *Computational Linguistics* 15: 97–108.

Pierrehumbert, J. B. 1980 The Phonetics and Phonology of English Intonation. Ph.D. Dissertation, MIT, Cambridge, MA.

Selkirk, E. O. 1984 *Phonology and Syntax: The Relation between Sound and Structure.* MIT Press, Cambridge, MA.

Sproat, R. W. and Liberman, M. Y. 1987 Toward Treating English Nominals Correctly. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics,* 140–146.

Streeter, L. A. 1978 Acoustic Determinants of Boundary Perception. *Journal of the Acoustical Society of America* 64: 1582–1592.

APPENDIX A

Human (H) and system (S) productions of syntactically varied sentences. Primary boundaries are marked by "‖", secondary by "|", tertiary by "!".

1. H: *The name | of the character ‖ is not pronounced.*
   S: *The name ! of the character | is not pronounced.*
2. H: *The left-hand power unit ! on each shelf | in the forty-eight channel module ‖ operates ! the echo cancellers.*
   S: *The left-hand power unit | on each shelf | in the forty-eight channel module ‖ operates | the echo cancellers.*
3. H: *Phoneme characters ‖ give more control ‖ over the particular sounds ‖ that are generated.*
   S: *Phoneme characters | give ! more control | over the particular sounds ‖ that are generated.*
4. H: *The connection must be determined ‖ for the left-hand power unit ‖ on each shelf.*
   S: *The connection ! must be determined | for the left-hand power unit ‖ on each shelf.*
5. H: *I need a man ‖ to fix the sink.*
   S: *I need a man | to fix ! the sink.*
6. H: *The techniques | that we had implemented ‖ were tested | on a larger computer.*
   S: *The techniques | that we had implemented ‖ were tested | on a larger ! computer.*
7. H: *Everyone | who had participated | in the attempt ‖ was considerably affected.*
   S: *Everyone | who had participated ! in the attempt ‖ was considerably affected.*
8. H: *The method | by which one converts a word | into phonemes ‖ is provided | in chapter seven.*
   S: *The method | by which one converts ! a word | into phonemes ‖ is provided ! in chapter seven.*
9. H: *In these instances ‖ it may be desirable | to use phonemic characters ‖ each time | that it appears ‖ on the input text.*
   S: *In these ! instances | it may be ! desirable ‖ to use ! phonemic characters ‖ each time that it appears ! on the input text.*

10. H: *The thrust | was now from the south ‖ which Mallory had deemed impossible.*
    S: *The thrust | was | now ! from the south ‖ which Mallory | had deemed ! impossible.*
11. H: *The destruction | of the good name | of his father ‖ bothered him.*
    S: *The destruction | of the good name ! of his father ‖ bothered him.*
12. H: *Every event | of that dreadful time ‖ is seared ‖ into my memory.*
    S: *Every ! event | of that ! dreadful time ‖ is seared ! into my memory.*
13. H: *Everest was discovered ‖ during a survey of India ‖ in 1852.*
    S: *Everest was discovered ‖ during a survey ! of India | in 1852.*
14. H: *He told the director ‖ to give the names | of the characters ‖ to Ivan.*
    S: *He told ! the director ‖ to give | the names ! of the characters | to Ivan.*
15. H: *It may be impossible ‖ to give that machine ‖ the proper workout.*
    S: *It may be ! impossible ‖ to give that machine | the proper workout.*
16. H: *Eventually | he will realize ‖ that his cigars are bothering ‖ the other passengers.*
    S: *Eventually ‖ he will realize ! that his cigars | are bothering ! the other passengers.*
17. H: *She was given | more difficult books ‖ by her uncle.*
    S: *She was given ! more difficult | books by her uncle.*
18. H: *You could easily | break that vase ‖ if you aren't careful.*
    S: *You could easily break | that ! vase ‖ if you aren't careful.*
19. H: *The president | asked the group ‖ what they were capable | of doing.*
    S: *The president | asked ! the group ‖ what they were | capable ! of doing.*
20. H: *What book | on the subject ‖ would you recommend | to the group?*
    S: *What book ! on the subject ‖ would you recommend to the group?*
21. H: *I can usually read | a lot faster ‖ than Roger.*
    S: *I can usually read | a lot faster ! than Roger.*

Appendix B

Test sentences used by G&G with their prosodic marking converted to our notation. *S* represents our system's productions, *G&G* represents the human productions.

1. S: *When the new lawyer | called up ! Reynolds ‖ the plan ! was discussed ! thoroughly.*
   G&G: *When the new lawyer ! called up Reynolds ‖ the plan ! was discussed | thoroughly.*
2. S: *In addition ! to his files ‖ the lawyer ! brought | the office's best adding machine.*
   G&G: *In addition ! to his files ‖ the lawyer brought | the office's ! best ! adding machine.*
3. S: *By making ! his plan | known ‖ he brought out | the objections ! of everyone.*
   G&G: *By making ! his plan known ‖ he brought out | the objections ! of everyone.*
4. S: *That a solution ‖ couldn't be found seemed | quite clear ‖ to them.*
   G&G: *That a solution ! couldn't be found ‖ seemed ! quite clear | to them.*
5. S: *Not quite all ! of the recent files ‖ were examined | that ! day.*
   G&G: *Not quite ! all | of the recent ! files ‖ were examined ! that day.*
6. S: Too many parse problems
   G&G: *That ! the matter ! was dealt with | so fast ‖ was a shock ! to him.*
7. S: *John ! asked | the strange ! young ! man ‖ to be ! quick ‖ on the task.*
   G&G: *John ! asked ! the strange young man ‖ to be quick | on the task.*
8. S: *Closing ! his client's ! book ‖ the young expert | wondered ! about this ! extraordinary story.*
   G&G: *Closing his client's book ‖ the young expert | wondered about ! this ! extraordinary story.*
9. S: Parse problems with *wh* trace in infinitival relatives; change wording slightly: *The expert | who didn't know ! what to tell us ‖ sat back ! in despair.*
   G&G: *The expert | who couldn't see ! what to criticize ‖ sat back ! in despair.*
10. S: *After the cold winter ! of that year ‖ most people | were ! totally fed up.*
    G&G: *After the cold winter | of that year ‖ most people ! were totally ! fed up.*
11. S: *The agent | consulted ! the agency's book ‖ in which they offered ! numerous tours.*
    G&G: *The agent | consulted ! the agency's book ‖ in which ! they offered numerous tours.*
12. S: Parse problems with *surprisingly* and the main verb; change wording slightly: *She discussed | the pros and cons ‖ to overcome | surprisingly apprehensive feelings.*
    G&G: *She discussed ! the pros and cons ‖ to get over her surprisingly | apprehensive ! feelings.*
13. S: *Our disappointed woman | lost ! her optimism ‖ since the prospects ! were too limited.*
    G&G: *Our disappointed woman | lost her ! optimism ‖ since the prospects ! were too limited.*
14. S: *Since she was ! indecisive ! that day ‖ her friend | asked her to wait.*
    G&G: *Since ! she was | indecisive | that day ‖ her friend ! asked her to wait.*

Notes

1. Bachenko et al. (1983) outlines the main features of the syntactic framework we assume. Any syntactic approach that provides the

lexical and constituent information discussed in Section 2 should be sufficient for the phrasing analysis we present, but this idea remains to be explored.

2. e.g. Cooper (1976), Cooper and Paccia-Cooper (1980), Klatt (1975), and Streeter (1978). Downing (1970), like Langendoen, explicitly assumes a direct connection between syntax and prosodic phrasing with "some aspects at least of surface structure ... determined exclusively by the necessity of providing input to the phonological rules that specify prosodic features" (p. 204).
3. In characterizing this phrasing as discourse-neutral, we are assuming, contra Bing, that length is independent of discourse.
4. Both examples are from Martin (1970). Bierwisch (1966) also includes length, in terms of stressed syllable count, as a factor in the prosodic phrasing of German auxiliary + verb strings. Length as a factor in prosodic phrasing is mentioned in Selkirk (1984) and Nespor and Vogel (1986), but without any systematic account of its effects.
5. G&G's test corpus consists of 14 sentences that were originally used in Grosjean et al. (1979).
6. A more accurate account of adjunction, such as that in Selkirk (1984), would limit the list of adjoinable words to those that have both a strong and a weak form. For example, the word *him* in *see him* may be realized as the strong form [hIm], where the vowel carries stress, or it may take the weak unstressed form [m̩]. This word is therefore a candidate for adjunction. In contrast, a preposition such as *among* has no weak form and would be nonadjoinable. Results that we discuss below support this approach. Our current system, however, relies solely on lexical category and tree structure to identify adjunction possibilities.
7. These are essentially the *phi* phrases of G&G.
8. Below, in the discussion of constituent length, we will claim that word count should actually be stressed syllable count. In the current system, however, length depends on phonological word count.
9. Problems with final phrase boundaries are discussed in more detail in Bachenko et al. (1986), where the analysis presented assumes the version of final bundling proposed in G&G. We return to this type of example below.
10. Prosodic events finer than the pause were marked but ignored for this study. A stress foot contains one stressed syllable and zero or more unstressed syllables. In our current implementation, length is measured by phonological word count, since stress foot information is not readily available to us.
11. Conjunctions of lexical items, e.g. *some subtle and horrible crime, these windows and shutters,* are rarely split into two phrases.
12. Klatt (1987) presents a useful review of text-to-speech technology and performance.
13. Fitzpatrick and Bachenko (1989) provide a more detailed discussion of the differences between parsing for speech and parsing for text understanding.
14. Hindle (1983); lexical lookup and category disambiguation are done by the stochastic parser described in Church (1988).
15. We have as yet no account of pre-verbal embedded sentences, though an analysis that does not involve an embedded S node is available (Koster, 1978).