TABLE OF CONTENTS

1. List of papers

This thesis consists of a summary and the following papers:

I. ENGLISH SENTENCE STRESS, GRAMMATICAL FUNCTIONS AND CONTEXTUAL COREFERENCE
Merle Horne
Studia Linguistica 39: 51-66 (1985)

II. INFORMATION FOCUS: ASSIGNMENT AND PHONOLOGICAL IMPLICATIONS
Merle Horne
In: L. Evensen (ed.), Nordic Research in Text Linguistics and Discourse Analysis, 155 - 70. University of Trondheim: TAPIR (1986)

III. FOCAL PROMINENCE AND THE 'PHONOLOGICAL PHRASE' WITHIN SOME RECENT THEORIES
Merle Horne
Studia Linguistica 40: 101-21 (1986)

IV. TOWARDS A QUANTIFIED, FOCUS-BASED MODEL FOR SYNTHESIZING SENTENCE INTONATION IN ENGLISH
Merle Horne
Working Papers (Lund Univ., Dept. of Linguistics) 31: 41-86 (1987) (Also to appear in Lingua)

V. 'CONTRASTIVE PROMINENCE' AND SYNTACTIC PARALLELISM
Merle Horne
Manuscript (Lund Univ., Dept. of Linguistics) (1987)

2. Introduction

This thesis presents the development of a grammatically and textually based model for describing and synthesizing English declarative sentence intonation. Within the past decade, the area of intonation synthesis (and speech synthesis in general) has become the center of intense research due primarily to the possibilities made available for this work by developments in computer hard- and software. Text-to-speech research has blossomed during this period, and currently, there are several systems in operation, e.g., KTH text-to-speech (see Carlson & Granström 1976), MITalk system (see Allen et al. 1987), Edinburgh text-to-speech system (see Ladd 1987, Monaghan 1987). The wealth of papers presented on this subject at the 11th International Congress of Phonetic Sciences, Tallinn, 1987 (see e.g. Eady & Dickson 1987) and the European Conference on Speech Technology, Edinburgh, 1987, bear witness to the importance placed on the research that will lead to a better understanding of the rule systems that lie behind the production of human speech in a discourse context. As this thesis shows, discourse rule systems cannot be equated with those needed for accounting for sentence grammar phenomena. The model presented here has not as yet been implemented in any program but it is designed so that it could be, given the possibility to automatically extract the syntactic and contextual parameters that are assumed to condition intonation patterning in English.

3. The notion of 'sentence stress'

Paper I presents a general overview of previous work on predicting the placement of 'sentence stress', as well as a

first approximation of our own algorithm for assigning the highest degree of prominence in a sentence.

It has been generally assumed in the phonological literature that a given sentence in isolation, out of context, will have one particular word that bears more prominence than all the others. This prominence has usually been referred to as 'sentence stress' in the American literature and 'nucleus' in the British literature. As regards explanations for the particular position of this sentence stress, one can divide the approaches roughly into a syntactic approach and a semantic/pragmatic approach.

The syntactic approach is represented for example, by Chomsky & Halle (1968) and Bresnan (1971). Chomsky & Halle extended their Nuclear Stress Rule (NSR) which accounts for stress assignment in English phrases to account for sentence prominence by allowing it to apply cyclically to surface structure (the NSR assigns sentence stress to the last primary stressed word of a sentence). Bresnan (1971) attempted to account for certain exceptional cases to the NSR (i.e. cases where the sentence stress fell on a non-final word) by claiming that the NSR applied after all the syntactic transformations on each transformational cycle. Bresnan's account, however, was criticized by Berman & Szamosi (1972) on several accounts, e.g. it does not account for cases where 'sentence stress' falls on the subject, as e.g. in "'The sky is falling', cried Chicken Little". In a reply, Bresnan (1972) admitted that syntax alone does not condition all instances of sentence stress placement. In a further attempt to save the NSR, however, she introduced a new rule, that of Topical Stress Assignment that could (in some unspecified way) account for sentences with sentence stress on

the subject. This work, where she introduced semantic notions, was followed by the highly critical article by Bolinger (the main proponent of the semantic approach to sentence stress) in 1972 who maintained that "the distribution of sentence accents is not determined by syntactic structure but by semantic and emotional highlighting" (1972:644). However, despite Bolinger's pessimistic claim that "Accent is predictable (if you're a mind reader)" most researchers still shared the feeling that there was more grammatical patterning behind the dynamics of sentence stress than Bolinger would lead one to believe (for a more recent account of sentence stress in terms of a pragmatic notion of 'Dominance', see Erteschik-Shir & Lappin 1983).

4. Discourse and information focus

Attention shifted to the context of discourse and attempts to understand sentence stress by relating it to discourse functions. This approach had, in fact, been advocated several years earlier by Gunter (1966) who showed the importance of studying sentences in context in order to understand the dynamics of sentence intonation. Bruce (1977) also demonstrated the significance of taking a sentence-in-context perspective in the analysis of Swedish word accents. Firbas (1979), working in a Prague School 'functional sentence perspective', studied prosody with respect to the concept of 'communicative dynamism', theme/rheme dynamics in a discourse. The term 'information focus', instead of sentence stress, is used by Halliday (1967) to describe the prominence given to new information "...in the sense that the speaker presents it as not being recoverable from the preceding discourse" (1967:204).

Moreover, 'deaccenting' due to contextual factors is also taken up and discussed more in the literature (see Vanderslice & Ladefoged 1972). Chomsky (1971), Jackendoff (1972) and Quirk et al. (1972) also use the term 'focus' for that constituent of a sentence that is not 'presupposed'. Presupposition is thus sentence grammar's way of attempting to deal with what is in fact context-related information needed for accounting for various externally conditioned phenomena in isolated sentences.

Our goal has been to develop a discourse-based model of information focus assignment and projection that theoretically could be implemented in a text-to-speech program. The development of rule systems for the assignment of focus with respect to discourse, is something that has not been dealt with to any great extent in the linguistic literature. A number of trends have, however, been discussed. Schmerling (1976) presents what appears to be a strong tendency in discourse data and formulates it as her Principle II which says that Predicates receive more prominence than arguments. Ladd (1978) and Bing (1980), in their more syntactic approach, express this tendency in terms of syntactic categories, i.e. noun phrases (unless 'deaccented' due to discourse reasons) receive more prominence than verb phrases. As far as specific rules are concerned, Contreras (1976), for example, presents an analysis of theme/rheme in Spanish based on a hierarchy of semantic roles. Moreover, Gussenhoven (1983), develops a set of rules that assign 'focus accents' to semantic constituents: 'Predicates', 'Arguments' and 'Conditions' after determining a number of 'focus domains'. The work presented in this thesis follows the discourse-oriented approach of these researchers by attempting to systematize more facts about English sentence intontion.

5. Model presentation

Our own approach has been to deal both with the grammar/discourse side and the acoustic/phonetic side of information focus by developing a model that both assigns focus and describes its phonetic realization with respect to the parameter of pitch (F_0). It is geared to a text-to-speech situation and as such, its initial goal is to limit its scope to 'non-expressive' intonation, since expressive intonation is, at least at the current state of our knowledge, governed by factors which are not amenable to description by rule.

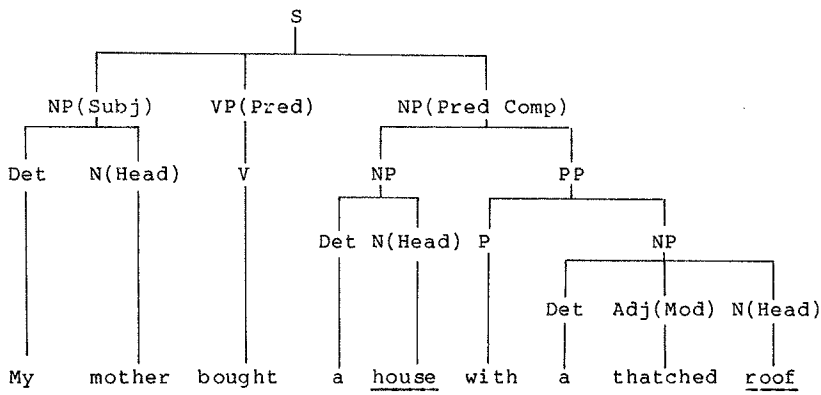
Our model is not based on a grammatical description in terms of semantic constituents nor in terms of syntactic categories (NP,VP), or lexical categories (N,V, Adj), but rather is related to the level of grammatical functions (Subject, Predicate, Predicate Complement). In Paper I, it is claimed that the position of sentence stress (greatest focal prominence) is conditioned by a hierarchy of grammatical functions interacting with a discourse parameter of coreferentiality[1], viz:

Predicate Complement > Subject > Predicate

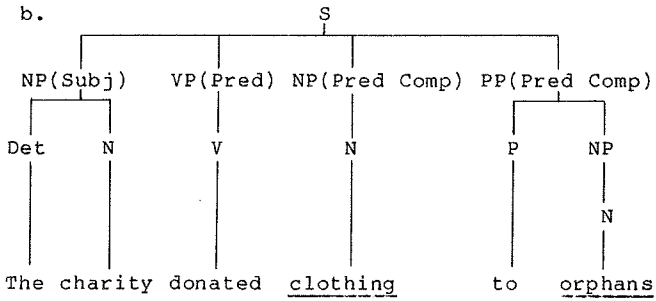
so that, in an all-new sentence (e.g. one that occurs at the beginning of a text), the sentence stress would fall on the predicate complement if there was one, otherwise on the subject. If there was no subject (e.g., in an imperative sentence), then the sentence stress would fall on the predicate. Notice that with this hierarchy, sentences with most prominence on the subject are not treated as deviant cases (see also Figure 1 below which shows how the hierarchy can be built into an algorithm for assigning information focus).

Projection of the sentence stress within the constituent realizing the grammatical function is handled by a subroutine in the hierarchy which projects sentence stress onto the heads of phrases making up the constituent in an all-new sentence (where 'head' is to be interpreted in the structuralist sense of 'center of an endocentric construction'). Provision is also made for projecting the stress onto the modifier in a subsequent part of the text when, for instance, the head of a constituent could be coreferential with something in a preceding part of the discourse. Moreover, as is discussed in Paper II, the assignment of focus is assumed to apply recursively to headed phrases within a given constituent realizing a grammatical function as well as iteratively on the level of grammatical functions so as to project focus correctly on e.g. the underlined words in the sentences in (1):

(1) a.



(recursive application of focus assigning rule)



(iterative application of focus assigning rule)

This type of rule application obviates the need for readjustment rules such as those developed by Cullicover & Rochemont (1983) for flattening out surface structure into prosodic structures that limit the number of prominence levels that can be derived (in their case, by node counting).

In Paper I, we were mainly interested in predicting the position of greatest focal prominence in a sentence (sentence stress). There are, however, other lesser degrees of prominence within a sentence and in Paper II, it was suggested that the hierarchy developed earlier to predict the greatest degree of sentence prominence (sentence stress) could be used to assign other prominences. We referred to these also as focal prominences since they are also associated with information that cannot be extracted from earlier parts of the text. It was thus assumed that the hierarchy, Pred.Comp. > Subj. > Pred. would be associated with levels of prominence so that in an all-new sentence containing all three grammatical functions, the Predicate Complement would be associated with the highest degree of prominence, the Subject with a lesser degree of prominence, and the Predicate with a tertiary degree[2].

6. Realization of focus as pitch (F_0)

These theoretical levels were later found to be associated with actually occurring F_0 values. In Paper III, empirical data are presented that show how the different levels of abstract prominence can be related to varying heights of F_0 peaks described with reference to the baseline of a phonological grid such as that used by Gårding (1983) for describing sentence intonation. We have not dealt with the parameter of duration as regards its role in the realization of focus, since it is generally assumed that pitch is the principal indicator of focus in English. However, a complete account of focus would, of course, have to take into consideration the cooccurring effects of duration (see e.g. Bannert 1986, Touati 1987). A flow-diagram showing how the grammatical function hierarchy was built into an algorithm for assigning information focus and describing its realization in terms of pitch (F_0) is presented in Figure 1.

Figure 2 presents a grid analysis of the F_0 curve obtained for the last sentence of the text fragment in (2). (3) gives a parse of this sentence done in the framework of referent grammar (see Sigurd 1987a,b). Referent grammar, with its built-in numbered referents, provides an appropriate syntactico-semantic representation for the focus-assigning algorithm to process:

(2) A: I'm just about finished writing my new book.

B: Oh, could you let me in on how it's going to end?

C: Yea, sure. A Mormon will marry a mayor.

(3) s(subj(np(nr4,nom(mormon,sg,indef))),

pred(v(vr6,nom(marry,fut))),

pred comp(np(nr5,nom(mayor,sg,indef))))

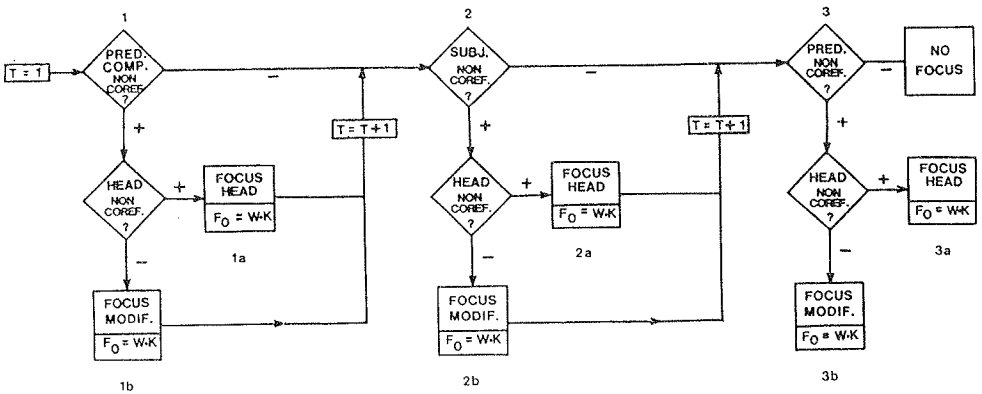


Figure 1. Model (flowchart) for assigning information focus to constituents on the basis of grammatical functions and the coreferential status of the lexical material realizing a particular function. The input to the model is a given sentence (S). Focus is realized as pitch (F_0) according to the equation $F_0 = W.K$ where F_0 here refers to the relative height of a given pitch obtrusion, W designates the width of the grid within which F_0 moves and K is a variable ranging over a number of prominence levels defined as fractions of the distance from the baseline to the topline of the grid. In Fig. 2, K assumes the values 1 (for the first focussed constituent), 0.8 (for the second focussed constituent), and 0.4 (for the third focussed constituent). For the syntheses done in Paper IV, however, the values were 1, 0.75, and 0.5, respectively. The box, T=T+1, is a counter which adds 1 each time focus is assigned. T is used in determining the coefficient, K: If T=1, -> K=1, if T=2 -> K=0.75, and if T=3 -> K=0.5. The diagram is to be read as follows: 1.: check to determine if there is a predicate complement that is non-coreferential with something in the preceding part of the text. If there is one, check whether it is the head that is non-coreferential. If this condition is met, focus the head, assigning it a level of prominence where $F_0 = W.K$ (1a). If the head is coreferential, assign the modifier focal prominence instead (1b). Go to the subject (2) and repeat the same routine, and then go to predicate (3), again repeating the same routine.

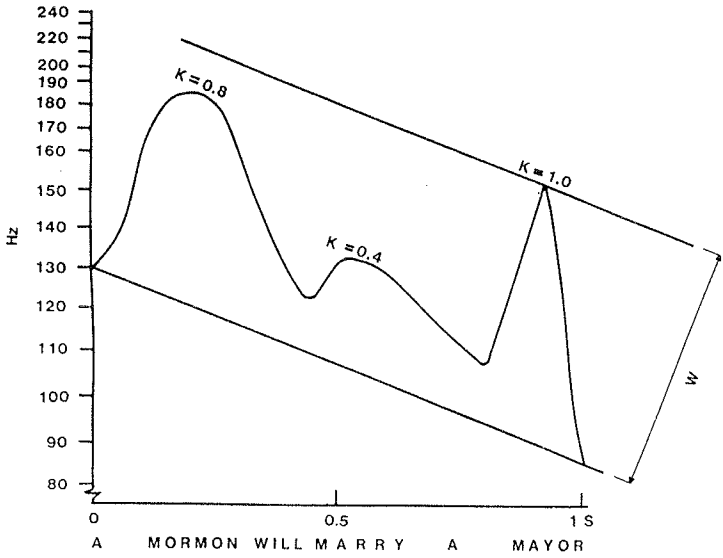


Figure 2. Actually occurring F_0 curve obtained for a reading of the last sentence in (2) where the Subject, Predicate and Predicate Complement are focussed according to the model in Figure 1.

With respect to the preceding context, the sentence A Mormon will marry a mayor in (2) contains no coreferential material. Consequently, the constituents realizing the Subject, Predicate and Predicate Complement are all assigned focal prominence in accordance with Figure 1: the Predicate Complement (mayor) receives more prominence than the Subject (Mormon), which in turn is assigned more prominence than the predicate (marry).

The analysis of the F_0 contour in Figure 2 was based on the following assumptions: We hypothesized that the baseline of the grid corresponded to the speaker's sentence initial and final voice frequency. These were determined to be approximately 130Hz and 85Hz, respectively after comparison with other utterances produced by the same speaker (an American male). The topline of the grid was drawn parallel to the baseline and passes through the top of the peak corresponding

to what the model in Figure 1 would predict as receiving the greatest degree of prominence (in this case, the object, mayor). This highest peak was assumed to define the width (W) of the grid and the other peaks were defined as fractions of the distance from the baseline to the topline of the grid. Figure 3 presents the same sentence in another contextual environment, i.e. one in which the predicate and the predicate complement are given. Focus is realized only on the subject, but the amount of prominence was observed to be equal to that on the object in Figure 2 when analyzed in terms of the grid.

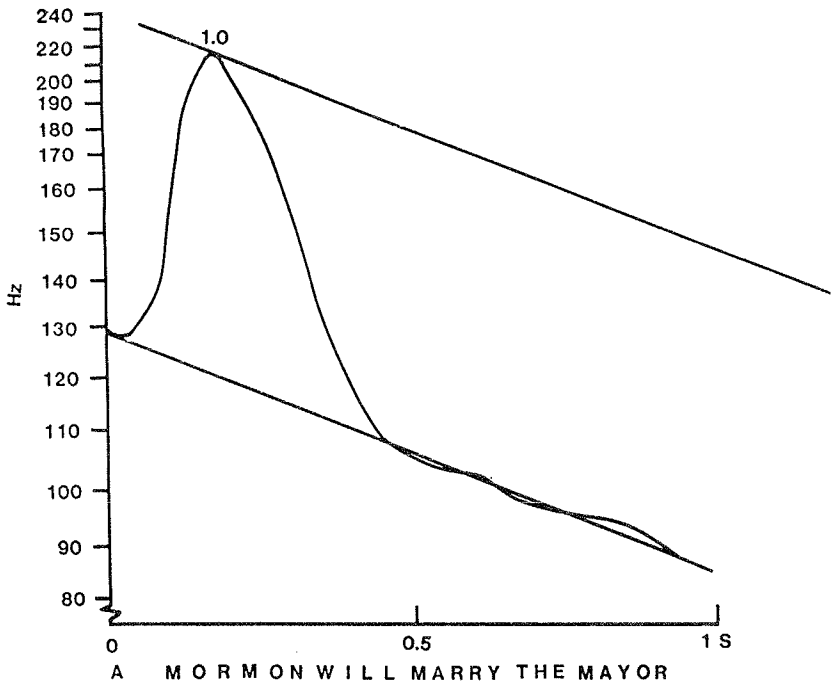


Figure 3. Actually occurring F_0 curve obtained for same sentence as in Figure 2, but where only the subject is focussed according to the algorithm in Figure 1.

7. Pitch generating algorithm

Paper IV elaborates on the empirical observations presented in Paper III, and develops a pitch generating algorithm for synthesizing focal prominence on the basis of the model.

The abstract grid was defined on the basis of the empirical data presented in Paper III with a certain amount of 'rounding off' taking place as regards the speaker's F_0 range and the values assigned to the various levels of focal prominence (see Figure 2). For example, we decided to fix the F_0 range at 1 octave (low point = 90Hz, high point = 180Hz). This range, which is not unnatural for a given speaker, also has the advantage of more or less eliminating the discrepancies that could potentially arise in translating between a linear scale (which is what our F_0 editing program works on) and a logarithmic scale, which is assumed to better correspond to the way speakers perceive pitch. Furthermore, the relative degrees of prominence given in Figure 2 were also rounded off so that the predicate was assigned a level corresponding to .5W, the subject, a level equal to .75W and the predicate complement, a prominence level corresponding to W in an all new sentence. Moreover, in our subsequent analyses by synthesis, we decided to attribute a phonetic reality to the baseline, i.e. it was decided that the baseline would be realized phonetically over stretches of nonfocussed material. The topline, however, is not ascribed any phonetic reality; it functions solely as a reference line for computing F_0 obtrusion levels. Figure 4 presents the phonological grid used for defining levels of F_0 prominence associated with information focus.

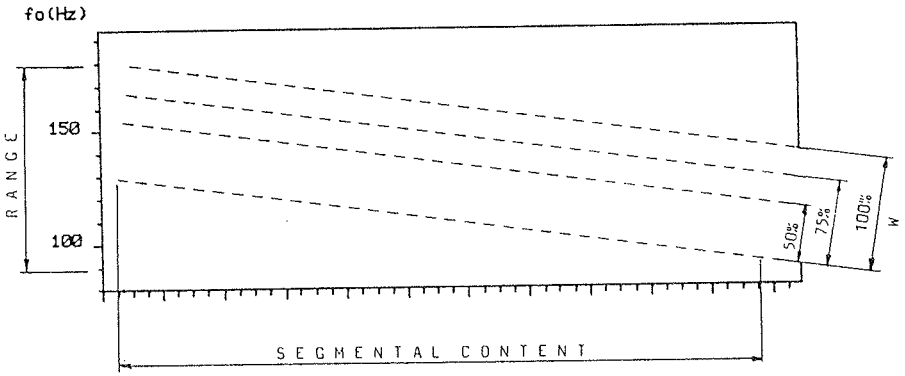


Figure 4. Phonological grid used for synthesizing F_0. The F_0 range extended between 90 and 180 Hz. The beginning and end points for a given sentence were set at 130 Hz and 90 Hz, respectively. According to Figure 1, the first focussed constituent receives a level of prominence equal to 'W', the second, a level of prominence equal to .75W, and the third, a prominence level equal to .50W.

An informal analysis by synthesis experiment and a test of the derived rules on a fragment of text resulted in the algorithm for pitch generation presented in Figure 5. An example of how an intonation contour would be generated is presented in Figure 6 for the sentence Nine million is still owing me, where nine and owing are focussed according to the model in Figure 1.

A few comments on the prosody model in Figure 5 will be made here:

In point 2 in the flow chart is given a rule for assigning a level of prominence corresponding to .25W to the heads of syntactic phrases. This rule was arrived at on the basis of our analysis-by-synthesis experiment. The need for its existence in the rule system would, furthermore, seem to support Beckman and Pierrehumbert's (1986) speculation on the existence of some kind of left-dominant 'accentual phrase' as a

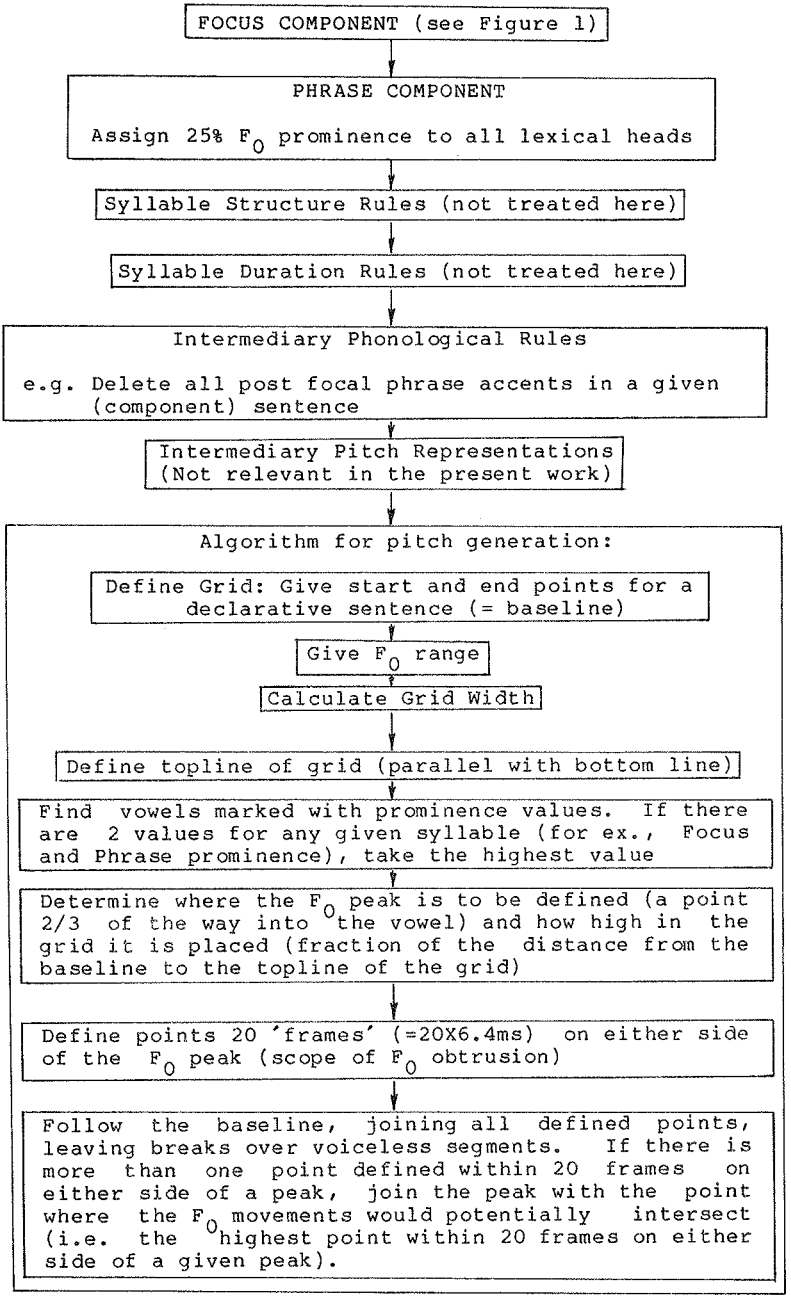


Figure 5. Flow-chart for English sentence prosody systematizing rules for synthesizing F_0 contours.

Derivation of sentence in Figure 6 below following prosody model in Figure 5.

```
    F=W                    F=.75W
1)  Nine million is still owing me

        P=.25W              P=.25W
    F=W                    F=.75W
2)  Nine million is still owing me
5)  Not applicable
7)  Baseline defined in Figure 6a
8)  F0 range defined in Figure 6a
9)  Grid width (W) calculated in Figure 6a
10) Topline of grid defined in Figure 6a

        P=.25W
    F=W                    F=.75W
11) Nine million is still owing me
12) Define F0 peaks in grid (X's in Figure 6a)
13) Define scope of F0 obtrusion (*'s in Figure 6a)
14) Generate F0 contour (Figure 6b)
```

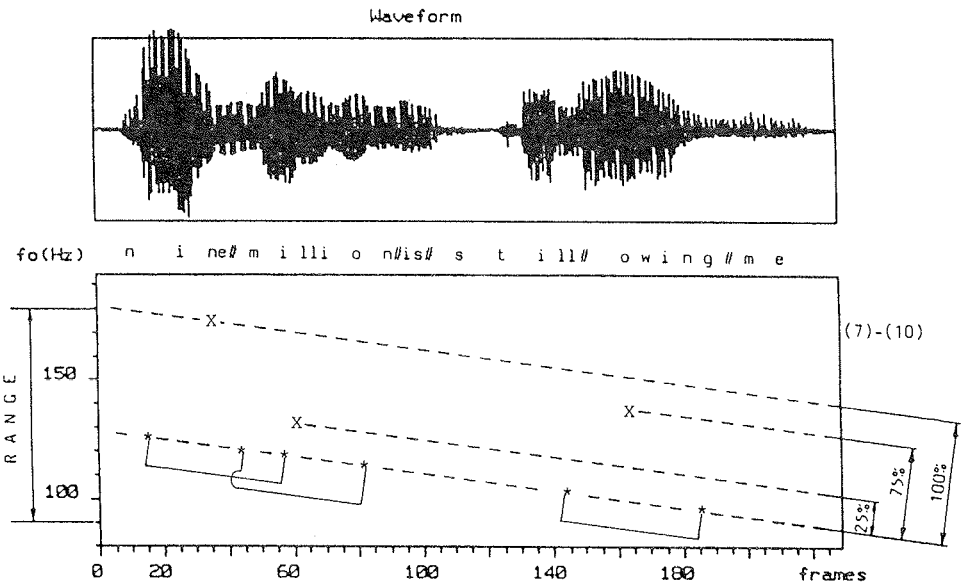


Figure 6a. Partial derivation of F_0 curve (after point 13 in flow-diagram in Figure 5). See above for a clarification of the figure. Derivation continued in Figure 6b.

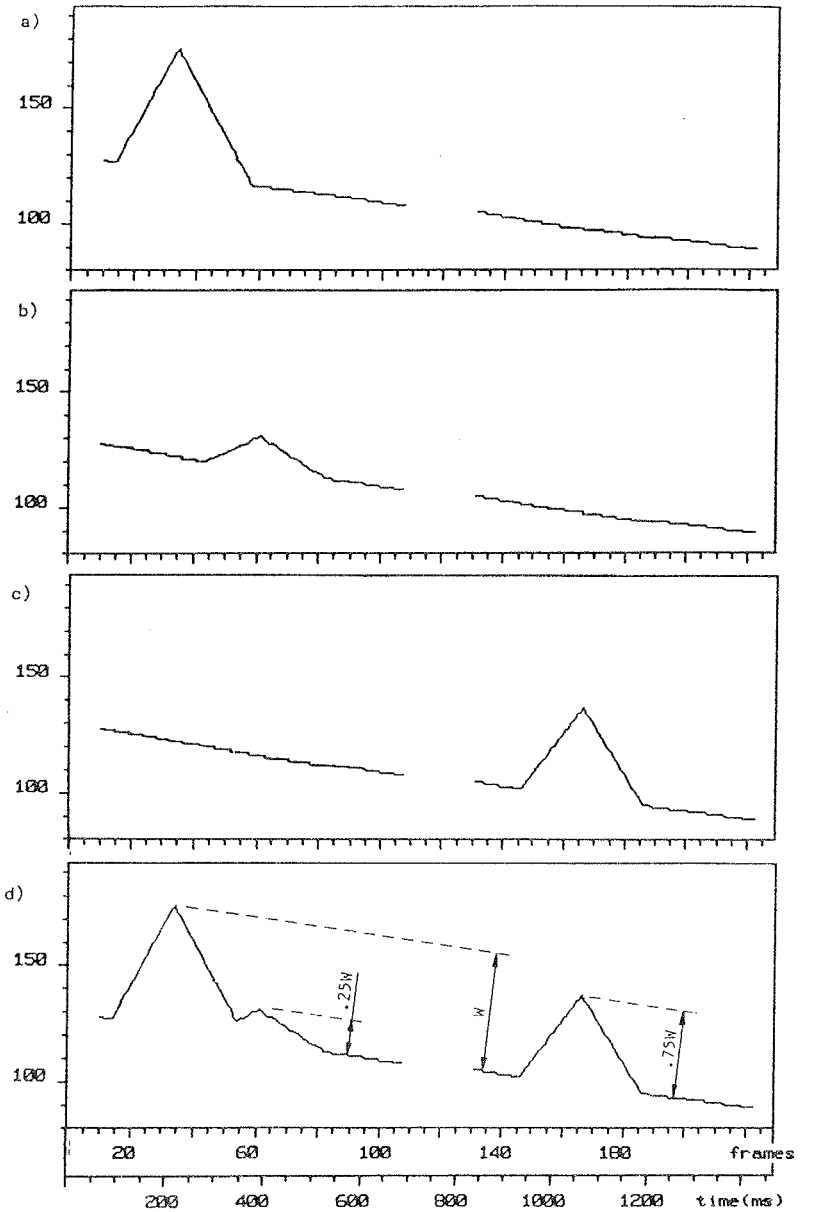


Figure 6b. Potential stages in the synthesis of the F_0 curve of sentence in Figure 6a. Notice that the first two F_0 obtrusions overlap. The final output in (d) is obtained by connecting the highest points in the intermediary curves (a-c).

unit of English prosody. That is to say, there is perhaps some structural reason why it is possible to have a flat F_0 curve after an early focal F_0 movement but not when the focus comes late in a sentence. It is possible that there are well-formedness constraints on the form of an intonational contour like there are on the prosodic patterns of words. Just as there is a constraint that says that a word cannot begin with two light syllables in English, it is possible that there is an analogous restriction on the sentence level that prohibits a sentence from beginning with two phrases without any F_0 prominence. Further research is needed to clarify this point.

The definition of a sloping baseline at Point 7 assumes that 'declination' is an important parameter in the description of English sentence intonation. This issue has been the topic of much debate and has never really been resolved (see Cohen et al. 1982, Bruce 1984, Ladd 1984). Our data would, however, seem to support the observations of Maeda (1976) who found "a constant amount of declination for each speaker, and hence a slower rate of decline in longer utterances" (cited from Ladd 1984:57).

At point 11 in the flow diagram is a rule that chooses the highest prominence value for synthesis in the case where both a focal and a phrasal prominence had been assigned to one and the same word. Although the rule effects the correct output, it is possible, that from a production point of view, the assignment of prominence should be envisaged in another way, viz. that the assignment of a prominence value blocks the assignment of any further prominence by another rule.

8. Implications for phonological theory: the 'phonological phrase'

Paper III discusses the notion of 'phonological phrase' within some recent phonological analyses (Selkirk 1980,1984; Nespor & Vogel 1982,1983) in light of a model such as ours that includes a focus assigning component in the grammar. The notion of a syntactically defined 'phonological phrase' has been used in the literature to explain the phenomena of e.g. the 'Rhythm Rule' in English and 'Raddoppiamento Sintattico' in Italian.

8.1. Rhythm Rule

The so-called 'Rhythm Rule' is intended to account for the fact that when a phrase is uttered out of context, for example, the well-known thirteen men case, one gets the impression that in the word thirteen, 'stress' shifts from the final lexically stressed syllable -teen to the syllable thir- when followed by another word with lexical stress on the first syllable, e.g. men. That is to say, the potential 'clash' that arises when two stressed syllables lie next to each other is avoided by moving the 'stress' to the left.

Our view of the phenomenon, however, differs from this stress movement analysis. We maintain that what causes the impressionistic effect of a rhythm rule in English is simply the nonrealization of a word stress as a F_0 movement on -teen. In the phrase thirteen men, it is the word men whose stress would be realized as F_0 according to our focus assigning algorithm since men is the head of the constituent. The impression of stress on the first syllable is an automatic consequence of the fact that it is a 'heavy' syllable (has a

branching rhyme). As a matter of fact, listeners have a tendency to associate an equal amount of 'stress' with all the heavy syllables of unfocussed 'rhythm rule' words such as Tennessee or kangaroo (see Cooper & Eady 1986), so that the postulation of the addition of rhythmical prominence on the initial syllable of these words in nonfocal position seems to be unwarranted. We did this in Paper III (pg.108), following Schane (1979). We now feel that this 'Initial Constraint' discussed there is to be better regarded as a phonotactic constraint on English word structure and not as a process in the derivation of sentence intonation.

8.2. Raddoppiamento sintattico (RS)

Another process that has been considered to show the existence of the 'phonological phrase' is the sandhi process of Raddoppiamento Sintattico in Italian. This process has the effect of lengthening the initial consonant of $word_2$ in a sequence $word_1$ $word_2$ if the final vowel of $word_1$ is stressed and the affected consonant is followed by (Glide/Liquid) V. The process is considered by Nespor and Vogel to take place in a syntactically defined phonological phrase, Φ , which is determined as in (4):

(4) a. φ construction (φ = phonological phrase)

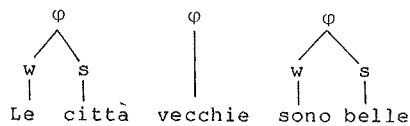
Join into φ any lexical head (X) with all items on its nonrecursive side within the maximal projection and with any other nonlexical items on the same side (e.g. prepositions, complementizers, conjunctions, copulas...)

b. φ constituency

φ branches in the same way as the syntactic tree

So, for example, the sentence in (5) would be divided up into phonological phrases as indicated in the associated tree:

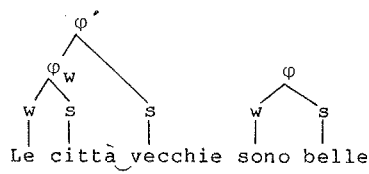
(5)



'Old cities are beautiful'

In this case, RS would not apply since città and vecchie belong to different phonological phrases. However, a rule of optional restructuring has the effect of creating a structure where città and vecchie fall within a single phonological phrase as in (6):

(6)



where application of RS is indicated by ' ‿ '

Just what the conditions are under which restructuring takes place are not given. It would seem to us, however, that the application vs. nonapplication of RS is intimately tied to the prosodic structure of a given utterance. That is to say, we would hypothesize that RS can apply when the second of the two words is more prosodically prominent than the first. The structure in (6) would, we claim, be associated with a context where focus is realized on vecchie, i.e., this wuld be the 'normal' case. The nonoccurrence of RS, as in (5), would be associated with a contrastive context in which focus was realized on città, as e.g. in sentence B in the text fragment in (7):

(7) A: Did you say old CITADELS are beautiful?
B: No, I said old CITIES are beautiful.

We feel that the process of RS culd be better stated in terms of prominence relations than in terms of some rather ad hoc phonological phrase. RS takes place when $word_2$ has more (focal or phrasal) prominence than $word_1$.

8.3. Monosyllabic Destressing

Monosyllabic Destressing (as discussed by Selkirk 1972,1984) is a third process that appears to be conditioned by prominence relations within the sentence. That is to say, we feel that the domain for this process is a given grammatical function and that the conditioning environment is that the monosyllable be followed by a more prominent word in that constituent. This would account for the weakening of in in sentence (8a) as well as the absence of destressing in (8b), where the monosyllable comes at the end of the constituent

realizing the predicate, and (8c) (here the monosyllable is focussed, i.e. has more prominence than the material that precedes):

(8) a. He boxed in ([ɐn]) the crowd

b. What were you thinking of ([ɔv]) last night?

c. I don't recall the title, but I should ([ʃud]) remember it

9. Contrastive Prominence

Paper V presents some preliminary research on 'contrastive prominence' conditioned by syntactic parallelism. Intonation patterns associated with contrastive prominence cannot be assigned and synthesized by our prosodic model as it stands. The goal of Paper V is, therefore, to propose how these patterns could be accounted for. Three different rules, which are assumed to apply before the focus assigning algorithm in Figure 1, were formulated to account for the F_0 patterns in sentences such as those in (9):

(9) a. $\text{Peter}_{(i)\text{Subj}}$ $\text{kicked}_{(j)\text{Pred}}$ $\text{Ulrika}_{(k)\text{Pred.Comp.}}$

and then $\text{SHE}_{(k)\text{Subj}}$ $\text{kicked}_{(j)\text{Pred}}$ $\text{HIM}_{(i)\text{Pred.Comp.}}$

b. I looked ON the table, UNDER it and AROUND it.

c. I didn't say UNarm; I said REarm.

In (9a), the parallelism involves pairs of sentences (potentially n-tuples of sentences) where one particular grammatical function is associated with the same referent in both sentences. The words associated with the referents in the

second sentence which do not fill the same grammatical function as they did in the first are assigned contrastive prominence.

In sentence (9b), the syntactic parallelism involves iterative enumeration of a given syntactic phrase. The words realizing the nonidentical referents of the phrasal categories are assigned contrastive prominence.

Finally, (9c) presents a case of syntactic parallelism involving a given phrase which is associated with opposing values of polarity but which fills the same grammatical function in the two sentences. Here, contrastive prominence is also assigned to the words realizing the nonidentical referents within the phrase.

The analysis of the parallel sentence structures conditioning contrastive prominence is also used to shed some light on the problem of noncontraction of the copula in English in sentences such as (10):

(10) Willie {is / 's} as English as Malin {is / *'s} Norwegian

It is suggested that the impossibility of contracting the copula after Malin in (10) can be explained if one assumes that a prosodic boundary is inserted after such words contrasted according to the rules referring to syntactic parallelism ((10) being analogous to (9a)). The boundary can be assumed to block cliticization of the copula to the preceding word.

The different cases of syntactic parallelism exemplified in (9) are also realized by different F_0 patterns and these are also taken up and discussed in Paper V. Preliminary proposals for synthesis of the contrastive F_0 contours are presented as well.

10. Conclusion

The purpose of this study has been to obtain a better understanding of English declarative sentence intonation by examining it in a discourse context. A rule system, geared to a text-to-speech context, has been developed which assigns sentence prominences related to information focus. The realization of these prominences with respect to the phonetic parameter of pitch (F_0) has also been described with a pitch generating algorithm.

The dynamics of information focus in English have been shown to be accountable for by a hierarchy of grammatical functions interacting with a contextual notion of coreference. The hierarchy is superior to a linearly-based 'sentence stress' rule like the Nuclear Stress Rule since it allows for a more general statement of the dynamics of sentence prominence, e.g. it accounts for the placement of non-final as well as final sentence prominence. The notion of 'normal sentence stress' is thus irrelevant to such an approach to sentence prominence. The hierarchy, in fact, constitutes a mechanism for accounting for what have been termed 'default accents' in accounts of sentence prominence assuming a notion of 'normal sentence stress.' Projection of focal prominence is effected by a subroutine that assigns prominence either to the head or to the modifier of the constituent realizing a particular grammatical function depending on the coreferential status of the associated lexical items. The interaction of the grammatical functions with coreferential relations allows the model to apply to non-initial discourse sentences and assign prominence in a systematic way to new information. The issue of 'deaccenting' does not arise, furthermore, since the model assumes that F_0

prominence is not related to word-level information but rather that it is only assigned on the sentence level in e.g. focal contexts. Furthermore, cyclical application of the focus assigning algorithm within a headed construction as well as iterative application within a particular grammatical function allow the model to assign more than one focal prominence within a given constituent. This kind of rule application avoids the necessity of special readjustment rules flattening out constituent trees to limit the number of potential degrees of prominence allowable in a given sentence.

The model assumes three degrees of prominence associated with the three grammatical functions (Subject, Predicate Complement, Predicate). These have further been found to be associable with the parameter of pitch (F_0). The degrees of prominence can be associated with varying heights of F_0 peaks with respect to the baseline of a phonological 'grid' describing a speaker's F_0 range and declination constants. An analysis-by-synthesis test of the model later showed the need for a further degree of prominence on prefocal lexical heads. Thus four degrees of declarative sentence prominence have been postulated. A pitch generating algorithm was subsequently developed and applied to a fragment of discourse.

The notion of 'contrastive prominence' was also discussed and the various instances were unified and analyzed under the notion of 'syntactic parallelism'. Preliminary rules were proposed for the assignment as well as for the synthesis of contrastive prominence. The rules for assigning contrastive prominence are assumed to apply before the general focus assigning algorithm.

The problem of noncontraction of the copula in English was

also related to the issue of contrastive prominence and explained by the presence of a prosodic boundary before the copula which functions to block encliticization.

A number of other phonological issues were also taken up and discussed in light of a grammar which includes a focus assigning component. The status of the notion of 'phonological phrase' was examined in the light of a number of processes which have been claimed to warrant its existence.

The so-called 'Rhythm Rule' in English is one of these. The phenomenon was reanalyzed and explained as the nonrealization of a word stress as focal prominence. The impressionistic shift of stress is maintained to result from the heavy syllable structure of the affected word.

The sandhi process of Raddoppiamento Sintattico in Italian is another phenomenon that would appear to be influenced by prominence relations on the sentence level, i.e. the process is claimed to be restricted to taking place in an environment where the two words involved in the process are related in such a way that $word_2$ has more sentential prominence than $word_1$.

A third process which has been used to motivate the existence of a phonological phrase is Monosyllabic Destressing. This process, however, is also explainable by relating it to sentence prominence relations. Monosyllabic Destressing can be simply accounted for by situating it in a sentence-level perspective; it is seen as taking place in a constituent where the monosyllable is followed by a word receiving more sentential prominence.

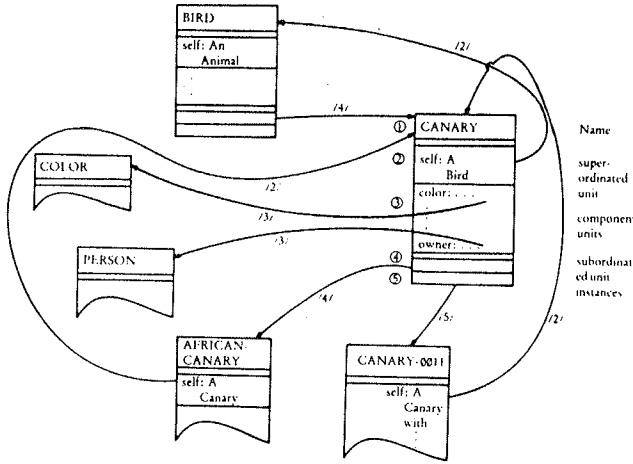
Footnotes

1. The model described here makes the assumption that it is to some extent possible to recognize identity relationships between text referents (see e.g. Hirst 1983). This is so far an area that has received relatively little priority in text-to-speech programs due to practical reasons. We see it, however, as theoretically stimulating to include this thorny area in our model.

Existing text-to-speech systems generally rely on overtly marking words that do not receive focal prominence due to coreference, e.g. Eady & Dickson 1987. Monaghan 1987 and Ladd & Monaghan 1987, however, have attempted to handle certain cases of 'deaccenting' by using lexical and syntactic deaccenting diacritics placed, e.g. on "semantically empty" NP's such as street and building.

Discourse-oriented anaphora theories such as that of Sidner discussed in note 2 use search methods for matching referents based on hierarchies of 'frames' of reference (see Hirst 1983 for a review). Frames can be thought of as connected structural units which play an important role in recognition and reasoning (Metzing 1981). For example, the coreferential relationship between canary and bird in the text fragment below would be resolved by means of a frame representation such as that in Figure 7 (reproduced from Metzing 1981:336):

My best friend gave me a canary$_i$ for my birthday, but my husband forced me to give the bird$_i$ away. He was allergic to it$_i$.



There are several structural elements of a frame unit which are linked in special ways: There is the name of a frame unit (①); this unit is pointed to from a more general unit (/4/) as well as from a more special unit (/2/) and from an instantiation (/2/); this unit is described in terms of another unit (prototype unit) (②) whose properties are inherited (and pointed to by /2/). Additional component properties of an object may be described in terms of other units (③), pointed to by /3/. And finally, there are pointers to subordinated units (④, /4/) and to instances (⑤, /5/).

Figure 7. Frame representation reproduced from Metzing (1981:336)

2. It should be pointed out here that our hierarchy for assigning focal prominence receives support from independent work within artificial intelligence on referent resolution (anaphor comprehension). In an influential program developed by Sidner (1983), the PAL (Personal Assistent Language Understanding Program) system, it has been proposed that 'focus' can be used in selecting and ordering the set of potential antecedents for an anaphoric element. By 'focus', Sidner means 'discourse topic', "the something which the communication is about". According to Sidner (1983:274), "the focus and the assumed shared knowledge can be used as one of

the chief constraints on the choice of the co-specification of anaphoric expressions". The algorithm that Sidner presents for choosing an initial discourse 'focus', although expressed in terms of thematic positions, in essence boils down to the hierarchy we have suggested, i.e. Predicate Complement > Subject > Predicate. We reproduce Sidner's (1983:287) algorithm here for the interested reader:

"Expected Focus Algorithm
Choose an expected focus as :
(1) The subject of the sentence if the sentence is an is-a or a there-insertion sentence
This step presumes information from a parse tree about what the subject, and the verb are and about whether the sentence is there-insertion.
(2) The first member of the default expected focus list (DEF list), computed from the thematic relations of the verb is as follows:
-order the set of phrases in the sentence using the following preference schema:
-theme [i.e. usually Object (MH)] unless the theme is a verb complement, in which case the theme from the complement is used
-all other thematic positions with the agent last
-the verb phrase
This step requires a list of the surface order of the noun phrases and a data structure which indicates which noun phrases fill which thematic slots in the verb."

The algorithm applies to the first sentence in a discourse and predicts an anaphor's cospecifier and then an inferring process confirms the prediction. For example, in the text in (11), the algorithm chooses the object, bear over the verb phrase as a potential cospecifier of it in the noninitial sentences:

(11) i. Mike captured a bear
ii. Everyone said it made a lot of noise
iii. but I was asleep and didn't hear it

Although Sidner does not take up the issue of prosody since she deals with written texts, we feel that our proposed

hierarchy finds support in the fact that it can also be used in referent resolution programs. In fact, one could postulate that the various degrees of prosodic prominence given to the different constituents provide the listener with further cues for keeping track of discourse referents.

References

Allen, J., Hunnicutt, M.S., and Klatt, D. 1987. From text to speech. The MITalk system. Cambridge: Cambridge University Press.

Bannert, R. 1986. Independence and interdependence of prosodic features. Working papers (Dept. of Ling., Univ. of Lund) 29: 31-60.

Beckman, M. and Pierrehumbert, J. 1986. Intonational structure in Japanese and English. Phonology yearbook 3: 255-300.

Berman, A. and Szamosi, M. 1972. Observations on sentential stress. Language 48: 304-25.

Bing, J. 1980. Aspects of English prosody. Bloomington; Indiana Univ. Linguistics Club.

Bolinger, D. 1972. Accent is predictable (if you're a mind reader). Language 48: 633-44.

Bresnan, J. 1971. Sentence stress and syntactic transformations. Language 47: 257-81.

_____ 1972. Stress and syntax: a reply. Language 48: 326-42.

Bruce, G. 1977. Swedish word accents in sentence perspective. Lund: Gleerups.

_____, 1984. Aspects of F_0 declination in Swedish. Working Papers (Dept. of Ling., Univ. of Lund) 27: 51-64.

Carlson, R. and Granström, B. 1976. A text-to-speech system based entirely on rules. Proceedings of the Int. Conf. on acoustics, speech and signal processing, 686-8. New York: IEEE. IEEE Catalog No. 76-CH-1067-8 ASSP.

Chomsky, N. 1971. Deep structure, surface structure and semantic interpretation. In: D. Steinberg and L. Jakobovitz, Semantics: an interdisciplinary reader in philosophy, linguistics and psychology, 183-216. Cambridge: Cambridge University Press.

_____ and Halle, M. 1968. The sound pattern of English. New York: Harper & Row.

Cohen, A., Collier, R., and t'Hart, J. 1982. Declination: construct or intrinsic feature of speech pitch? Phonetica 39: 254-73.

Contreras, H. 1976. A theory of word order with special reference to Spanish. Amsterdam: North Holland.

Cooper, W. and Eady, S. 1986. Metrical phonology in speech production. Journal of memory and language 25: 369-84.

Cullicover, P. and Rochemont, M. 1983. Stress and focus in English. Language 59: 123-65.

Eady, S. and Dickson, B. 1987. Speech synthesis of sentence focus in English declaratives. In Proceedings XIth Int. Cong. Phonetic Sciences, Aug. 1-7, 1987. Tallinn, Estonia. Vol. 3, 262-5. Academy of Sciences of the Estonian S.S.R.

Erteschik-Shir, N. and Lappin, S. 1983. Under stress: a functional explanation of English sentence stress. J. Linguistics 19: 419-53.

Firbas, J. 1979. Post-intonation-centre prosodic shade in the modern English clause. In Greenbaum, S., Leech, G., and Svartvik, J. (eds.), Studies in English linguistics for Randolph Quirk, 125-33. London:Longman.

Gårding, E. 1983. A generative model of intonation. In: A. Cutler and D. Ladd (eds.), Prosody: models and measurements, 11-25. Berlin:Springer.

Gunter, R. 1966. On the placement of accent in dialogue: a feature of context grammar. J. of linguistics 2:159-79.

Gussenhoven, C. 1983. Focus, mode and the nucleus. J. of linguistics 19: 377-417.

Halliday, M. 1967. Notes on transitivity and theme in English. Part 2. J. of linguistics 3: 177-274.

Hirst, G. 1983. Anaphora in natural language understanding: a survey. Berlin: Springer.

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

Ladd, D.R. 1978. The structure of intonational meaning. Bloomington: Indiana Univ. Press.

_____. 1984. Declination: a review and some hypotheses. Phonology yearbook 1:53-74.

_____. 1987. A model of intonational phonology for use in speech synthesis by rule. In Laver & Jack, 21-4.

_____ and Monaghan, A. 1987. Modelling rhythmic and syntactic effects on accent in long noun phrases. In Laver & Jack, 29-32.

Laver, J. and Jack, M. (eds.) 1987. Proceedings: European conference on speech technology, Edinburgh. Edinburgh: CEP Consultants.

Maeda, S. 1976. A characterization of American English intonation. PhD Diss., MIT.

Metzing, D. 1981. Frame representation and lexical semantics. In H.-J. Eikmeyer and H. Rieser, Words, worlds and contexts, 320-42. Berlin: Walter de Gruyter.

Monaghan, A. 1987. A system for left-to-right intonation specification from text. In Laver and Jack (eds.), 25-8.

Nespor, M. and Vogel, I. 1982. Prosodic domains of external sandhi rules. In H. Van der Hulst and N. Smith (eds.), The structure of phonological representations, Part I, 225-55. Dordrecht:Foris.

_____. 1983. Prosodic structure above the word. In A. Cutler and D. Ladd (eds.), Prosody: models and measurements, 123 - 40. Berlin: Springer.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. 1972. A grammar of contemporary English. London: Longman.

Schane, S. 1979. The rhythmical nature of English word stress. Language 55: 559-602.

Schmerling, S. 1976. Aspects of English sentence stress. Austin: University of Texas Press.

Selkirk, E. 1972. The phrase phonology of English and French. Ph.D. Diss., M.I.T. Reproduced by Indiana Univ. Linguistics Club, 1981.

_____. 1980. On prosodic structure and its relation to syntactic structure. Indiana Univ. Linguistics Club.

_____. 1984. Phonology and syntax: the relation between sound and structure. Cambridge, Mass.: MIT Press.

Sidner, C. 1983. Focusing in the comprehension of definite anaphora. In A. Brady and R. Berwich (eds.), Computational models of discourse, 267-330. Cambridge, Mass.: MIT Press.

Sigurd, B. 1987a. Referent grammar (RG) in computer comprehension, generation and translation of text (SWETRA). Working papers (Dept. of Ling., Univ. of Lund) 31: 137-67.

_____. 1987b. Referent grammar (RG)--a generalized phrase-structure grammar with built-in referents. Studia linguistica 41:2 (to appear).

Touati, P. 1987. Structures prosodiques du suédois et du français. Profils temporels et configurations tonales. Lund: Lund University Press.

Vanderslice, R. and Ladefoged, P. 1972. Binary suprasegmental features and transformational word-accentuation rules. Language 48: 819-38.

Acknowledgements

This thesis, like all others, was not produced in a vacuum. A considerable number of people have helped and influenced me during the course of its development and it gives me great pleasure to acknowledge them here.

First of all, I would like to express my gratitude to my thesis director, Thore Pettersson, for his interest, enthusiasm, and guidance during my thesis research. His office door has always been open for discussion and debate!

I am also indebted to Bengt Sigurd for helpful suggestions and constructive criticism as well as for insightful discussions on Referent Grammar!

Gösta Bruce and Eva Gårding are acknowledged for their stimulating 'Friday Seminars' in phonetics where I became fascinated with the Lund model of prosody. My thesis topic was a direct result of the input I received there. I would also like to thank the 'guys' from the phonetics lab for helping me in countless ways, and in particular, David House, for letting me analyze his F_0 contours, and Paul Touati, for inspiration. Lars Eriksson and Gustav Jönsson are thanked for technical assistance during recording and analysis sessions.

My friend and colleague, Jan-Olof Svantesson is warmly thanked for valuable comments and discussion on many occasions.

I would also like to express my appreciation to the Swedish state for the generous financial support they have provided over the last four years as well as for a grant making the publishing of this thesis possible.

And last but by no means least, Olle....THANK GOODNESS FOR YOU!

I

Merle A. Horne

ENGLISH SENTENCE STRESS, GRAMMATICAL FUNCTIONS AND CONTEXTUAL COREFERENCE

1.0 Introduction*

The goal of this paper is to provide a better understanding of the dynamics of English sentence stress by examining it within a discourse grammar framework. It will be maintained that the position of sentence stress is governed by two interacting factors: a hierarchy of grammatical functions and the coreference status of the lexical items realizing these functions with respect to previous parts of the discourse. This approach differs from previous generative treatments, where sentence stress is described solely with reference to the syntactic constituent structure (Chomsky & Halle 1968; Bresnan 1971, 1972). Conceptually, the present model is influenced by the Praguean Functionalist Sentence Perspective, an approach which has not been without impact on American linguists (Chafe 1974:119–29; Liberman 1978:167).

We will first review some previous attempts that have been made to deal with English sentence stress in the phonological literature before presenting our own approach. Most analyses within the generative framework have been based on the assumption that it is possible to predict the relative degrees of suprasegmental prominence within a sentence using only information from the labelled bracketed surface structure.

1.1 Chomsky & Halle (1968)

In Chomsky & Halle (1968), for example, it is maintained that sentence stress is assigned by the Compound Rule and the Nuclear Stress Rule, which can be formalized as in (1) (p. 18):

* Thanks are due to Gösta Bruce, Thore Pettersson, Bengt Sigurd, and Paul Touati for useful comments on an earlier version of this paper and to David House for assisting me with phonetic analyses of the data.

$$\text{(1)}\quad \begin{bmatrix}1\ \text{stress}\\ V\end{bmatrix} \rightarrow [1\ \text{stress}] / \left\{ \begin{matrix} \left[\_\_\_ \cdots \overset{1}{V} \cdots \right]_{NAV} \\ \left[\overset{1}{V} \cdots \_\_\_ \cdots \right] \end{matrix} \right\} \quad \begin{matrix}\text{a) Compound Rule}\\ \text{b) Nuclear Stress Rule}\end{matrix}$$

The effect of rule (1) is to place primary stress on the first stressed vowel of a compound (e.g. *$\overset{1}{\text{black}}\overset{3}{\text{board}}$*) and on the last stressed vowel of a phrase (e.g. *$\overset{2}{\text{black}}\ \overset{1}{\text{board}}$*), each word having previously been assigned stress by rules of word phonology. Furthermore, by a general convention of stress assignment, each time an element receives primary stress, all other stresses within the domain of application are reduced by one. Rule (1), along with all other stress rules, have been assumed to apply cyclically to surface structures. As an example of how the two ordered subrules of (1) interact to produce a prominence contour, the sentence *John's blackboard eraser was stolen* is analyzed in (2) (see Chomsky & Halle 1968:22):

(2) $_S[_{NP}[_D[\text{John's}_D]_N[_N[_A[\text{black}_A]_N[\text{board}_N]_N]_N[\text{eraser}_N]_N]_{NP}]_{VP}[_V[[\text{was}][\text{stolen}]_V]_{VP}]_S]$

John's	black	board	eraser	stolen	
1	1	1	1	1	
	1	2			(1a)
	1	3	2		(1a)
2	1	4	3		(1b)
3	2	5	4	1	(1b)

One problem associated with the SPE system is that, due to the convention of stress reduction and to the fact that the NSR can apply to its own output, complex sentences will receive prominence contours containing indefinitely weak stress, such as the sentence mentioned by Chomsky & Halle (1968:23): *My friend can't help being shocked at anyone who would fail to consider his sad plight,* where *sad plight* would end up with the contour *$\overset{8}{\text{sad}}\ \overset{1}{\text{plight}}$* after 7 applications of the NSR. Chomsky and Halle (1968:23) note this shortcoming of their analysis and 'make it quite clear that the rules discussed above give accurate results only for very simple constructions'. They suggest, furthermore, that 'the problem of extending this description to a wider class of cases may be nontrivial' (1968:24).

The problem involving multiple degrees of stress was avoided in later phonological descriptions by working on a more abstract level of metrical structure with representations consisting of structures with alternating patterns of S[trong] and W[eak] syllables. The problem of determining sentence stress still remains, however (see below, § 3.3.), since metrical trees are, like Chomsky and Halle's stress contours, based on the syntactic surface structure and the principles embodied in the Compound and Nuclear Stress Rules.[1]

1.2 Bresnan (1971)

In what was the first detailed generative treatment of English sentence stress, Bresnan (1971) pointed out that there was a large class of exceptions to the NSR. For example, under normal conditions, the following sentence-final items do not receive final stress (258):

(3) Anaphoric pronouns:
   (a) Robert discussed it
   (b) *Robert discussed it

(4) Indefinite pronouns:
   (a) The girl picked some
   (b) *The girl picked some

Other anaphoric elements, even when grammatically definite, are not assigned sentence stress:

(5) Mary knows a man$_i$ who excells at judo, and she adores the man$_i$.

As regards these exceptions to the NSR, Bresnan assumes that 'by some means or another, anaphoric and indefinite elements are not assigned primary stress' (1971:258). She does not deal with the derivation of the stress pattern of such sentences, but instead attempts to account for another set of apparent exceptions to the NSR. These involve sentences which can be analyzed as having undergone movement or deletion of NP objects, e.g.:

(6) (a) George has plans to leave
   George has plans[$_S$George leave plans
   (b) George has plans to leave
   George has plans[$_S$George leave

(7) (a) Mary liked the proposal that George left
   Mary liked the proposal [$_S$that George left the proposal

---

[1] '... we simply define a relation on each pair of sister nodes in the syntactic structure, the output depending on certain local properties of that structure. To be specific, this theory's version of the NSR and CSR can be stated as follows:

(8) in a configuration [$_C$AB]$_C$:
   (a) NSR: If C is a phrasal category, B is strong
   (b) CSR: If C is a lexical category, B is strong iff it branches.' (Liberman & Prince 1977:257.)

(b) Mary like the proposal that George leave
Mary liked the proposal [$_S$that George leave

(8) (a) John asked what Helen had written
John asked [$_{S'}$ COMP$_{+WH}$ [$_S$Helen had written something$_{+wh}$

(b) John asked what books Helen had written
John asked [$_{S'}$ COMP$_{+WH}$ [$_S$Helen had written some books$_{+wh}$

According to Bresnan, all these exceptions to the NSR are 'predictable without any special modifications in that rule, given one assumption: **the nuclear stress rule is ordered after all the syntactic transformations on each transformational cycle**' (Bresnan 1971:259). She assumes, furthermore (1971:259), that the NSR cycles only on NP and S, not on VP. Consequently, the derivation of the sentence stress patterns in (6)–(8) becomes straightforward. In (6a), for example, the embedded sentence has an object, *plans*, which is assigned primary stress on the lower cycle, with concomittent lowering of stress on *George* and *leave*. On the upper cycle, a syntactic rule applies, deleting *plans* in the embedded S. The NSR subsequently applies on the upper cycle and assigns primary stress to the rightmost [1 stress], which now happens to be the object *plans* in the matrix sentence. In (6b), on the other hand, there is no underlying object that gets deleted or moved from its sentence-final position, with the result that the verb, *leave*, is reassigned [1 stress] on the upper cycle. Derivations for the other examples proceed in an analogous fashion.

1.3 Berman & Szamosi (1972)

Bresnan's proposal was criticized by Berman and Szamosi (1972) on several points. They note, for example, that in many cases Bresnan's cyclical hypothesis makes the wrong predictions. For instance, if fails in cases where, instead of the object the subject is deleted, e.g.:

(9) Let me tell you about something strange that happened

where following Bresnan's proposal, primary stress should fall on the verb, and in

(10) Mary liked the proposal that was made

where Bresnan's cyclical hypothesis would predict

(10′) *Mary liked the proposal that was made
                                    1

Although they do not develop any alternative for describing English sentence stress, Berman and Szamosi propose the following surface generalization which they claim governs its placement to some extent; they note that 'regardless of the "derivational history" of a sentence, if it ends in an NP, this NP receives primary stress' (1972:309). They note, however, that there are a number of exceptions, most notably, anaphoric items, e.g.

(11) The children didn't want to go to bed, so
    (a) John scolded the governess
                        1.
    (b) John scolded the bastards
            1

Moreover, Berman and Szamosi feel that the different stress contours of phrases like *plans to leave* (cf. (6)), corresponding to different meanings 'remain valid counterexamples to a theory of stress-assignment which relies solely on surface structure' (1972:310). Sentences like (12) are also exceptions to Berman and Szamosi's surface structure generalization, as was later pointed out by Bresnan (1972:331):

(12) Peter had clams for dinner
                1

Berman and Szamosi come to the conclusion, therefore, that 'neither the surface level application nor the cyclical application of the NSR works. Can the NSR be made to work at all? Probably not, especially because of the existance of a large number of cases which indicate that primary stress assignment (let alone the entire stress contours) often depends on factors that only marginally, if at all, involve structure' (1972:311–12).

1.4 Bresnan (1972)

In a reply to Berman and Szamosi, Bresnan (1972) admits that not only structural, but also semantic factors are involved in determining the placement of primary stress. She does not abandon the NSR, however, but rather proposes a new rule, that of Topical Stress Assignment, an optional rule which is assumed to apply before the NSR. This rule of Topical Stress Assignment is assumed to assign stress in the exceptional cases to the NSR. For example, Bresnan notes (1972:328) that in simple declarative sentences ending in a predicative, the primary stress is often most naturally placed on the subject:

(13) (a) The sun[1] is shining
(b) A proposal[1] was made
(c) A new book[1] has appeared
(d) An umbrella[1] was lost
(e) Work[1] must be done
(f) Guests[1] arrived

Topical Stress Assignment is assumed to account for the patterning of sentence stress in (13). It is also assumed to handle Berman and Szamosi's counterexamples to the NSR, e.g.

(10) Mary like[d] the[1] proposal that was made (A propósal was made)

(14) What books[1] have arrived? (Some bóoks have arrived)

(15) There is work[1] to be done (Wórk (must) be done)

The inclusion of a rule of topical stress assignment in addition to the NSR makes the grammar of sentence stress that much more complicated. It is all the more dissatisfying due to the fact that no details are presented as to the conditions of application of the Topical Stress Assignment Rule.

2.0 A proposal: Discourse Coreference and Grammatical Functions (DisCoGramFunc)

We think that Berman and Szamosi's idea that it is surface structure that is relevant to determining the place of sentence stress can be made to work if we take an 'enriched' surface structure as input. We are of the opinion that, to the extent that syntactic information is relevant to the assignment of sentence stress, it is grammatical functions, not constituent structure that provides the most insightful basis for understanding the dynamics of sentence stress. The position of sentence stress is, we would maintain, highly predictable on the basis of what we would propose is a natural hierarchy of grammatical functions based on their potential for attracting stress and the coreferential status of the lexical items realizing these functions with respect to preceding parts of the discourse.[2] In other words, we claim that the

[2] For the present, we will not attempt to formulate specific rules determining contextual coreference but rather follow Firbas' (1979:127) guide rule:

position of greatest prosodic prominence in a given English sentence within a discourse is based not on the NSR, but rather on a hierarchy of grammatical functions of the following type:

(16) Predicate Complement > Subject > Verb

That is to say, in a given sentence, the predicate complement[3] is most likely to receive stress if it is contextually noncoreferent; if it is contextually coreferent, then sentence stress will fall on the subject if *it* is contextually noncoreferent. If the subject is also contextually coreferent, sentence stress with fall on the verb.

This hierarchy is based on semantic notions. The fact that subjects and objects tend to receive more prominence than verbs has been observed by a number of scholars.[4] For example, Schmerling (1976:82) formulates the idea

---

'An element is context dependent [i.e. contextually coreferent M.H.] if it conveys information derivable from the preceding verbal context and/or if it refers to a perfectly obvious item belonging to the situational context of immediate experience shared by the speaker and listener, context dependence or independence ultimately depending on the communicative purpose entertained by the speaker at the point that, at the moment of utterance, has been reached in the development of discourse. It should be evident that in regard to context dependence, context is here understood in the narrowest way.'

[3] By 'predicate complement' we mean, following Lyons (1969:345), a 'word or phrase (other than the verb itself) which is an obligatory constituent of the predicate: for instance, the object of a transitive verb ..., nominal or adjectival expressions which combine with the 'copula' in such sentences as *Mary is a beautiful girl* and *Mary is beautiful*'..., and such 'adverbial' expressions as *in Central Park* or *on Sunday* in sentences like *The parade was in Central Park* or *The demonstration was on Sunday*'. It will also be assumed that adverbials having selectional restrictions with the lexical verb are included under this category, for example, *go*: Locative adverbial, *behave*: Manner adverbial (see Allerton and Cruttenden 1978, Dik 1978, and Quirk *et al.* 1972). Further research is required as regards specification of these restrictions for individual verbs.

[4] It has been noted that, as regards child language, subjects and objects are acquired before verbs. It is interesting to speculate that this may be related to their greater likelihood of receiving prosodic prominence in discourse. McNeill (1970:66) notes, for example, that 'Greenfield observed the appearance of subjects and objects towards the end of the holophrastic period, but no verbs'. Moreover, in two-word sentences, the verb is often omitted, whereas the subject and/or object are always present. This is explained by McNeill (ibid.) as being due to the relational character of the verb:

'... all possible two-word combinations occur in child speech-subject-object, verb-object and subject-verb. However, these combinations occur with unequal frequency in Bloom's records, declining in the order given ... The most general interpretation of verbless sentences is to regard verbs as being words that stand for grammatical relations, and that, like prepositions, they tend to appear only after combinatorial patterns with non-relational words are established. Such an interpretation is consistent with the progress in child grammar, from the

in her Principle II: 'The verb receives lower stress than the subject and direct object, if there is one; in other words, predicates receive lower stress than their arguments, irrespective of their linear position in surface structure'. A similar observation is made by Firbas (1979:129):

> '... in regard to a context independent object, subject complement, and adverbial functioning as a specification, not as a setting, the verb will as a rule not exceed any of these elements in CD [CD = *communicative dynamism,* a term within Praguean Functionalist Sentence Perspective which refers to the relative extent to which a given element contributes towards the amplification (development) of the discourse]. In the absence of these elements, it will not even exceed a context independent subject in CD. This is because in the development of discourse the primary function of the verb is an introductory one. It consists in introducing into the discourse notions conveyed by the context independent elements. This explains why the verb comparatively rarely comes to carry the highest degree of communicative dynamism and to function as the IC [IC = intonation center]'.

The fact that the predicate complement precedes the subject in the hierarchy can be related to the fact that in English, the rheme (or new information) tends to be placed at the end of the sentence, the normal position of the predicate complement, whereas subject position tends to be reserved for the theme (or old information).

The hierarchy in (16) should, in fact, be further specified in order to account for the internal structure of a given phrase, since it is not always the case that it is automatically the *head* that receives the stress. In cases where it is only the head that is contextually coreferent and not the whole phrase, it is the *modifier* that will be assigned sentence stress, as in (17b):

(17) (a) My mother gave me a [yellow [**dress**$_i$]]$_j$
(b) I really wanted a [**blue** [one$_i$]]$_k$

Compare this with (18b) where the whole object phrase is coreferential with *yellow dress,* thus causing sentence stress to be placed elsewhere:

(18) (a) My mother gave me a [yellow [**dress**$_i$]]$_j$
(b) I'm going to **wear** [the gorgeous [thing$_i$]]$_j$ today.

---

holophrastic period on to express grammatical relations first through combinations and last through special words.'

McNeill does not take up prosodic prominence which certainly constitutes a factor making a particular grammatical function within a sentence more perceptually salient for the child and thus perhaps easier to learn.

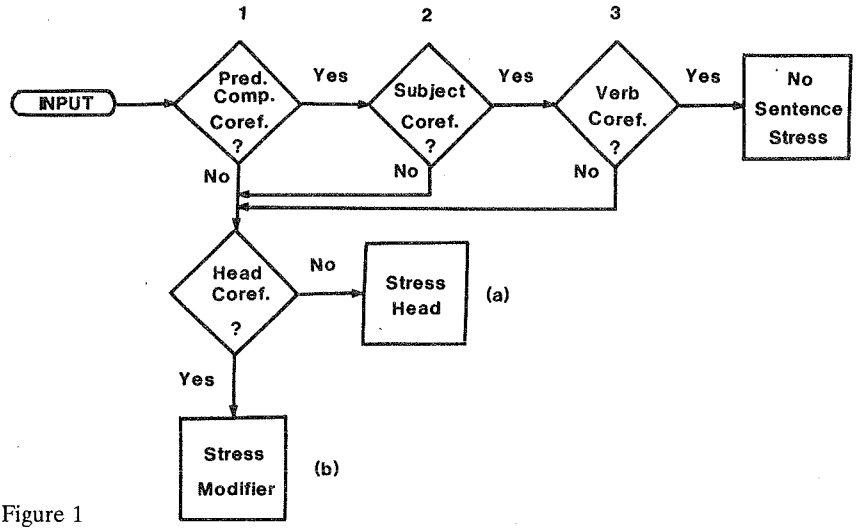


Figure 1

With these facts in mind, the process of assigning sentence stress can then be formally represented by means of a flowchart such as that in Figure 1.

Figure 1 is to be interpreted as follows:

To assign sentence stress in a given piece of dialogue, proceed in the following way: (1) check the Predicate Complement if there is one; if the whole phrase is noncoreferential with respect to the preceding context, assign the Head sentence stress if it is likewise noncoreferential (1a). If the head *is* coreferential, however, assign stress to the Modifier (1b). If the whole Predicate Complement is contextually coreferent on the other hand, go to (2), Subject. The same procedure is repeated here as in the case of the Predicate Complement. Should the procedure be carried to the Verb position and a 'Yes' response be obtained even here as regards the coreferential status of the whole verb phrase, then the sentence is marked 'No Sentence Stress'. Such a sentence is in some respect 'devient' as regards 'normal' sentence stress dynamics, e.g. a sentence which provides no 'new' information, such as (19b):

(19) (a) Kids_i like_j icecream_k
(b) They_i like_j it_k

2.1 Comparison with NSR

It is interesting to note that the characteristic 231 stress pattern derived by the NSR for simple SVO sentences reflects the ordering in (16), e.g.:

(20) Rats like cheese
2 3 1
S V O

According to the NSR, however, all sentences receive sentence stress on the rightmost element containing [1 stress]. In reality, this is not the case as we have seen above, e.g.:[5]

(11) The children didn't want to go to bed, so
(a) John scolded the governess
1
(b) John scolded the bastards
1

(13) (b) A proposal was made
1

Moreover, if one listens to a given piece of English dialogue, it becomes very difficult to accept Chomsky and Halle's (1968:25) claim that it is the listener's knowledge of the Nuclear Stress Rule that determines the stress contour of a given set of sentences within a context. One would tend to agree with Berman and Szamosi (1972:312) that it is perhaps misleading to carry over devices that worked well for the description of word-level phonology to the description of larger phrasal units and that 'some entirely novel ideas are necessary'.

2.2 Data

We have taped a number of episodes of the American television program 'Dynasty' in order to get some idea of how sentence stress patterns in English. Although it is too early to draw any definite conclusions, it would appear that the Nuclear Stress Rule fares very badly. A model such as that in (16′) seems to work much better in the analysis of sentence stress. Following is an example of the patterning of sentence stress in a piece of discourse from 'Dynasty'. The word or words that bear most prosodic prominence in each sentence have been written in bold letters. The subscript *c* has been used to indicate coreferent items that are not immediately obvious to the reader, e.g. things that had been talked about before the present dialogue began. Instances of obvious anaphoric coreference are indicated by the subscripts *i* and *j*:

[5] For sentences cited in isolation, names appear to act as contextually coreferent, even if they are mentioned only once. This is a peculiarity of names that has been noted by several scholars (see Ladd (1978:90–1) for a discussion of this phenomenon).

(21) Blake: (a) Dammit, Taft, we have a **crisis**$_i$ here.
(b) I **need** that information$_c$ within the next two hours.
(c) Now Taft, if the **sun** comes up on this one$_i$
and it's$_i$ not **settled**
we can all meet on the **unemployment** line.
(d) What is the **latest** obituary$_c$ from our mortician, Taft?
Taft: (e) It seems **definite**.
(f) The revolutionary government has confiscated the **oil**.
(g) Your **reserves** have been nationalized.
Blake: (h) Are you **hurrying** with those$_{c\ j}$?
Secretary: (i) As fast as I **can**, Mr. Laird.
Blake: (j) Andrew, I'm not **sure** I wanna go **through** with those papers$_j$
Taft: (k) Why **not**?
(l) It's a **gift**.
(m) There's no **law** that says a man can't give **a gift** to his **spouse**.
(n) Besides Blake, you have no other **choice**.
(o) I'm gonna have to **insist** that you **do** it.

The patterning of sentence stress in the dialogue in (21) seems to follow the hierarchy in (16). Thus, with reference to the flowchart in Figure 1, this patterning can be accounted for in the following way: in sentence (a), the object, a contextually noncoreferential noun, is stressed according to step (1a). In (b), however, the object, *that information,* has an antecedent earlier on in the dialogue and thus does not receive prominence. The subject, *I,* is also contextually coreferential. Consequently, the sentence stress falls on the verb, *need,* according to step (3a) in Figure 1. In the first conjunct in sentence (c), there is no predicate complement; consequently, the stress falls on the contextually noncoreferent subject, *sun,* according to step (2a). In the second conjunct, there is likewise no predicate complement; the subject, however, is coreferential with *crisis* in sentence (a); therefore, the stress falls on the verb, *settled* by (3a). In the main clause of sentence (c), the predicate complement, a noncoreferential locative adverbial, receives stress at (1a). In sentence (d), the predicate complement is not contextually coreferential; its head, however, *is* (i.e. it refers to 'news' discussed earlier in the dialogue); consequently, sentence stress falls on the modifier, *latest,* according to step (1b). The same patterning can be seen to exist throughout the remainder of the dialogue. We should point out here that we assume that each clause (S) of a sentence is submitted to analysis according to the model. Furthermore, we will assume that the assignment of sentence stress applies recursively at

each step in Figure 1 so as to account for instances where more than one stress is present (as in (21m)) (I owe this observation to Gösta Bruce).

We have also constructed possible texts where the sentence stress pattern follows the model in Figure 1. An example is found in (22), where identical subscripts indicate coreference. Numbers following the underlined words indicate at which point in Figure 1 sentence stress is applied. The text sounds very natural when read aloud:

(22) My mother$_i$ sent me some **money**$_j$ (1a) yesterday.
I really **needed** (3a) the cash$_j$.
I'd **spent** (3a) all my dough$_j$ and asked the dear$_i$ for a **loan**$_k$ (1a).
My brother$_l$ was of course **livid** (1a).
He$_l$ **refused** (3a) me a loan$_k$ earlier and the creep$_l$ called me a **spendthrift**$_m$ (1a).
That's absolutely **ridiculous** (1a)!
Now my **sister**$_n$**'s** (2a) spendthrift$_m$.
She$_n$'s a **real** (1b) squanderer$_m$!
⋮
⋮

3.1 Discussion

The model presented here subsumes what has sometimes been referred to in the literature as cases of 'anaphoric destressing' (Liberman 1978), 'deaccenting' (Bolinger 1972, Chafe 1974) or 'default accent' (Ladd 1978). The term 'deaccenting' seems conceptually inappropriate, since it implies a negative process, e.g. lowering pitch, whereas the approach taken here is to view sentence stress as the **addition** of some parameter(s) of phonetic prominence to an already existing word stress.[6] Deaccenting is assumed to apply in the following examples from Ladd (1978:52–3):

(23) A: John was mad because he got nothing but books for Christmas.
B: Oh, doesn't John **read** books?

(24) Harry wants a VW but his wife would prefer an **American** car.

where *books* in (23B) and *car* in (24) are assumed to be deaccented by some means or another. 'Default accent' is a better term, but what it describes is the natural output of the model presented here, i.e. sentence stress in a dialogue context. Note that according to our model, the sentence stress on

[6] This is the position taken, for example, by Bruce (1977).

read in (23B) would be placed at step 3a in Figure 1, and that on *American* (Modifier) in (24), at step 1b.

The hierarchy in (16) can account for the troublesome examples for Bresnan's and Berman and Szamosi's accounts of sentence stress such as

(6) (a) John has **plans** to leave

where *plans* is the object (stress according to step 1 in Figure 1),

(6) (b) John has plans to **leave**

where *to leave* is the object, *has plans* being analyzed as a complex verb with the meaning 'intends' (cf. Chomsky and Halle 1968:24) (stress according to step 1 in Figure 1), and

(12) John had **clams** for dinner

where *clams* is the object (stress according to step 1 in Figure 1). The model presented here is more general than previous models of sentence stress in that it can also explain examples accounted for by the NSR, Bresnan's Topicalization Rule, and Berman and Szamosi's Surface Structure Generalization, e.g.:

(7) (a) Mary liked the **proposal** that George left (NSR)

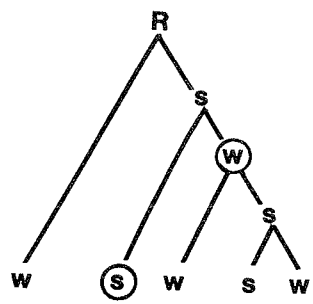
(10) Mary liked the **proposal** that was made (SS Generalization)

which, in the present model, are both assigned sentence stress by step 1 in Figure 1, and

(13) (f) **Guests** arrived (Topicalization Rule)

which, in our analysis is assigned sentence stress according to step 2 in Figure 1.

Notice that the present model also explains certain problematic cases for Liberman's (1978) metrical analysis of sentence prosody, e.g. (25)



If John were here, Sam'ld hit the bastard

With respect to this example, Liberman (1978:167–8) makes the following remarks:

> 'Presumably the reason that the constituent which contains the circled nodes has the form [sw], instead of the "null hypothesis" form [ws], is that anaphoric or redundant material is considered intrinsically **weak**, as opposed to anything that adds new information to the conversation. The other side of the coin is obviously that crucial or extra-important new information will want to be **strong**. Often it is hard to determine, in a particular case, what represents the result of destressing old information and what represents the result of stressing new information. One not unreasonable hypothesis would permit free assignment of either [ws] or [sw] to any phrasal constituent, but would stipulate that the pattern [sw] will be assumed to have some special pragmatic justification. The alternative is to postulate some set of features like [± contrastive], assigned to nodes, which influence prosodic and tonal phenomena in some way. The main difficulty in this area is that nobody has ever gotten the facts straight, to my knowledge.'

The fact that the verb *hit* receives sentence stress in Liberman's example is a natural outcome of the model presented here: both the object and subject are contextually coreferent[5] and consequently stress is assigned according to step (3a) in Figure 1.

The biggest problem in the application of our model lies in determining coreferential relations. In cases where anaphoric pronouns are used, the situation is very clear, e.g.:

(26) (a) Rats$_i$ like **cheese**$_j$
(b) It's$_j$ **good** for them$_i$

However, in more gramatically opaque cases of coreference or cases involving lexical relations such as synonomy, hyponomy and part-whole relations, the situation becomes more complicated, both for the listener in a concrete speech situation and for the linguist in attempting to account for the conditions under which such contextual identity is made. Consider, for example, the following two sentences uttered by Kurt Vonnegut in an interview on Swedish Radio:

(27) (a) When you laugh or cry, you are in fact throwing off (all sorts of things which have been released into (your bloodstream$_i$))$_j$ which will allow you to run or to stand and fight.
(b) And if you can do neither, then (your body)$_i$ has to get **rid** of (these chemicals)$_j$.

In this case, *these chemicals* in (27b) is coreferent with the phrase *all sorts of things which have been released into your bloodstream* and *your body* is related in a part/whole relation to *your bloodstream* in (25a). Note that it is the result of the speaker's placing sentence stress on the verb, *get rid* that we, as listeners, are able to infer such identity relations between what is being said and what has previously been said.[7]

Whether the concept of contextual coreference can be further developed in order to account for other cases of sense relationships involving coreference beyond the lexical level is an open question. For instance, in order to account for the so-called 'contrastive stress' patterns in sentences such as (28) (see Ladd 1978:78 ff.), the relation of coreference must be extended to cover identity of reference of semantic roles.

(28) Peter$_i$ kicked Ulrika$_j$ and then **she**$_j$ kicked **him**$_i$

In (28), the conditioning environment for sentence stress lies, not on the level of lexical referential identity, but rather on the level of identity of semantic roles, the stress pattern in this case serving to indicate a change in the 'actor/patient' relationship (see Enqvist 1979:140–1). It is possible that such nonlexical coreferential relations could be built into a surface syntactic model to the extent that they can be related to a change of grammatical functions in otherwise referentially identical surface structure strings. As it stands now, however, our model cannot account for such cases.

In conclusion, we should note that the dynamics of sentence stress is an issue of current interest in research on speech production and comprehension. We feel that the analysis presented here provides an insight into the direction in which further work in this area can proceed. Our model is, for example, well suited for application in computer models of verbal production such as that presented in Sigurd (1982).

References

Allerton, D. and Cruttenden, A. 1978. Syntactic, illocutionary, thematic and attitudinal factors in the intonation of adverbials. *J. of Pragmatics* 2. 155–188.

Berman, A. and Szamosi, M. 1972. Observations on sentential stress. *Language* 48. 304–325.

Bolinger, D. 1972. Accent is predictable (if you're a mind reader). *Language* 48. 633–644.

Bresnan, J. 1971. Sentence stress and syntactic transformations. *Language* 47. 257–281.

– 1972. Stress and syntax: a reply. *Language* 48. 326–342.

[7] See Nooteboom and Terken (1982) for experimental support for this notion.

BRUCE, G. 1977. *Swedish word accents in sentence perspective.* Lund: Gleerups.
CHAFE, W. 1974. Language and consciousness. *Language* 50. 111–133.
CHOMSKY, N. and HALLE, M. 1968. *The sound pattern of English.* New York: Harper & Row.
DIK, S. 1978. *Functional grammar.* Amsterdam: North-Holland.
ENQVIST, N. 1979. Marked focus: functions and constraints. In Greenbaum *et al.* 1979, pp. 134–152.
FIRBAS, J. 1979. Post-intonation-centre prosodic shade in the modern English clause. In Greenbaum *et al.* 1979, pp. 125–133.
GREENBAUM, S., LEECH, G. and SVARTVIK, J. (Eds.) 1979. *Studies in English linguistics for Randolph Quirk.* London: Longman.
LADD, D. 1978. *The structure of intonational meaning.* Bloomington: Indiana University Press.
LIBERMAN, M. 1978. *The intonational system of English.* Bloomington: Indiana University Linguistics Club.
LIBERMAN, M. and PRINCE, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8. 249–336.
LYONS, J. 1969. *Introduction to theoretical linguistics.* Cambridge: Cambridge University Press.
MCNEILL, D. 1970. *The acquisition of language. The study of developmental psycholinguistics.* New York: Harper & Row.
NOTTEBOOM, S. and TERKEN, J. 1982. What makes speakers omit pitch accents? An experiment. *Phonetica* 39. 317–336.
QUIRK, R., GREENBAUM, S., LEECH, G., and SVARTVIK, J. 1972. *A grammar of contemporary English.* London: Longman.
SCHMERLING, S. 1976. *Aspects of English sentence stress.* Austin: University of Texas Press.
SIGURD, B. 1982. Commentator: a computer model of verbal production. *Linguistics* 20. 611–632.

II

INFORMATION FOCUS: ASSIGNMENT AND PHONOLOGICAL IMPLICATIONS

Merle A. Horne

Abstract

Data are presented in support of the claim that the conditioning of information focus within discourse is determined by the interaction of two parameters: a hierarchy of grammatical functions and the coreferential status of the lexical material filling these functions with respect to preceding parts of the discourse. A process model is presented for the assignment of focus and ramifications for the phonological component are discussed.

1. Introduction

A problem that has plagued phonologists for many years (and still is) is that surrounding the dynamics of focus placement and its phonological realization. The intractability of the problem has, of course, been partly due to the fact that phonologists have on the whole limited themselves to the study of sentence grammar. As Enkvist (1979:151) has pointed out, however, one must "venture beyond sentence grammar" in order to study the phenomena associated with focus. That is to say, focal prominence is a discourse phenomenon and cannot be adequately accounted for by limiting oneself to data consisting of isolated sentences. The interesting question then becomes: Is it in any way possible, using the concepts available within linguistic theory to develop a model that would, to a statistically significant degree, predict where grammatically conditioned focus would be placed? Such a model could, for instance, be integrated into a program for speech production (see Sigurd 1982) or perhaps be used in teaching prosody. We should note that we are for the present purposes abstracting away from focus conditioned by such factors as emphasis or corrective ends, and concentrating on what has been termed 'information focus'

or what we are assuming to be grammatically conditioned focus.

2. Background

Early work on focus or what was termed 'sentence stress' by generative phonologists was limited to the domain of the isolated sentence. Normal or 'unmarked' sentence stress was assumed to be determined on the basis of word stress and abstract syntactic structure, i.e., the rightmost lexical item containing primary stress was assigned sentence stress. Deficiencies in this analysis were soon pointed out (cf. Bresnan 1971, 1972; Berman & Szamosi 1972; Bolinger 1972). For instance, the Nuclear Stress Rule could not correctly place sentence stress or focus in such 'unmarked' cases as those in (1):

(1) a) WORK must be done.
b) The kids had FISH for dinner.

since it blindly looked for the last stressed item in a sentence, in this case, done and dinner. Placement of focus on these words, however, automatically leads to a marked or contrastive reading.

The situation did not improve much with the dawn of metrical phonology, since the NSR was integrated into this theoretical framework as well (see Liberman & Prince (1977); Giegerich (1983). Ladd (1978), however, noted that certain cases of so-called 'default accent' could be accounted for by assuming a change in the placement of metrical S's that determine the degrees of phonological prominence. For instance, the pair of sentences in (2) is cited by Ladd (1978: 53) as a typical case of default accent:

(2) A: John was MAD because he got nothing but BOOKS for Christmas.
B: Oh, doesn't John LIKE books?

where, in the B) sentence, BOOKS is deaccented and focus moves in this instance to the left sister, the verb LIKE. Just how focus gets placed on BOOKS in the first place is not

explained nor for that matter can Ladd explain the focus placement in the sentences in (1).

In her latest work on the interaction of phonology and syntax, Selkirk (1984) does away with the level of metrical structure and assumes that all information regarding sentence prosody can be captured by means of metrical grid structure. Assignment of focus is assumed to take place at the level of logical form. She insists that "the location of focus relevant prosodic prominence within a focused constituent is not determined by principles such as the NSR; second, that it is the argument structure of the phrase and sentence that is crucial in determining focus properties of higher constituents on the basis of the prominences within them" (1984:203,206). Here, however, as in Ladd's analysis, focus accent is assumed to be freely assigned.

3. Model Presentation

Our own work has been concerned with the attempt to systematize certain facts about focus and to develop a model that will make the initial focus placement. We assume, like Selkirk, that it is at some level of semantic/syntactic structure that the assignment of focus takes place. We hypothesize, furthermore, that the dynamics of focal accent placement within a discourse is determined by the interaction of two parameters: a hierarchy of grammatical functions and the coreferential status of the lexical items filling these functions with respect to preceding parts of the discourse. We thus view the linguistic function of focus to be twofold: On the one hand, it serves to give information concerning subcategorization relationships between the predicate and its arguments within a given proposition. This could perhaps be interpreted as the highest degree of 'valency' with the predicate in the sense of Enkvist (1976), that is to say, an 'obligatory' argument receives focus. On the other hand, focal prominence (or rather its absence) gives information on the lexical relationship between newly introduced terms and terms already introduced into the discourse situation.

All other things being equal, arguments will receive more prominence than predicates. This fact has been observed by Schmerling (1976), for instance. Furthermore, as far as English is concerned, a predicate complement, i.e. a semantically obligatory nonsubject argument, will, all things being equal, receive more prominence than the subject argument. Compare, for example, the sentences in (3):

(3) a) A man bought a BOOK.
b) A MAN appeared.
c) *A MAN bought a book.
d) *A man APPEARED.

As 'news' sentences, i.e., as possible answers to the question "What happened?", (3) (ab) are possible answers, whereas (3) (cd) are not, i.e. they presuppose a certain amount of 'given' information; c), for example, could be construed as a potential reply to a question such as: "Who bought a book?", whereas d) could be construed as forming part of a corrective reply such as: "She didn't say 'A man disappeared'; she said:____". Thus we see that contextually coreferent material produces the same effect on the position of focus as the absence of a given grammatical function, that is to say, when both the subject and the object for example are present and have not been previously mentioned in the discourse, the focal prominence falls on the object. However, when there is no object or when the object is contextually coreferent, then the focal accent falls on the subject, as in b). And, of course, should there be no subject or object in a given sentence (as in the case of an imperative sentence) or if both the subject and object are contextually coreferent, then the focal prominence will end up on the verb or predicate. In light of these observations, we might attempt to represent the potential for assignment of focus on the basis of a hierarchy such as that in (4):

(4) PREDICATE COMPLEMENT>SUBJECT>PREDICATE

As it stands, however, the picture is somewhat simplified since it fails to indicate where focus would be realized

in cases where a given grammatical function was filled by a complex term, for example an attributively modified noun. In Horne (1985), we chose to explain the observed patterning on a subphonological level, using the syntactic head-modifier relation. That is to say, all things being equal, it is of course the head of the given construction that is assigned the focal accent:

(5) a) A big black DOG appeared.
b) I bought some new RECORDS today.

The 'head-modifier' relationship can be seen to be very enlightening in explaining the default cases. (Ladd (1984), furthermore, avails himself of the head-modifier relation to explain the origin of compound words in English.) This becomes clear when one examines a piece of connected text. It is here, also, that the second parameter, what we have termed 'contextual coreference' comes into play in conditioning the position of focus. Contextual coreference includes many different kinds of lexical relationships, everything from repetition of a lexical item and pronominalization to relationships such as synonymy, metonymy and part-whole relationships (see Allerton 1978:140-1). It is this parameter which takes the upper hand in a discourse situation. Consider, for instance, the pairs of sentences in (6) which demonstrate how inadequate the NSR really is accounting for focus placement in such a situation:

(6) a) i. My mother gave me a $[\text{yellow } [\text{FORMAL}_i]]_j$.
ii. I really wanted a $[\text{BLUE } [\text{dress}_i]]_k$.
b) i. My mother gave me a $[\text{yellow } [\text{FORMAL}_i]]_j$.
ii. I'm going to WEAR the $[\text{gorgeous } [\text{gown}_i]]_j$ today.

As the sentences in (6) demonstrate, it is not possible to account for the placement of focus solely by means of the NSR. The second sentence in each pair of examples illustrates this well, since the NSR would automatically place prominence on the rightmost lexical item in the respective phrases. In the approach taken here, it is the object phrases in the i) sentences that would be assigned focus accor-

ding to the hierarchy in (6). Moreover, the prominence falls on the head of the object phrase. In the ii) sentences, however, we notice that the situation is somewhat different in the two cases. In the a) case, the modifier receives prominence, whereas in the b) case it is the predicate that becomes focussed. This difference can be explained, however, by reference to the concept of contextual coreference. In a), the object phrases are not coreferential but their heads <u>are</u>. In a)i, the head receives prominence and in a)ii, it is the modifier that is accented. In b), on the other hand, the object phrases as wholes are coreferential. In b)i, as in a)i, it is the head that receives focal accent. In b)ii, however, the prominence shifts to the predicate. Thus when focal accent is assigned according to the hierarchy in (4), one must assume that the whole phrase in question is first scanned to determine its coreferential status with respect to the preceding text. If it is the case that the entire phrase is coreferential, then one moves to the subsequent step in the hierarchy. If it is only the head that is coreferential, however, then it is the modifier that will attract the focus. With all these facts in mind, the process of focus placement can be formally represented by means of a flow chart such as that in Figure 1 (we assume in what follows that the model applies within each clause (S) of a given sentence:

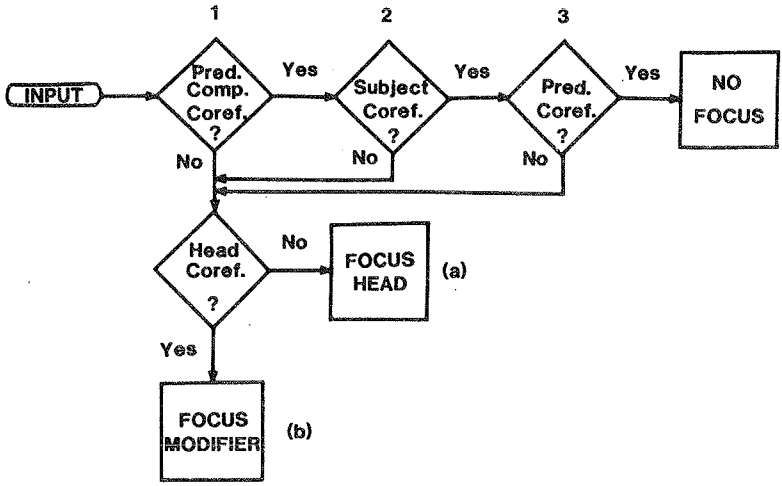


Figure 1

Figure 1 is to be interpreted as follows:

To assign focus in a given fragment of discourse, proceed in the following way: (1) check the Predicate Complement if there is one; if the whole constituent is noncoreferential with respect to the preceding context, assign the Head sentence stress if it is likewise noncoreferential (1a). If the head is coreferential, however, assign stress to the Modifier (1b). If the whole Predicate Complement is contextually coreferent on the other hand, go to (2), Subject. The same procedure is repeated here as in the case of the Predicate Complement. Should the procedure be carried to the Predicate position and a 'Yes' response be obtained even here as regards the coreferential status of the whole Predicate phrase, then this indicates that the sentence contains no new information as regards lexical content. This does not mean, however, that the sentence does not qualify to receive other kinds of focal prominence as the

example in (7) shows:

(7) John$_i$ hit$_j$ MARY$_k$ and then SHE$_k$ hit HIM$_i$.

The model in Figure 1 can account for the focus on MARY but not on the pronouns SHE and HIM since coreferential items do not attract focal accent in the 'normal' case. Focus in this instance differs from the 'normal' case in that it requires certain conditions of contextual identity for its appearance. Focal accent in such cases involving parallel structures is applied to corresponding arguments in adjacent sentences that have semantically identical predicates. The function of focus here, as Enkvist (1978,141) has noted is to indicate a "shift in semantic roles". Just how contrastive accent is to be derived in a given case is an interesting question but notice that the model in Figure 1 provides a natural point to proceed to account for them: a 'Yes' response at stage (3) in the model characterizes precisely those cases where we find focus accents on contextually coreferent material such as in (7). A logical solution would be to proceed to examine the patterning of semantic functions of the arguments in such parallel cases. Should they differ, then a special subroutine would assign focal prominence to the material marked with the relevant semantic function, e.g.:

(7´) John$_i$ Ag hit$_j$ Mary$_k$ Go and then she$_k$ Ag hit$_j$ him$_i$ Go.

In this case, she and him would be assigned prominence since their semantic roles, Ag(ent) and Go(al), respectively, differ from those assigned to their antecedents in the preceding discourse. Another case where contextually coreferent material receives focal prominence due to conditioning on the semantic level is that involving a change in the polarity value of the predicate. When this involves going from the negative to positive value for instance, this change is marked by focal prominence which is realized on the first auxiliary verb, eg.:

(8) a. John doesn't like SPAGHETTI.

b. He DOES like spaghetti.

Here again, our model in Figure 1 could not predict the focus in (8) b. since there is no contextually noncoreferent material. If one assumes, however, that a semantic feature of polarity is assigned to the predicate and is available to the focus projection rules just like other semantic function features, then these examples fall together with those in (7).

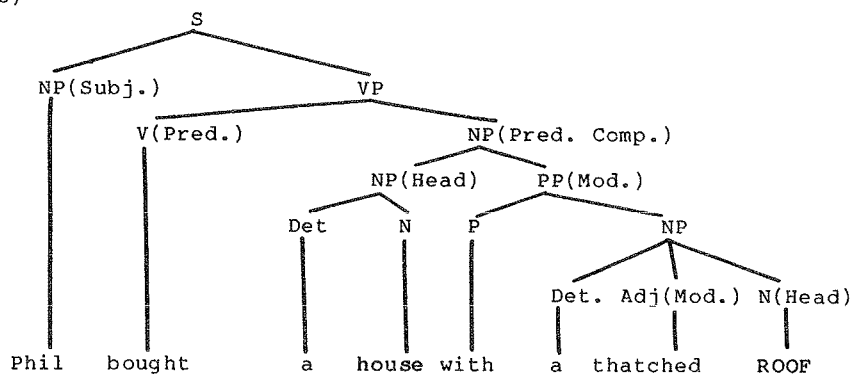
4. Phonological implications of focus assignment

Before we proceed to illustrate the functioning of our model, a few details should be discussed concerning the phonological ramifications of information focus assignment. Consider, for example, a sentence such as that in (9):

(9) Phil bought a house with a thatched ROOF.

which can be assigned the syntactic structure in (10a):

(a) (10)



b) FOCUS ASSIGNMENT

i. (1a)(recursive)		x	x
ii. Phrasal Rule			x

According to the model in Figure 1, focus would be assigned to the Predicate Complement at Stage 1. One question then becomes: how is focus to be realized in a case where the

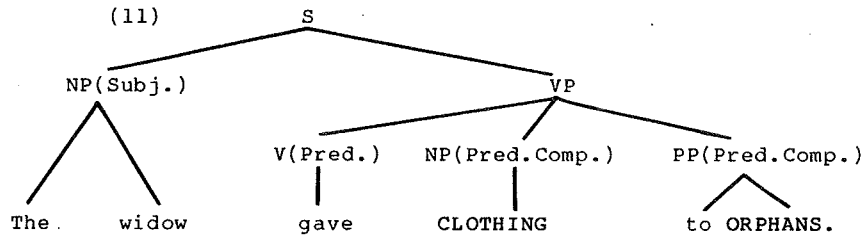
constituent realizing the grammatical function contains more than one head? We will make the assumption that all the heads in the given constituent are assigned prominence. This decision, of course, requires us to include an additional rule in the phonological component that gives even greater prominence to the rightmost focussed element. Such a phrasal rule has been generally assumed, however (see Newman 1946; Selkirk 1984) and would also seem to be uncontroversial among phoneticians (see Bruce 1982, Gårding 1983). Thus in the case of the model in Figure 1, we will consider that the subroutine (a-b) applies recursively within each headed construction realizing a grammatical function and assigns phonological prominence which can be represented by means of a grid as in (10b i.). Subsequent application of a phrasal rule (10b ii.) would assign additional prominence to the rightmost focally prominent element.

It could be argued that it is not necessary to assign prominence to each head, but rather assume just a phrasal rule that assigned prominence to the rightmost noncontextually coreferent lexical item in the given constituent. This would have as a consequence that house, for example, in (10) would receive no focal prominence whatsoever. Such a situation does not seem to correspond to reality, however, in the case where house is to be considered as new information. Auditorily, there is a secondary degree of prominence given to house in this instance which is not perceived when house is construed as old information. A very definite pitch obtrusion is, furthermore, observed in the intonation curves of corresponding elements in similar sentences in Swedish, Greek and French in Gårding (1983).

Moreover, as well as applying recursively within each headed construction, notice that the assignment of phonological prominence must be assumed to occur iteratively on the level of grammatical functions in order to account for cases where there is more than one instance of a given function realized in a particular sentence. This is the case, for instance, in sentences where the Predicate Complement consists of two obligatory arguments, such as the sentence in (11) (subsequent application of the phrasal rule on the

VP level would give ORPHANS more prominence in relation to CLOTHING):

(11)



a) FOCUS ASSIGNMENT		x	x
b) Phrasal Rule			x

Another point that should be made here is that it could be argued that the hierarchy in (4) should in practice be interpreted differently. That is to say, that in a given clause, different degrees of focal prominence should be assigned to the different grammatical functions. This was hinted at in Horne (1985). That is to say, instead of just assigning one grammatical function focus, one could rank all the functions in a given sentence according to the hierarchy in (4). Thus we could derive prominence contours such as those in (12), ranking, of course, only contextually noncoreferent constituents:

(12) a) Policemen attacked innocent people.

S	P	PC
2	3	1

b) A flying star suddenly appeared.

S	P
1	2

Such an interpretation of the hierarchy is no doubt the correct one, since it intuitively seems necessary in the case of 'all-new' sentences where it can be argued that both the subject and predicate complement have focal prominence (see Gussenhoven 1983). Phonetically, however, the

Predicate Complement has more highlighting than the subject (see Gårding 1983).

5. Data

In order to illustrate the model presented here the following fragment of discourse taken from the American television program, 'Dynasty', will be used. The subscript c has been used to indicate coreferent items that are not immediately obvious to the reader, e.g. things that had been talked about before the dialogue began. Instances of obvious anaphoric coreference are indicated by the subscripts i and j. The most prominent word in each clause has been written in bold letters:

(12)

Blake: (a) Dammit, Taft, we have a **CRISIS**$_i$ here.
(b) I **NEED** that information$_c$ within the next two hours.
(c) Now Taft, if the **SUN** comes up on this one$_i$ and it's$_i$ not **SETTLED**, we can all meet on the **UNEMPLOYMENT** line.
(d) What is the **LATEST** obituary$_c$ from our mortician, Taft?

Taft: (e) It seems **DEFINITE**.
(f) The revolutionary government has confiscated the **OIL**.
(g) Your **RESERVES** have been nationalized.

Blake: (h) Are you **HURRYING** with those$_{c\ j}$?

Secretary: (i) As fast as I **CAN**, Mr. Laird.

Blake: (j) Andrew, I'm not **SURE** I wanna go **THROUGH** with those papers$_j$.

Taft: (k) Why **NOT**?
(l) It's a **GIFT**.
(m) There's no **LAW** that says a man can't give a **GIFT** to his **SPOUSE**.
(n) Besides Blake, you have no other **CHOICE**.
(o) I'm gonna have to **INSIST** that you **DO** it.

The patterning of (main) focal accent placement in the dialogue in (12) would appear to conform to the conditions specified in the flowchart in Figure 1. For example, in sentence a), the object (Pred. Comp.), that information, has an antecedent earlier on in the dialogue and thus does not receive prominence. The subject, I, is also contextually coreferential. Consequently, focal accent falls on the predicate, **NEED**, according to step (3a) in Figure 1. In the first conjunct in sentence c), there is no predicate complement; consequently, the accent falls on the contextually noncoreferent subject, sun, according to step (2a). In the second conjunct, there is likewise no predicate complement; the subject, however, is coreferential with crisis in sentence a); therefore, the accent falls on the predicate, settled by (3a). In the main clause of sentence c), the predicate complement, a noncoreferential locative adverbial, receives accent at step (1a). (The fact that all listeners of the dialogue agree that there is a 'secondary' accent on meet would lend support to the suggestion above that the grammatical functions be ranked according to the hierarchy in (4), in this case assigning 1 to the Predicate Complement and 2 to the Predicate.) In sentence d), the predicate complement is not contextually coreferential; its head, however, is (i.e. it refers to 'news' discussed earlier on in the dialogue); consequently, focal accent falls on the modifier, latest, according to step (1b), in Figure 1 or according to the prominence assignment rule of Figure 2. The same patterning can be seen to exist throughout the remainder of the dialogue.

6. Conclusion

Within recent years, there has been a considerable amount of research being directed towards the understanding of information focus (cf. Wells & Local 1983; Selkirk 1984; Gibbon & Richter 1984 (in particular the interesting articles by Fuchs and Ronat); Bolinger 1985 and the works cited therein). This is not surprising, in view of the increasing interest in fields such as discourse analysis, artificial

intelligence and speech synthesis. The structuring of information is a complex and dynamic process. We hope to have made a contribution to the understanding of this process by factoring out what we feel to be the 'linguistic' correlates of information focus. Much research remains to be done as far as determining for example what linguistic factors determine relations of contextual coreference (see, for ex. Ronat 1984). As regards phonology, the model presented here demonstrates just how dependent the phonological component is on semantic/syntactic information in explaining prosodic phenomena.

References

Allerton, D. (1978). The notion of 'givenness' and its relations to presupposition and to theme. Lingua 44: 133-68.

Berman, A. and M. Szamosi (1972). Observations on sentential stress. Language 48: 304-25.

Bolinger, D. (1972). Accent is predictable (if you're a mind reader). Language 48: 633-44.

- (1985). Two views of accent. J. of Linguistics 21: 79-138.

Bresnan, J. (1971). Sentence stress and syntactic transformations. Language 47: 257-81.

- (1972). Stress and syntax: a reply. Language 48: 326-42.

Bruce, G. (1982). Developing the Swedish intonation model. Work. Pap. Linguist. Univ. Lund. 22: 51-114.

Enkvist, N.E. (1976). Notes on valency, semantic scope, and thematic perspective as parameters of adverbial placement in English. In: Reports on Text Linguistics: Approaches to Word Order, N.E. Enkvist and V. Kohonen (eds.). Åbo: Meddelanden från Stiftelsens för Åbo Akademi Forskningsinstitut, No. 8.

- (1979). Marked focus: functions and constraints. In: Studies in English Linguistics for Randolph Quirk, S. Greenbaum, G. Leech & J. Svartvik (eds.). London: Longman.

Fuchs, A. (1984). 'Deaccenting' and 'Default Accent'. In: Gibbon & Richter.

Gårding, E. (1983). A generative model of intonation. In: Prosody: Models and Measurements, A. Cutler and D. Ladd (eds.). Berlin: Springer.

Gibbon, D. and H. Richter (1984). Intonation, Accent and Rhythm. Studies in Discourse Phonology. Berlin: W. de Gruyter.

Giegerich, H. (1983). On English sentence stress and the nature of metrical structure. J. Linguistics 19: 1-28.

Gussenhoven, C. (1983). Focus, mode and the nucleus. J. Linguistics 19: 377-417.

Horne, M. (1985). English sentence stress, grammatical functions and contextual coreference. Studia Linguistica (in press).

Ladd, R. (1978). The Structure of Intonational Meaning. Bloomington: Indiana University Press.

- (1984). English compound stress. In: Gibbon & Richter.

Liberman, M. and A. Prince (1977). On stress and linguistic rhythm. Linguistic Inquiry 8: 249-336.

Newman, S. (1945). On the stress system of English. Word 2: 171-87.

Ronat, M. (1984a). Le liage prosodique: l'intonation prosodique: l'intonation comme anaphore. In: Recherches sur l'anaphore. Département de recherches linguistiques, Université de Paris 7.

- (1984b). Logical form and prosodic islands. In: Gibbon & Richter.

Schmerling, S. (1976). Aspects of English Sentence Stress. Austin: University of Texas Press.

Selkirk, E. (1984). Phonology and Syntax: the Relation between Sound and Structure. Cambridge: MIT Press.

Sigurd, B. (1982). Commentator: a computer model of verbal production. Linguistics 20: 611-32.

Wells, W. and Local, J. (1983). Deaccenting and the structure of English intonation. Linguistics 21: 701-15.

III

Merle A. Horne

FOCAL PROMINENCE AND THE 'PHONOLOGICAL PHRASE' WITHIN SOME RECENT THEORIES

1. Introduction*

The purpose of this paper is to discuss the status of the prosodic category 'phonological phrase' as discussed in some recent phonological analyses (Selkirk 1980, 1984; Nespor & Vogel 1982, 1983). It will be argued that it is unnecessary to assume such a category in the phonology if one assumes that prosodic rules have direct access to information concerning focus structure. The prosodic phenomena assumed by the above mentioned authors to be restricted to occurring within the unit termed 'phonological phrase' can instead be analyzed as due to prominence relations determined on the basis of focus assignment and general rhythmic constraints. These prominences are further assumed to be realized on the unit (stressed) 'syllable'.

2. Information focus assignment

In order to better understand the discussion of the phonological phrase, we will outline here the factors relevant for assigning information focus. Details are to be found in Horne 1985, 1986. The model will be summarized and elaborated on here. The placement of information focus (i.e. grammatically and textually conditioned focus, not expressive focus) in English can be formally represented by means of a flow chart such as that in Figure 1 (we assume that the model applies within each clause (S) of a given sentence:

* I am grateful to A. Botinis, D. House, T. Pettersson, B. Sigurd and P. Touati for comments on earlier versions of this paper. D. House and G. Jönsson are thanked for assistance with recording and instrumental analyses. B. Sigurd and O. Söderman are also acknowledged for their help in designing the flowchart in Figure 1.

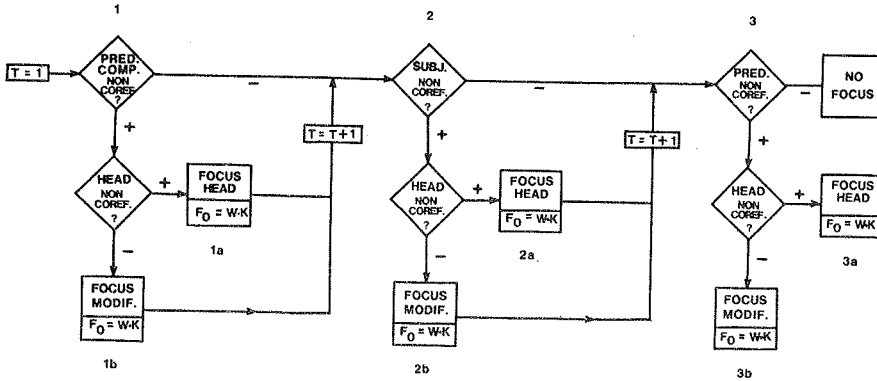


Figure 1. Model (flowchart) for assigning information focus to constituents on the basis of grammatical function and number of focused constituents. Preliminary values for the realization of focus as pitch (F_0) are also included. W designates the width of the grid within which F_0 moves. K is a variable ranging over a series of 'target values', fractions (extending from 0–1) of the distance from the baseline (0) to the topline (1). Here, K is assigned the values 1 (for the first focused constituent), 0.8 (for the second focused constituent) and 0.4 (for the third focused constituent).

In order to illustrate the model, we will show how it could be applied to derive one prosodic parameter, the intonation contour (F_0). As data, we will consider the F_0 contour of the last sentence of the dialogue in (1), spoken by an American male, which is reproduced in Figure 2. As a working hypothesis, we have assumed that, in Figure 2, the baseline of the grid [1] corresponds to the speaker's initial and final voice frequency. These were determined to be 130 Hz and 85 Hz, respectively, after comparison with several other utterances spoken by the same speaker. The topline of the grid has provisionally been drawn parallel to the baseline and passes through the top of the peak corresponding to what the model in Figure 1 would predict as receiving the greatest prominence (in this case, the Object *mayor)*. The values 1.0, 0.8, and 0.4 are 'target values', fractions (extending from 0–1) of the distance from the baseline (0) to the topline (1) (see Pierrehumbert 1981 who uses similar target values):

(1) A: I'm just about finished writing my new book.
B: Oh, could you let me in on how it's going to end?
A: Yea, sure. A Mormon will marry a mayor[2].

[1]For a discussion of the notion 'grid' see Gårding 1983.
[2]Following Bruce (1977), we have restricted ourselves to using only sonorant segments in order to obtain an Fo contour as free from local pertubations as possible.

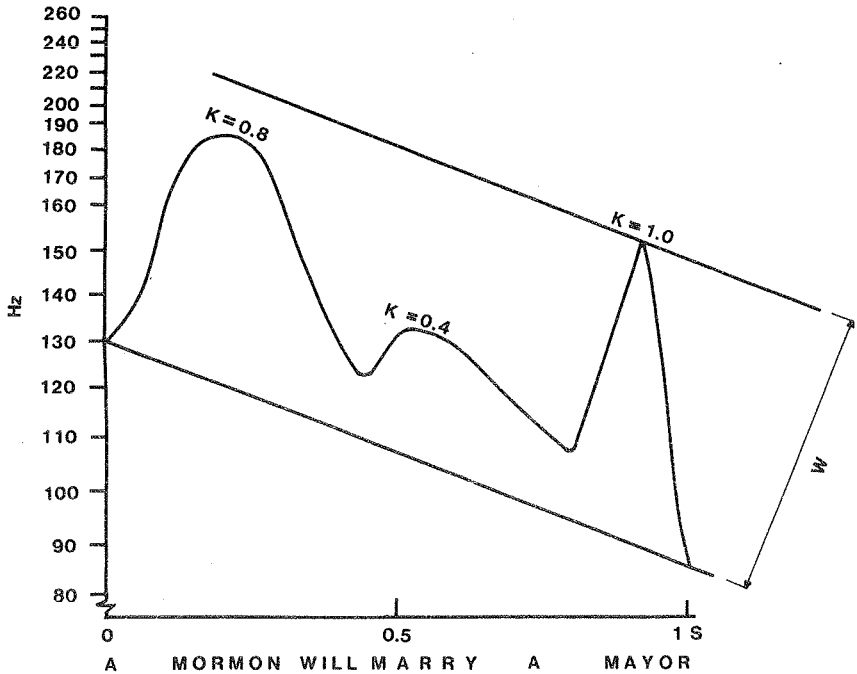


Figure 2

Figure 1 is to be interpreted as follows: To assign information focus within a given fragment of discourse, proceed in the following way: (1) check the Predicate Complement if there is one; if the whole constituent is noncoreferential with respect to the preceding context, focus the Head if it is likewise noncoreferential (1a). In Figure 2, the head *mayor* receives the highest degree of focal prominence, in this case a 'target' of 1. If the Head is coreferential, however, assign focus to the Modifier (1b). Proceed then to (2), Subject. The same procedure is repeated here as in the case of the Predicate Complement, but the amount of prominence assigned to the Subject is the grid width (W) multiplied by 0.8 in the case where the Predicate Complement has already been assigned focal prominence[3]. Subsequently, the Predicate is examined for its coreferential status. In Figure 2, the predicate is assigned a target of 0.4. As can be seen, the amount of prominence given to the various functions is directly proportional to their

[3]This target value is somewhat high we believe, since it includes a contribution from what could be analyzed as 'initial juncture' as well as from focus assignment. Further experimental work is needed before the various target values can be fixed.

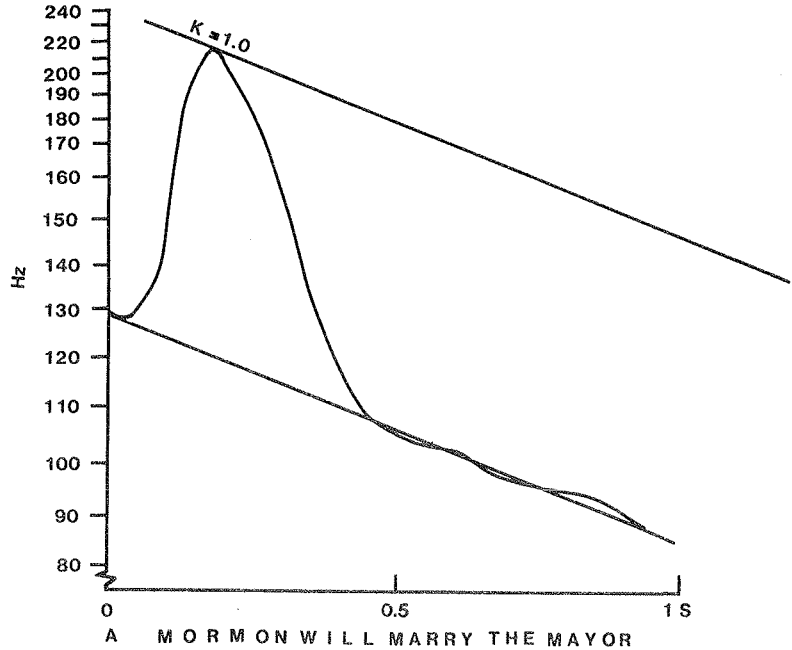


Figure 3

position in the hierarchy i.e. a Predicate Complement receives more prominence than a Subject which in turn receives more prominence than a Predicate. For the sake of illustration, Figure 2 can be compared with Figure 3 in which the intonation contour of the last sentence of the dialogue in (2) is presented. This sentence differs from the corresponding one in (1) in that the Predicate Complement and the Predicate are coreferential with material in the preceding part of the dialogue. Consequently, this coreferential material receives no focal prominence whatsoever, while the Subject receives the same amount of prominence as the Predicate Complement in (1), i.e. a target of 1:

(2) A: My new book is about a mayor$_i$ living in Malmö. He$_i$ meets an interesting person there and gets married$_j$.
B: Oh, could you let me in on who marries$_j$ him$_i$?
A: Yea, sure. A Mormon will marry$_j$ the mayor$_i$.

3. Phonological Phrase

The term 'phonological phrase' as introduced by Chomsky and Halle (1968)[4] was developed by Selkirk (1980) and Nespor and Vogel (1982, 1983) in order to explain certain prosodic phenomena within the phonology of English and Italian[5]. In particular, it was maintained by Selkirk that the so-called Rhythm Rule (or Iambic Reversal) and the Monosyllabic Rule in English were restricted to operating within the domain of the phonological phrase, while Nespor and Vogel claimed that the rule of Raddoppiamento Sintattico also had as its domain the phonological phrase. The phonological phrase is, however, defined somewhat differently in the two cases. We shall examine these processes in what follows.

3.1. Selkirk (1980): Rhythm Rule

Selkirk (1980) defines the phonological phrase as follows:

(3) The Phonological Phrase: Constituency
i) An item which is the specifier of a syntactic phrase joins with the head of the phrase.
ii) An item belonging to an 'non-lexical' category (cf. Chomsky 1965) such as Det, Prep, Comp, $Verb_{aux}$, Conj, joins with its sister constituent.

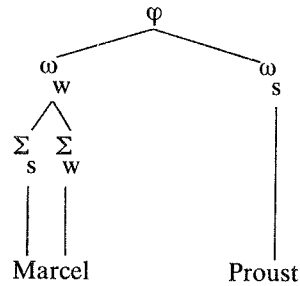
Evidence for the existence of the phonological phrase is derived from the behaviour of certain phonological rules. One of these is the so-called Rhythm Rule which is assumed to be responsible, for example, for

[4]Some related terms are 'tone units' (Crystal 1969; Svartvik 1982), 'speech chunks' (Sigurd 1984), or 'prosodic phrases' (Gårding & House 1985). These units are characterized by Gårding & House, for ex. as "a part of an utterance in which accents or tones are organized in a common unbroken intonation movement" (1985:205) [my translation, MH]. The examples in Fig. 2 and 3, for instance, would thus be considered to constitute one tone unit. The division of an utterance into tone units is a late procedure, taking place after the assignment of focal prominence and as Crystal (1969) notes, factors such as constituent length and speech tempo come into play in determining where an individual speaker will place tone unit boundaries. One could predict, however, that a tone unit boundary could potentially be placed before each focussed element.

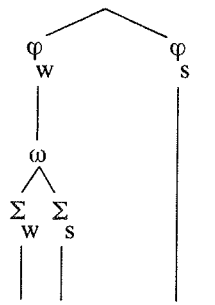
[5]Chomsky and Halle assumed that phonological phrases were formed upon the application of certain readjustment rules to the syntactic surface structure in order to create an appropirate input to the phonological component:
"The readjustment rules will modify the surface structure in various ad hoc ways demarcating it into phonological phrases, eliminating some structure and replacing some occurrences of # by +" (1968:13).

changing the ws pattern of thirteen to sw when it is followed by a s as in thirteen men. In order to illustrate how the phonological phrase functioned in restricting the application of the Rhythm Rule, the difference between the rhythmic structure of the two utterances in (4) is compared:

(4) a.



b.



where Σ = foot, ω = prosodic word, φ = phonological phrase.
In (4a), *Marcel Proust* is assumed to exhibit the rhythmic structure sws. This demonstrates that the Rhythm Rule has applied here, since the basic rhythmic form of *Marcel* is ws. In (4b), however, the utterance *Marcel proved* is assumed to exhibit the rhythmic structure wss. This difference, or the nonapplication of the Rhythm Rule in (4b) is accounted for by assuming that this process is restricted to applying within the phonological phrase as defined in (3).

This account does not hold, however, since we would expect that the same pattern would surface in the case of the contrast between the phrase *kangaroo court* and *kangaroos kick,* in the fragment:

I'm studying kangaroos$_i$. *Did you know that kangaroos*$_i$ *kick a lot?*

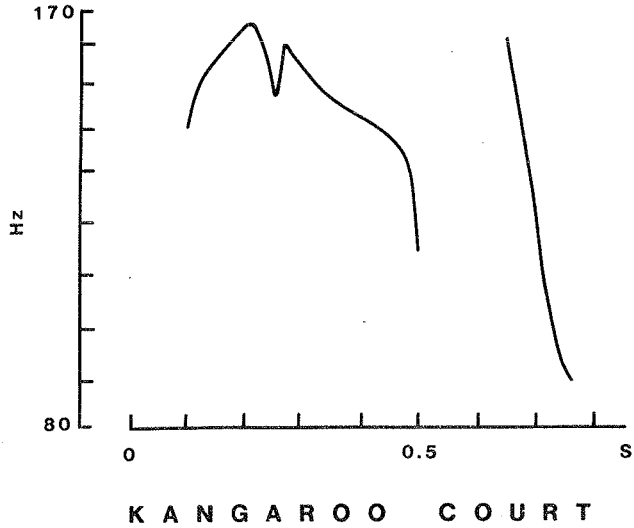


Figure 4

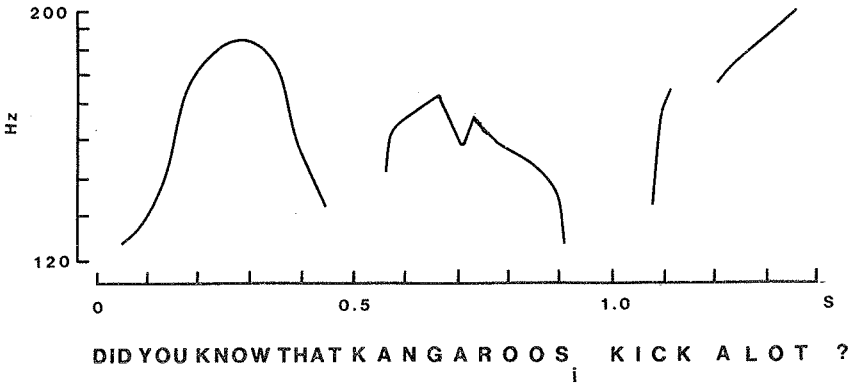


Figure 5

That is to say, we would expect a contrast between ws in *kangaroo court* (w s) and ss in *kangaroos kick* (s s) (the basic rhythmic structure of *kangaroo* being sws), the lack of application of the Rhythm Rule in the latter instance being due to the fact that *kangaroo* and *kick* belong to different phonological phrases. Such is not the case, however, for *kangaroos kick* also has the rhythmic structure ws, i.e., the Rhythm Rule applies in this case as well, giving *kangaroos kick.* (w s) In way of illustration, compare the almost identical F_0 contours of *kangaroo* in Figures 4 and 5 which bear witness to

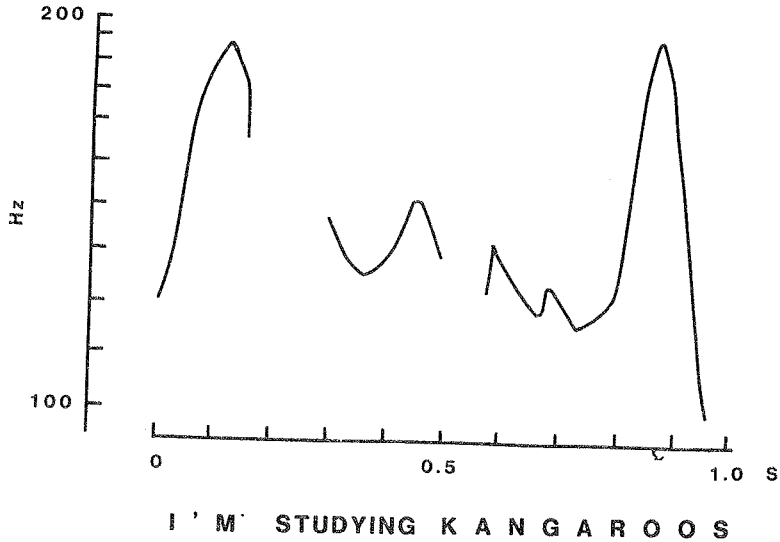


Figure 6

the application of the Rhythm Rule. These can in turn be compared with the contour of *kangaroo* in Figure 6 where the prominence on the final syllable shows that the Rhythm Rule has not applied.

3.2. Reinterpretation of the Rhythm Rule

On the basis of the above evidence, therefore, it is clear that it is not a level of syntactic structure that conditions the so-called Rhythm Rule but rather it would appear to be conditioned by a relation of prominence holding between the two lexical items involved. In order for the "Rhythm Rule" to apply, the right-hand element must be more prominent than the element to its left. What has been termed the "Rhythm Rule" is, therefore, better interpreted as the nonrealization of an 'inherent'[6] word stress followed by the addition of rhythmical prominence on the initial syllable. The attempt to restrict the phenomenon to a syntactically determined unit termed the phonological phrase does not work as we have seen above. In . . . *kangaroos kick* . . ., the "Rhythm Rule" is not blocked despite the fact that the two words belong to different 'phonological phrases' as defined in (3). This, we would maintain, is because *kick* is more prominent than *kangaroos* since it is in focus, *kangaroos* being out of focus since they were already mentioned earlier on in

[6]See Rischel 1983:56. See Also Bruce 1983 for a discussion of similar phenomena in Swedish.

the text. In other words, the inherent stress on *kangaroo* is not realized in this instance since no focal prominence (or rhythmic prominence) is assigned to it. This state of affairs can be represented as in (5) (where S and * are to be regarded as abstract units on which focus and rhythmic prominences (x) can be realized). The rhythmic prominence on the first syllable is due to an independent rhythmic constraint of English discussed for example by Schane (1979) and termed the Initial Constraint. This constraint says that a word-initial unstressed (weak) syllable becomes strong (i.e. receives rhythmic prominence) if followed by another weak syllable:

(5)

Kangaroos kick				
S	S	S	S	syllable
		*	*	inherent word stress
			x	focus (cf. Fig. 1, step 3a)
x				Initial Constraint

The failure of the "Rhythm Rule" to apply in . . . *Marcel proved* in (4b) can then be explained as being due to the fact that *Marcel* is more prominent than *proved.* This is in agreement with the hierarchy for assigning focal prominence as presented in Figure 1; that is to say, a noncoreferential subject receives more prominence than a noncoreferential predicate[7].

Thompson (1980) has also demonstrated the effect of focus on the application of the "Rhythm Rule". Although he does not attempt to assign information focus on the basis of grammatical parameters such as we have done (Horne 1985, 1986) and although his notion of focus includes expressive focus as well, the point of his argument is in agreement with the predictions made by our model concerning relative prominence. Thompson (1980:153) presents the data in (6–8) in order to demonstrate the interaction between the "Rhythm Rule" and + FOCUS (/ = his 'foot' boundary):

(6) a. /How do you / get to / work?
b. I / take the / San Mateo / Bridge.

[7]This is somewhat problematic, however, since it is commonly the case that proper names appearing in subject position are construed as 'given' and therefore not focussed (see Horne 1985). It could be that the awareness on the part of the speaker of the French origin of *Marcel* leads him to realize the final stress on *Marcel*, even though the word is construed as given. It certainly is not impossible to pronounce *Marcel* without stress on the final syllabe given a clear case of contextual coreference, as in:

There's the new player, Marcel$_i$ Brown. John said that Marcel$_i$ **fights** a lot.

(7) a. I / understand you / take the / Dumbarton / Bridge to / work.
b. / No I / take the San Mat / eo Bridge.

(8) a. I / understand you / take the / Berkeley / Ferry to / work.
b. / No I / take the / San Mateo / Bridge.

As Thompson notes (1980:153):

"Iambic reversal has applied in (1b) [= 6b], but highlighting [i.e. emphatic or contrastive prominence, MH] and the accompanying +FOCUS have blocked its application and bleached the foot boundary before *Bridge* in (2b) [= 7b]. But in (3b) [= 8b], with two items highlighted, it seems we are back to normal in some sense, with the foot structure being parallel to that in (1b) [= 6b]."

According to our model (see Figure 1), both (6b) and (8b) could be analyzed as in (9a), whereas (7b) would receive an analysis as in (9b):

(9) a.

San Mateo Bridge	
S S SS S	syllable
* * *	inherent word stress
———————	
x	Focus (1a)
x	Initial Constraint

b.

San Mateo Bridge	
S S SS S	syllable
* * *	inherent word stress
———————	
x	Focus (1b)
x	Initial Constraint

On our account, the reason so-called Iambic Reversal (Rhythm Rule) has applied in (6b) = (9a) and (8b) = (9b) is that *Bridge* has more prominence than *Mateo* due to the fact that it is focussed in accordance with Figure 1. Thus the inherent stress on *Mateo* is not realized. In (7b) = (9b), on the other hand, it is *Mateo* that is more prominent than *Bridge, Bridge* being contextually coreferential with a previous mention of the lexical item. Thus the inherent stress on *Mateo* is realized in this case. In both cases, however, we would predict that *San* would receive rhythmic prominence due to the Initial Constraint, i.e. *San's* inherent stress is realized in both instances.

3.3. Nespor and Vogel 1982:Rhythm Rule

Nespor and Vogel (1982) maintain, however, that the notion of the phonological phrase is still needed in English in order to account for what they consider to be a peculiarity of the "Rhythm Rule". It should be pointed out, however, that they define the phonological phrase somewhat differently than Selkirk. According to Nespor and Vogel, phonological phrases are determined in the following way (1982:228–9):

(10) a. Φ constructon (Φ = phonological phrase)
Join into Φ any lexical head (X) with all items on its non recursive[8] side within the maximal projection and with any other non lexical items on the same side (e.g. prepositions, complementizers, conjunctions, copulas . . .).
b. Φ constituency.
Φ branches in the same way as the syntactic trees.

Note that the basic difference between Selkirk's and Nespor and Vogel's definition of the phonological phrase lies in that the latter have stipulated that the specifiers that are joined to the head of a phrase must be on the nonrecursive side of the head. Nespor and Vogel further note that "under certain syntactic conditions", restructuring may optionally apply to create a larger phonological phrase. Just what these conditions are, however, is not specified by them. Optional restructuring is defined as follows:

(11) *Optional Φ Restructuring.*
A non branching Φ which is the first complement of X on its recursive side loses its label and is joined to the Φ containing x under a new node labelled Φ' (Nespor and Vogel 1982:230).

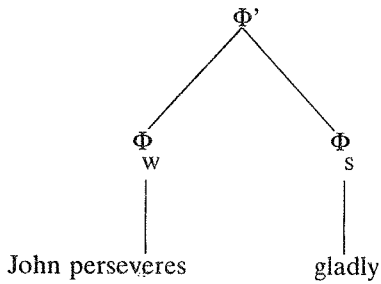
On the basis of the above definitions, Nespor and Vogel claim that the "Rhythm Rule" does not apply when the complement of the word which potentially could undergo the rule is branching. It is maintained, for instance, that there is a difference in the rhythmic structure of *persevere* in (12a) and (12b):

(12) a. John perseveres gladly
b. John perseveres gladly and diligently

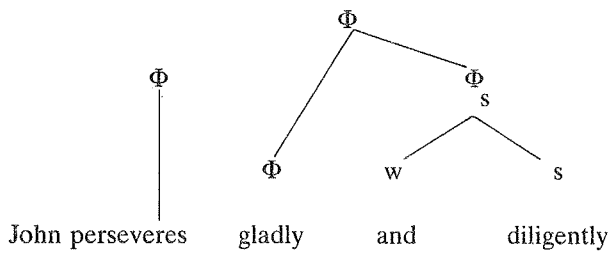
[8]I.e. the side opposite the recursive side. By recursive side they mean the side with respect to the head of syntactic phrase categories a language chooses for its complements (see Nespor & Vogel 1982:239–40).

It is maintained that the "Rhythm Rule" applies in (12a) but not in (12b) where the complement is branching. This is accounted for by assigning (12a) a structure where *persevere* and *gladly* belong to the same Φ, i.e. a 'restructured Φ', whereas in (12b), *gladly* and *diligently* belong to separate Φ's thus blocking the "Rhythm Rule", as in (13):

(13) a.



b.



Personally, I cannot detect any difference in the rhythmic structure of *persevere* in the two cases (nor could several other native speakers of English that I questioned), that is to say, given that the sentences are uttered "out of the blue" with no pause between the verb and the adverb. The stressed syllable *-vere* is definitely not as prominent in this case as it is when the word is cited in isolation or for example in the sentence *He perseveres* where *persevere* has focal prominence. This is what we would expect from our model (see above, Section 2), for in both sentences (13a) and (13b), *perseveres,* the Predicate, receives less prominence in comparison with *gladly,* the Predicate Complement. The prominence on the first syllable, *per-,* can in both cases be attributed to the Initial Constraint discussed above:

(14)

perseveres	gladly	
S SS	S S	syllable
*	*	inherent word stress
	x	focus (1a)
x		focus (3a)
x		Initial Constraint

3.4. Selkirk 1984: Silent Pauses and the Rhythm Rule

In Selkirk (1984) the notion of the phonological phrase is quickly dismissed and a new construct, that of the 'silent pause' is introduced in order to account for the "Rhythm rule" phenomena:

"The term *phonological phrase* has been used to apply to a (putative) level of English prosodic structure falling between the intonational phrase and the prosodic word . . . The English phonological phrase has been thought to have a role in the timing of the utterance, with an influence both on its rhythmic properties and on its division into pauses. We now think that the existence of this unit in English is highly suspect, for syntactic timing (silent pauses in the grid) gives a representation of the disjuncture or separation between syllables that is more appropriate to the description of such rhythmic phenomena. Indeed, we would now explicitly deny that the existence of a level of phonological phrase is well motivated in English (1984:29)."

Like the notion of phonological phrase, however, silent grid positions (silent demibeats) are determined on the basis of syntactic structure:

(15) *Silent Demibeat Addition* (Selkirk 1984:314):
Add a silent demibeat at the end (right extreme) of the metrical grid alligned with
a) a word
b) a word that is the head of a nonadjunct constituent
c) a phrase
d) a daughter of S

The difference between *Marcel Proust* and *Marcel proved* would then be represented in the following way (see Selkirk 1984:184):

(16) a.

```
                x                      x      Pitch Accent
        x       x         x            x      Word Stress
     x  x       x         x   x        x      Heavy/Initial Syll.
     x  x (x)   x   →     x   x (x)    x      Syll./Silent Demi-Beats
   Marcel     Proust    Marcel   Proust
```

b.

```
                x                         x
       x        x               x         x
    x  x        x           x   x    x    x
    x  x (x)xx  x   →       x   x (x)xx   x
   Marcel     proved     Marcel         proved
```

In (16a), *Marcel Proust,* Beat Movement (= Rhythm Rule) applies at level 3 giving more prominence to *Mar.* In (16b) on the other hand, Beat Addition applies to give a (silent) beat on the second metrical level thus undoing the clash and blocking Beat Movement. The basic difference between Selkirk's two analyses, therefore, is that whereas the "Rhythm Rule" is blocked in (b) in the earlier analysis due to the fact that *proved* was contained in a different phonological phrase, in the latest analysis, it is blocked because a new rule Beat Addition inserts a silent beat between the clashing prominences on the next lower level:

(17) Beat Addition (Selkirk 1984:87):

```
                x
          →
       x  x     x  x
```

Beat Additon is: a) left-dominent
b) applies right to left

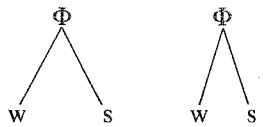
The same criticism can be directed at the grid account as was presented above against the phonological phrase account, i.e. there is no way the grid account can explain the "Rhythm Rule" phenomena in Fig. 5 nor in (16b) in the case where *Marcel* is not focussed due to contextual coreference.

4. Nespor and Vogel 1982: Raddoppiamento Sintattico

Another instance where the 'phonological phrase' has been evoked in order to account for prosodic sandhi phenomena is found in Nespor and Vogel (1982, 1983). They maintain that the rule of Raddoppiamento Sintattico (RS) in Italian is restricted to applying within the phonological phrase. The effect of the rule of Raddoppiamento Sintattico is to lengthen the initial consonant of $word_2$ in a sequence: $word_1$ $word_2$ if the final vowel of $word_1$ is stressed and the affected consonant is followed by $\left(\left\{\begin{matrix}\text{Glide}\\ \text{Liquid}\end{matrix}\right\}\right)$ V.

The process of raddoppiamento sintattico is seen to apply in the following cases (Nespor and Vogel 1982:229):

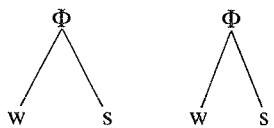
(18) a.



Ho‿mangiato da‿Carla

'I ate at Carla's'

b.



Ho‿creduto che‿venisse

'I believed that he would come'

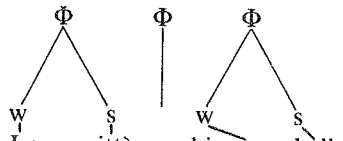
where the application of RS is indicated by '‿'.

What is not explicitly mentioned by Nespor and Vogel but something which can be deduced from the metrical trees is that $word_2$ is more prominent than $word_1$ (cf. Pratelli (1970) where this is explicitly mentioned). In the examples in (18), this can be assumed to be due to phrasal and/or focal prominence. Consequently, such data do not warrant the inclusion of a separate level of structure corresponding to the phonological phrase in the grammar since the phenomena under consideration can be explained by independently motivated prominence relations.

Further data presented by Nespor and Vogel only serve to make this point

all the more obvious. They note, for example, that phonological phrases do not necessarily correspond to constituents, as, for example, in (19):

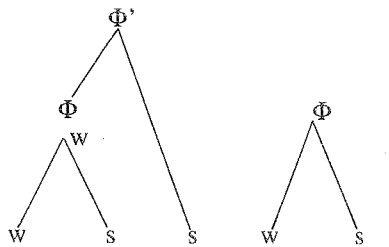
(19)



'Old cities are beautiful'

In this case, of course, no RS can apply as *città* and *veccie* belong to different phonological phrases according to Nespor and Vogel's definition. They note, however, that (19) could also undergo optional restructuring according to (11) and obtain the structure in (20):

(20)



'Old cities are beautiful'

In (20), RS can apply since *città* and *vecchie* now belong to the same phonological phrase. It should also be noted that in this case, *vecchie* would have more prominence than *città* since its bears the phrase (or focal) stress. A structure like that in (19) *would* be needed, however, by Nespor and Vogel in order to account for cases like those discussed by Pratelli (1970:48) involving 'contrastive' focus:

(21) Ci sont paesini morti? Anzi, *città* morte.
'Are they dead villages? On the contrary, dead *cities*'

In this case, *città* has, of course, more prominence than *morte*. This, in itself, can explain the lack of RS here as is no doubt also the case in (19), i.e. it should no doubt be translated as: 'Old *cities* are beautiful'. It seems quite unrealistic to attribute the nonapplication of the sandhi rule to membership in *ad hoc* phonological phrases when the explanation is directly available using focus information.

5. Monosyllabic Destressing

Another phenomenon that would appear to be sensitive to relative levels of prominence is that surrounding what has been termed Monosyllabic Destressing.

Monosyllabic Destressing is the process assumed to account for the weak forms of function words such as in (22a):

(22)

	a.	b.
can see	[kən, kn̩]	[kæn]
for many	[fər]	[fɔr]
in secret	[ən, n̩]	[ɪn]

In her early analysis, Selkirk (1972) assumed that Monosyllabic Destressing, like the "Rhythm Rule", was restricted to applying within the phonological phrase as defined in (3). In her latest (1984) analysis, however, a number of changes have been made in the treatment of monosyllables due partially to the fact that in the early treatment, focal prominence was, for example, totally ignored:

"We should point out here that in Selkirk 1972 there was no understanding of the role of pitch accents (or of intonation in general) in determining levels of stress, nor was there any understandning of the relation of those pitch/rhythm prominences to the focus properties of the sentence (Selkirk 1984:361)".

With respect to the prosodic representation of function words, Selkirk (1984:336–7) assumes that they are distinguished from "real" words as far as grid construction goes by:

"a) in their not being attributed the following silent demibeat [≈word boundary, MH] that their word status would merit them and

b) in their not being attributed a third-level "main word stress".

Consequently, the auxiliary verb, *can,* for example, would have the underlying prosodic representation as in (23a). Monosyllabic Destressing functions then to remove the second-level beat as in (23b). (The rule is assumed by Selkirk (1984:339) to apply cyclically on the phrasal domain and to precede application of Beat Movement on that domain.):

(23) a.

```
        x
    x   x
    x   x
  can fish
```

b.
x
x
x x
can fish

Selkirk (1984:336–72) then discusses three main environments where Monosyllabic Destressing is blocked:
(A) When the monosyllable is focussed as in (24a)
(B) When the monosyllable itself is part of a compound verb as in (24b)
(C) When the monosyllable is at the right end of a phrase as in (24c):

(24)
a. I don't recall the title but I SHOULD remember it.
b. They boxed in the crowd = They boxed the crowd in (vs. They boxed in the crowd (and not in the street).
c. i. What were you thinking *of* last night?
ii. She's not much taller than I *am*.

The above three cases are explained by Selkirk in the following way:
(A') As regards focussed monosyllables, they do not destress because they have prominence on the third level of the grid (Selkirk 1984:361).
(B') With respect to monosyllables that form part of compound verbs, these could be assumed to have lexical word stress and thus not eligible for destressing.
(C') Regarding monosyllables at the right end of a phrase, these are assumed not to destress due to the fact that a rule called Silent Demibeat Addition (cf.15) applies, introducing a silent demibeat on a higher phrasal domain as in (25) (see Selkirk 1984:368):

(25)
$$\begin{matrix} x \\ x \end{matrix} \quad {}_{\alpha}[\ldots {}_{fw}[\sigma]_{fw}]_{\alpha} \rightarrow {}_{\alpha}[\ldots {}_{fw}[\sigma]_{fw}]_{\alpha} \quad \begin{matrix} x \\ xx \end{matrix}$$

where fw = function word

Accordingly, another principle, the Grid Culmination Principle, accounts for why Monosyllabic Destressing does not apply:

(26) *Grid Culmination Principle*
A basic beat that is culminating may not be deleted (a beat *culminates* the metrical grid in d_i if it is alone on its metrical grid in d_i)

that is to say, the addition of a silent demibeat at the end of the phrase blocks destressing due to the Grid Culmination Condition.

As can be seen, the added silent demibeat serves the same function as the phonological phrase boundary in delimiting the domain of Monosyllabic Destressing. The Grid Culmination Principle, however, seems to us to offer a much less satisfactory explanation as to why the process is blocked than the explanation presented in Selkirk 1972. There, the rule is formulated as follows:

(27) $V \rightarrow [-\text{stress}]/\#W[C_o\text{-}C_o]\ [\ (\#)\ X\acute{V}Y(\tilde{\#})]\ Z\#]$
Condition $X \neq T$ ## U
where [(#)XV́Y(#)] is the affected monosyllable's head or a codependent i.e. "A monosyllable loses its stress when it precedes its head or a codependent in surface structure" (Selkirk 1972:31)

We are inclined to think that the key to understanding the phenomenon termed Monosyllabic Destressing is the stressed V in this rule. That is to say, the monsyllable must be followed by a more prominent element. This would account for all the nonreduced forms in (24). We feel that the analysis of monosyllables that Selkirk presents is overly complicated. There is no need to treat monosyllables differently from other lexical items as far as stress is concerned. They can be assigned inherent stress like all other words. The reason their vowel is reduced (or 'destresses') is, on our account due to the fact that this inherent stress is in general not realized due to cliticization. In the examples in (24), the inherent stress *is* realized due to focus or absence of cliticization.

It would thus appear that both Monosyllabic Destressing and the "Rhythm Rule" require the same prosodic conditions. This is not surprising since both processes can be regarded as belonging to the class of eurythmic phenomena. The domain for Monosyllabic Destressing can be considered to be that occupied by a given grammatical function within a clause, i.e. the same domain as the focus assignment rule given above in Figure 1.

6. Conclusion

The goal of this article was to examine the justification for setting up a special prosodic category, phonological phrase in a grammar. We hope to have shown that this is unnecessary in a phonology that has direct recourse to focal structure information. The same redundant status, we feel can be

assigned to the concept 'silent demibeat', at least to the extent that it is used to block the processes examined in this article. The phenomena associated with the Rhythm Rule, Monosyllabic Destressing and Raddoppiamento Sintattico can instead be seen to be associated with relative levels of prominence arising from the realization of abstract word stresses after application of (in all cases examined here) focus assignment rules.

University of Lund

References

BRUCE, G. 1977. *Swedish Word Accents in Sentence Perspective.* Lund: Gleerups.

——— 1983. On rhythmic alternation. *Working Papers, Lund University, Dept. of Ling.* 25. 32–52.

CHOMSKY, N. 1965. *Aspects of the Theory of Syntax.* Cambridge, Mass.: MIT Press.

CHOMSKY, N. and HALLE, M. 1968. *The Sound Pattern of English.* New York: Harper & Row.

CRYSTAL, D. 1969. *Prosodic Systems and Intonation in English.* Cambridge: Cambridge University Press.

GÅRDING, E. 1983. A generative model of intonation. In A. Cutler and D. Ladd (eds.) *Prosody: Models and Measurements.* 11–25. Berlin: Springer.

GÅRDING, E. and HOUSE, D. 1985. Frasintonation, särskilt i svenska. *Svenskans beskrivning* 15:205–21.

HORNE, M. 1985. English sentence stress, grammatical functions and contextual coreference. *Studia Linguistica* 39:51–66.

——— 1986. Information focus: assignment and phonological implications. In L. Evensen (ed.) *Nordic Research in Text Linguistics and Discourse Analysis.* 155–70. U. of Trondheim: TAPIR.

NESPOR, M. and VOGEL, I. 1982. Prosodic domains of external sandhi rules. In H. van der Hulst and N. Smith (eds.) *The Structure of Phonological Representations (Part I).* 225–55. Dordrecht: Foris.

——— 1983. Prosodic structure above the word. In A. Cutler and D. Ladd (eds.) *Prosody: Models and Measurements,* 123–40. Berlin: Springer.

PIERREHUMBERT, J. 1981. Synthesizing intonation. *J. Acoust. Soc. Am.* 70:985–95.

PRATELLI, R. 1970. Le renforcement syntaxique des consonnes en italien. *La linguistique* 6:39–50.

RISCHEL, J. 1983. On unit accentuation in Danish – and the distinction between deep and surface phonology. *Folia Linguistica* 17:51–97.

SCHANE, S. 1979. The rhythmic nature of English word accentuation. *Language* 55:559–602.

SELKIRK, E. 1972. *The Phrase Phonology of English and French.* Ph.D. Diss. MIT. Reproduced by Indiana University Linguistics Club (IULC) 1981.

——— 1980. On prosodic structure and its relation to syntactic strucutre. IULC.

——— 1984. *Phonology and Syntax: The Relation Between Sound and Structure.* Cambridge, Mass.: MIT Press.

SIGURD, B. 1984. Computer simulation of spontaneous speech production. In *Proceedings of COLING 1984,* 79–83. Assoc. for Computational Linguistics.
SVARTVIK, J. 1982. The segmentation of impromptu speech. In N. Enkvist (ed.) *Impromptu Speech: a Symposium.* 131–45. Åbo: Åbo Akademi.
THOMPSON, H. 1980. *Stress and Salience in English: Theory and Practice.* Ph.D. Thesis, Berkeley.

Charles F. Meyer

GRAMMATICAL AND PRAGMATIC EFFECTS ON EMPATHY CONSTRAINTS

Introduction

Much recent research has demonstrated that many linguistic phenomena cannot be accounted for unless one considers the interaction of grammatical rules (i.e. rules of syntax, semantics, and phonology) and pragmatic principles.[1] The study of this interaction has led to insights into all levels of linguistic structure, including discourse structure (Stubbs 1983 and Beaugrande and Dressler 1981), thematic structure (Halliday 1967–8 and Quirk et al. 1985:1360–77), semantic structure (Leech 1981:319–41 and Lyons 1977:725–86), syntactic structure (Lakoff 1974 and Green 1974), prosodic structure (Quirk et al. 1985:1355–60 and Schmerling 1976), and morphological structure (Gazdar 1980:60–3 and Lambert and Tucker 1976). In short, these (and other) studies have illustrated that complete "explanations" about language arise only when one explores all of the grammatical and pragmatic influences that shape language.

I wish to argue in this paper that this view of language can explain the idiosyncratic nature of the empathy constraints proposed in Kuno and Kaburaki (1977, hereafter K&K); that is to say, that the selective application of these constraints can be accounted for if one investigates the grammatical rules and other pragmatic principles that these constraints interact with. Specifically, I will demonstrate that (1) no empathy constraints override obligatory rules of the grammar that they interact with; (2) all empathy constraints override (in varying degrees) optional rules of grammar that they interact with; and (3) empathy constraints apply inconsistently when they interact with other pragmatic principles. That is to say, some empathy constraints override other pragmatic principles that they interact with; other empathy constraints, on the other hand, do not.

Because many of the claims made in this paper involve acceptability judgments, most of the data has been evaluated by groups of subjects.[2]

[1]I wish to thank Sidney Greenbaum, Frank Parker, and Kathy Riley for many helpful comments on an earlier draft of this paper. Of course I take responsibility for any remaining errors.

[2]All subjects were native speakers of English, and most had spent the majority of their lives in

IV

TOWARDS A QUANTIFIED, FOCUS-BASED MODEL FOR SYNTHESIZING SENTENCE INTONATION IN ENGLISH

Merle Horne

Abstract

An algorithm for assigning information focus within an English text (developed elsewhwere) on the basis of an interaction of grammatical functions and contextual coreferential relationships is phonetically quantified with respect to the parameter of pitch (F_0) and situated within a more embracing model of sentence prosody. The model is readily adaptable for implementation in a text-to-speech program.

The algorithm for assigning focal prominences serves as a basis for accounting for English sentence intonation. Levels of focal prominence are defined within an empirically determined sloping grid consisting of two parallel lines representing the direction and scope of a given speaker's nonemphatic declarative sentence intonation. An informal experiment based on analysis by synthesis is used to test the focus assigning model. The placement of prefocal phrasal prominences within the grid is also discussed and situated in the rule system of the prosody model. The resultant rules are then applied on a fragment of discourse. Derivations and synthesized F_0 curves are presented and discussed.

Introduction

Within recent years, there has been a considerable amount of research done in developing models for describing and synthesizing prosodic features (e.g. Bruce 1977, 1982; Bruce & Gårding 1978; Gårding 1977,1981,1983; Fujisaki and Hirose 1982; Ladd 1983; Olive and Liberman 1979, Pierrehumbert 1981; Sigurd 1984; Thorsen 1980). Some of these models have even been implemented in text-to-speech systems. None of them, however, includes in its phonological component rules for assigning prosodic prominences based on information focus, i.e. textually and grammatically conditioned focus. Rather, existing systems usually treat each sentence in isolation without regard to what information has been presented in earlier sentences and assign prominence on the basis of, for example, lexical categories (N, V, Adj), and/or rhythmical principles. Focus, to the extent that it is considered, is marked in each individual sentence by the analyser at the time of synthesis . The inclusion of a parameter of focus is, however, crucial for the optimal functioning of a text-to-speech system. The different mechanisms used to highlight new information as well as those used to refer to given information must be taken into consideration when writing rule systems for automatic speech processing. The aim of this paper is to propose how a phonological component including rules for assigning focal prominences could be implemented in a text-to-speech program.

In Horne 1985, 1986a,b, a model was developed for assigning information focus (i.e. grammatically and contextually conditioned focus). The output of this model is

a phonological representation where three different levels of focal prominence have been assigned to stressed syllables. Just how this type of representation could then be phonetically quantified will be developed below after a brief summary of the model.

Outline of Model for Assigning Information Focus

According to the model for assigning information focus (Figure 1) presented in Horne 1986b, focal prominence patterning in English can be accounted for on the basis of a hierarchy of grammatical functions interacting with contextual coreference relationships (cover term for coreference as well as identity of sense relationships such as synonomy, hyponomy, part-whole relationships). This model assumes, furthermore, that there are three degrees of focal prominence, corresponding to the three basic constituents of functional or logical structure: subject, predicate, predicate complement (a cover-term for object and VP (non-frontable) adverbials). Moreover, these grammatical functions are regarded as being hierarchically ordered, so that in an 'all new' SVO sentence, the predicate complement^{1} receives more prominence than the subject^{2} which in turn receives more prominence than the predicate^{3}. All these relations between grammatical functions are reflected in the flow-diagram in Figure 1. That is to say, the predicate complement in an 'all new' sentence receives more prominence than the subject, but in an intransitive sentence, the subject receives just as much prominence as the predicate complement in an SVO sentence. Note, furthermore, that the modifier in a head-modifier construction realizing a given grammatical function will

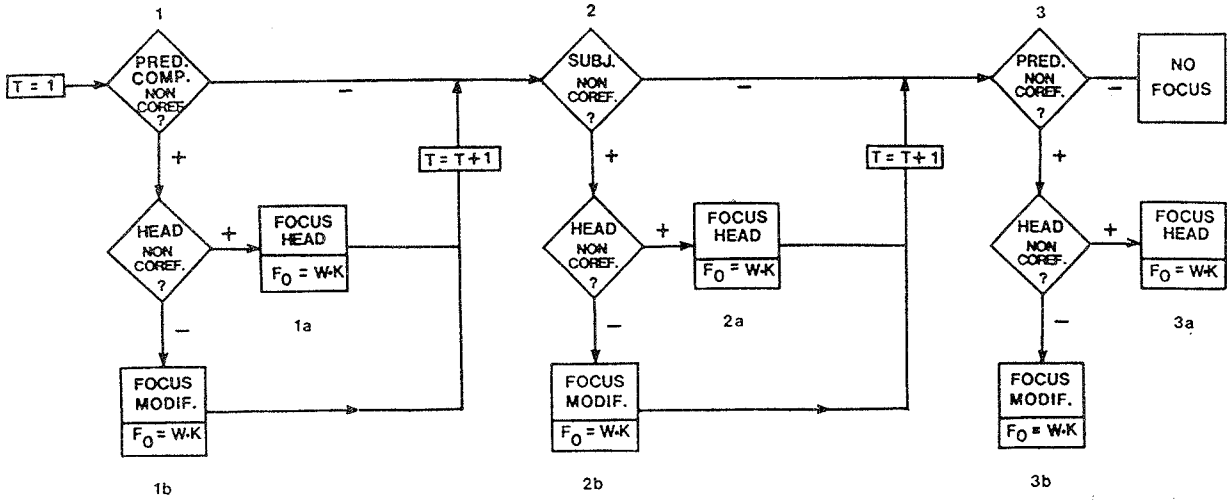


FIGURE 1. MODEL (FLOWCHART) FOR ASSIGNING INFORMATION FOCUS TO CONSTITUENTS ON THE BASIS OF GRAMMATICAL FUNCTIONS AND THE COREFERENTIAL STATUS OF THE LEXICAL MATERIAL REALIZING A PARTICULAR FUNCTION. THE INPUT TO THE MODEL IS A GIVEN CLAUSE (S). FOCUS IS REALIZED AS PITCH (F_0) ACCORDING TO THE EQUATION $F_0 = W \cdot K$ WHERE F_0 HERE REFERS TO THE RELATIVE HEIGHT OF A GIVEN PITCH OBTRUSION, W DESIGNATES THE WIDTH OF THE GRID WITHIN WHICH F_0 MOVES AND K IS A VARIABLE RANGING OVER A NUMBER OF PROMINENCE LEVELS DEFINED AS FRACTIONS OF THE DISTANCE FROM THE BASELINE TO THE TOPLINE OF THE GRID. IN FIG. 3, K ASSUMES THE VALUES 1 (FOR THE FIRST FOCUSSED CONSTITUENT), 0.8 (FOR THE SECOND FOCUSSED CONSTITUENT), AND 0.4 (FOR THE THIRD FOCUSSED CONSTITUENT). FOR THE SYNTHESES DONE IN THE PRESENT WORK, HOWEVER, THE VALUES USED WERE 1, 0.75, AND 0.5, RESPECTIVELY. THE DIAGRAM IS TO BE READ AS FOLLOWS: 1.: CHECK TO SEE IF THERE IS A PREDICATE COMPLEMENT THAT IS NON-COREFERENTIAL WITH SOMETHING IN THE PRECEDING PART OF THE TEXT. IF THERE IS ONE, CHECK AND SEE IF IT IS THE HEAD THAT IS NON-COREFERENTIAL. IF THIS CONDITION IS MET, FOCUS THE HEAD, ASSIGNING IT A LEVEL OF PROMINENCE WHERE $F_0 = W \cdot K$ (1a). IF THE HEAD IS COREFERENTIAL, ASSIGN THE MODIFIER FOCAL PROMINENCE INSTEAD (1b). GO TO SUBJECT (2) AND REPEAT THE SAME ROUTINE, AND THEN GO TO PREDICATE (3), AGAIN REPEATING THE SAME ROUTINE.

receive an amount of prominence equal to that of the head should the head be contextually coreferential with something in the preceding part of a given discourse.

The input to the model for assigning focal prominence is a syntactico-semantic representation generated by a computer-based referent grammar such as that developed by Sigurd 1987. Such a representation contains all the information needed by the model to assign focal prominence. For example, the last sentence in (1), analysed in Horne 1986, would, in addition to information about mode, have a representation such as that presented in (2):

(1) A: I'm just about finished writing my new book

B: Oh, do you think you could let me in on how it's going to end?

A: Yea, sure. A mormon will marry a mayor.

(2) s(subj(np(nr4,nom(mormon,sg,indef))),
pred(v(vr6,nom(marry,fut))),
obj(np(nr5,nom(mayor,sg,indef)))))

where nr4, nr5 are nominal referents and vr6 is a verbal referent. The existence of these referents is of crucial importance for the functioning of the focus assigning model. Figure 2a, for example, shows the phonetic realizationof F_0 when none of the referents have been mentioned in the preceding context, as in (1); in this case, all the lexical heads receive some F_0 prominence according to the model in Figure (1). On the other hand, consider the context in (3); here, both the predicate and the object in the last sentence,

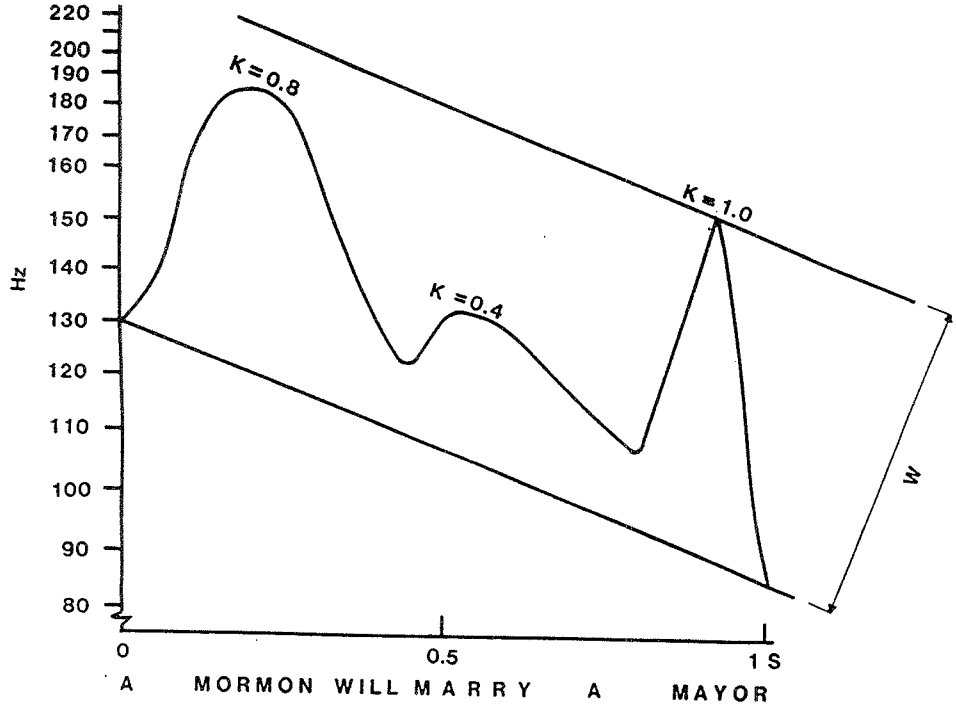


FIGURE 2a. ACTUALLY OCCURRING F_0 CURVE OBTAINED FOR A READING OF THE LAST SENTENCE IN (1) WHERE THE SUBJECT, PREDICATE AND PREDICATE COMPLEMENT ARE FOCUSSED ACCORDING TO THE MODEL IN FIGURE 1.

identical to those in (2) are contextually coreferent with previously mentioned lexical material. They consequently receive no focal prominence and the F_0 curve instead assumes a shape like that shown in Figure 2b (identical subscripts designate coreferential expressions):

(3) A: My new book is about a mayor$_i$ living in Malmö. He meets an interesting person there and gets married$_j$.

B: Oh, could you let me in on who marries$_j$ him$_i$?

A: Yea, sure. A Mormon will marry$_j$ the mayor$_i$.

Phonetic Quantification of the Model

The model described above constitutes a focus component which generates a phonological representation where levels of focal prominence are indicated. Just how this representation could be taken by the phonetic component and used in rules to generate an appropriate F_0 curve will be discussed in the present section.

In attempting to parameterize the output of the focus component (Figure 1), we have adopted, with some modification, the basic framework of the Lund model for prosody described for example in Bruce 1977, Bruce and Gårding 1978, Gårding 1981. This model was developed originally to analyze Swedish intonation, but is readily adaptable for describing the prosody of other languages (see Lindau 1986, Gårding 1981). The Lund model is designed to account for durational aspects of prosody as well, but in the present work, we will be concerned exclusively with the design of an algorithm for

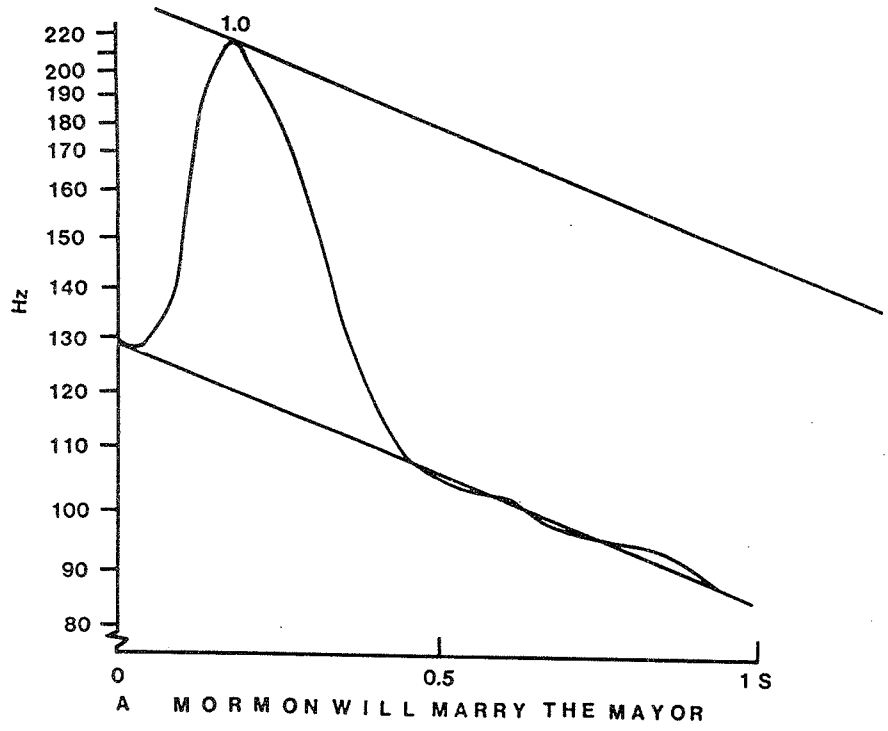
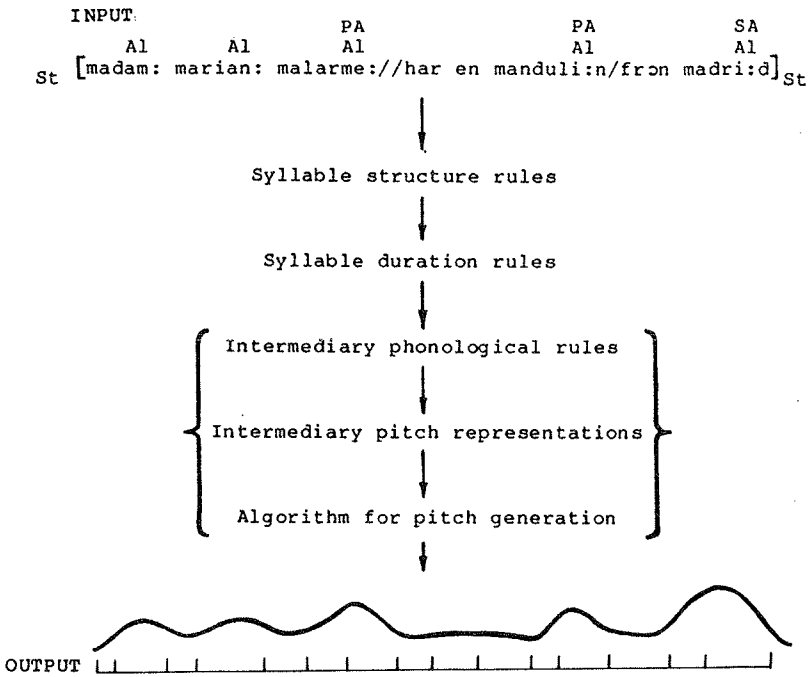


FIGURE 2b. ACTUALLY OCCURRING F_0 CURVE OBTAINED FOR A READING OF THE LAST SENTENCE IN (3) WHERE ONLY THE SUBJECT IS FOCUSSED ACCORDING TO THE MODEL IN FIGURE 1.

generating pitch contours in English. Figure 3, from Gårding 1981, shows the main components of the Lund model for prosody. We have enclosed in braces that part of the model that the present article intends to develop.

Defining the phonological grid

In Horne 1986b, preliminary values for the three levels of focal prominence were presented. They were based on measurements from actually occurring F_0 contours collected from one speaker of English, an American male. These values were specified as fractions of the distance from the baseline to the topline of a phonological 'grid', over-all contour lines within which a given sentence's intonation can be described (see Gårding 1981). This grid was drawn so that the baseline extended between the normal starting point (on an unstressed syllable) and end F_0 levels for this speaker. (See Figure 2a). In uttering this particular sentence, the speaker started at 130 Hz and ended at a level of 90 Hz. We joined these two points and the resulting line served as the baseline of the phonological grid for a declarative sentence. The topline of the grid was drawn parallel to the baseline so that it passed through the peak of the highest pitch obtrusion. With respect to the width of the grid, it was then observed that in relation to the height of the peak on the Object (set at 1.0 =100% of the width (W) of the grid), the Subject peak reached 0.8 of the distance from the baseline to the topline, and the Predicate, 0.4 of this same distance (see Pierrehumbert 1981 for a similar way of describing F_0 contours). These fractions were measured by hand using a ruler.



where Al = Accent 1 in Swedish (language specific)

PA = Phrase Accent

SA = Sentence Accent (our highest degree of focal prominence)

Figure 3. Lund model of prosody (from Gårding 1981)

The F_0 scale used in the analysis was logarithmic. It has been assumed that this scale corresponds better to the way speakers perceive F_0 than a linear scale (see Cohen et al. 1982:264). For the analyses done in preparing this article, however, we were obliged to use a linear scale, which is that available for pitch editing in the ILS program package at the Dept. of Linguistics, Univ. of Lund. We decided, however, to work within the range 90 - 180 Hz so that the relationships between levels of prominence expressed using the linear scale would be compatible with those using a semitone scale (see below, Figure 5 where we have compared the output of a given synthesis using the two different scales).

Generating pitch contours by the focus assigning model--an informal experiment

In order to arrive at appropriate values of focal prominence for plugging into the phonological representations, we decided to experiment with an arbitrary sentence consisting of exclusively sonorant sounds so as to obtain an unbroken F_0 curve :

(4) A young man will allay an ill lion

The sentence was recorded by the same American. We then began to edit the pitch contour of this sentence using the program mentioned above, leaving the segmental content undisturbed. Stylized F_0 curves composed of straight lines were used in the syntheses (cf. t'Hart 1982).

Grid. As in Figure 2a, we defined a baseline corresponding to

beginning and end F_0 points characteristic for this speaker (130 Hz, 90 Hz, respectively). The pitch range was set at 1 octave, the low point being 90 Hz and the high point, 180 Hz.; the topline of the grid was then drawn parallel with the baseline as before. This grid was then assumed to represent the speaker's non-emphatic F_0 range for a given declarative sentence. The relative degrees of prominence given in Figure 2a were then arbitrarily rounded off so that the predicate was assigned a level 50% of the way from the baseline to the topline, the subject, a level 75% of this distance, and the predicate complement, 100% of this distance in an all new sentence. Thus the abstract grid for a declarative sentence uttered by this particular speaker was defined as in Figure 4 (see Huber 1985 for an alternative way of interpreting the grid for Swedish).

Baseline vs. topline. In order to synthesize new pitch contours for this sentence, it was decided to first of all attribute a phonetic reality to the baseline. That is to say, we decided that this baseline would be realized phonetically over stretches of nonfocussed material. The topline, however, is not ascribed any phonetic reality; it functions solely as a reference line for computing F_0 obtrusion levels.

Analysis by synthesis. a) Sentences with an early focal prominence. Figure 5 shows the F_0 curve synthesized in the case where the sentence in (4) is assigned an all new reading (we have here represented the result of the synthesis using both a linear and a semitone scale for sake of comparison; as can be seen, the prominence relations, described as fractions

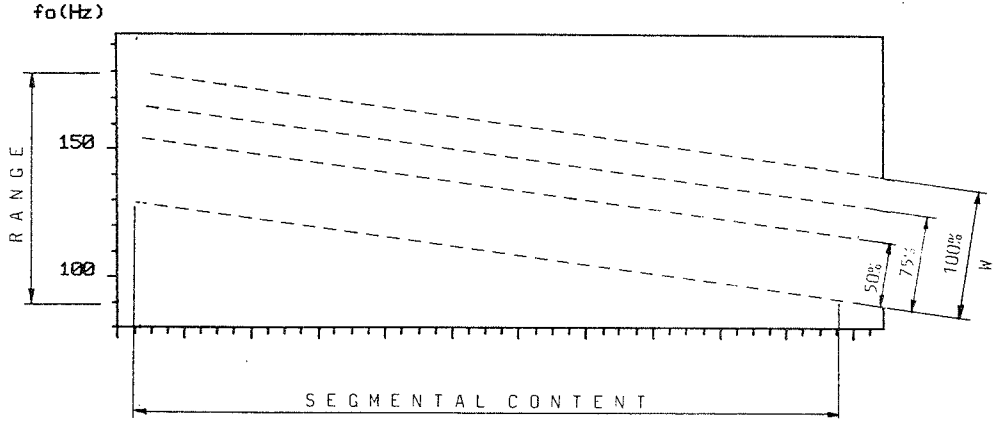


FIGURE 4. PHONOLOGICAL GRID USED FOR SYNTHESIZING F_0. THE F_0 RANGE EXTENDED BETWEEN 90 AND 180 HZ. THE BEGINNING AND END POINTS FOR A GIVEN SENTENCE WERE SET AT 130 HZ AND 90 HZ, RESPECTIVELY. ACCORDING TO FIGURE 1, THE FIRST FOCUSSED CONSTITUENT RECEIVES A LEVEL OF PROMINENCE EQUAL TO 'W', THE SECOND, A LEVEL OF PROMINENCE EQUAL TO .75W, AND THE THIRD, A PROMINENCE LEVEL EQUAL TO .50W.

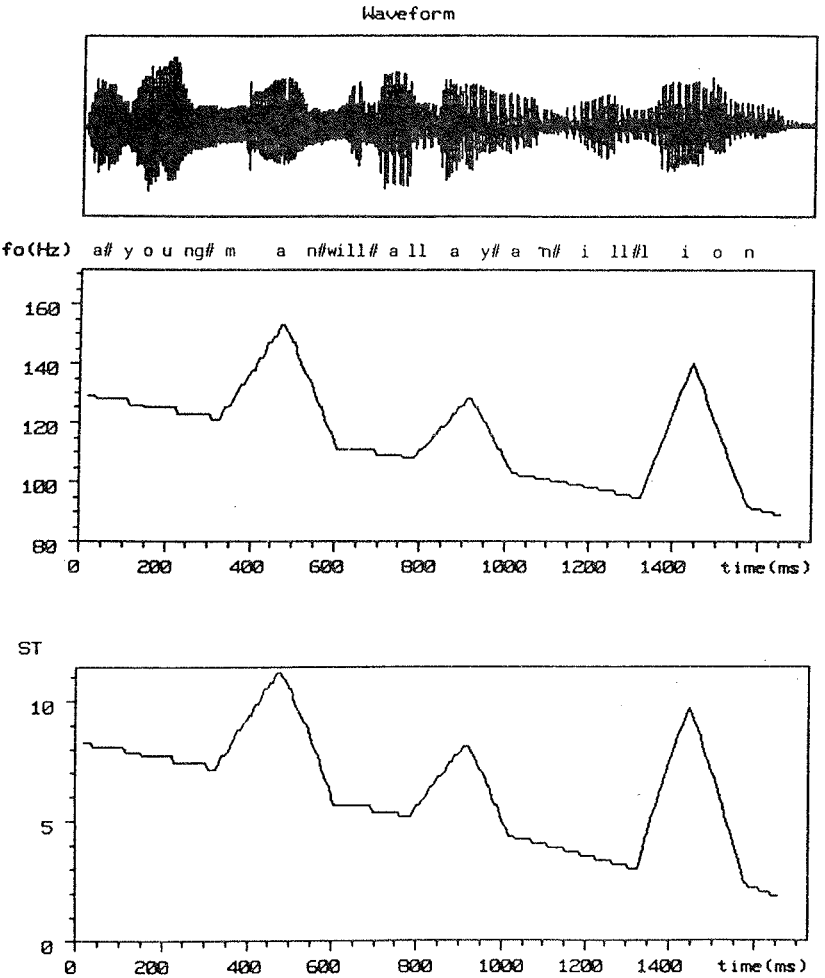


FIGURE 5. SYNTHESIZED F_0 CURVE OF SENTENCE 4 WITH FOCUS ON SUBJECT, PREDICATE, AND PREDICATE COMPLEMENT ACCORDING TO FIGURE 1. FOR SAKE OF COMPARISON, THE SYNTHESIS IS REPRESENTED USING BOTH A LINEAR SCALE (UPPER CURVE) AND A SEMITONE SCALE (LOWER CURVE). NOTE THAT THE RELATIVE PITCH LEVELS ARE ALMOST IDENTICAL IN THE TWO CASES.

of the distance from the baseline to the topline, are almost identical in this F_0 range). According to the focus assigning model in Figure 1, the object, 'lion', was assigned a pitch obtrusion extending from the baseline to the topline, the subject, an obtrusion reaching 75% of the way from the baseline to the topline, and the predicate, an obtrusion extending over 50% of this distance. The span of the obtrusion was the 'underlying' stressed syllable, with the peak coming towards the end of the vowel. This synthesis sounded quite acceptable. We then proceeded to synthesize contours corresponding to other potential outputs of the focus assigning component. Figure 6 shows that derived when the subject and predicate would be focussed, for example, when the sentence functions as the answer to a hypothetical question such as "What will happen to an ill lion?". Figure 7 displays the synthesis of the F_0 contour when only the subject is focussed, as for instance when the sentence is uttered as a response to the question "Who will allay an ill lion?". Both these syntheses also sounded very good.

b) Sentences with a late focal prominence. A poor result arose, however, when we synthesized the contour displayed in Figure 8, i.e. the predicted output of the focus assigning model when only the object is focussed. The long flat stretch before the late pitch obtrusion sounded very artificial. It is, in fact the case in naturally occurring speech that we rarely find a nondisturbed F_0 curve before focus. After focus, however, it is natural to find F_0 corresponding with the baseline. However, we were assuming at this point that the only perceptually important F_0 obtrusions would be those

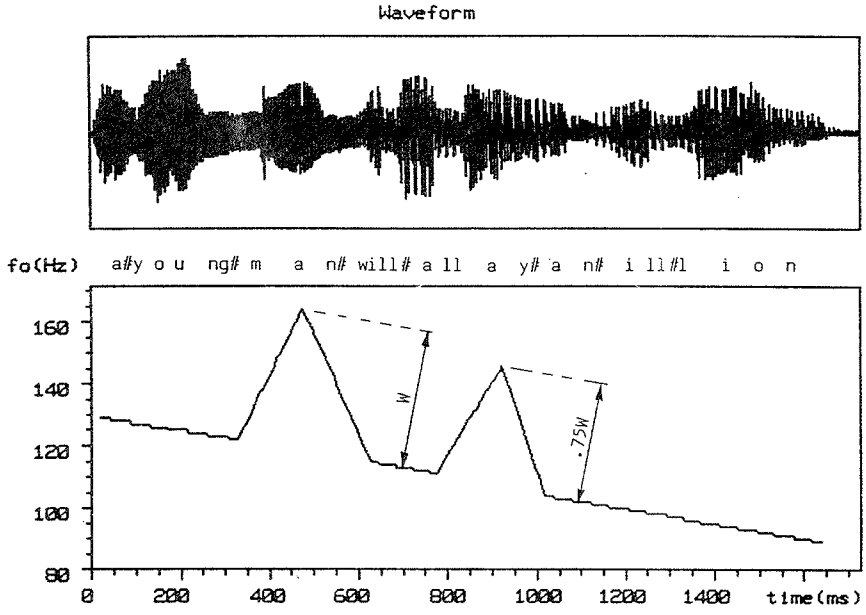


FIGURE 6. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE SUBJECT AND PREDICATE

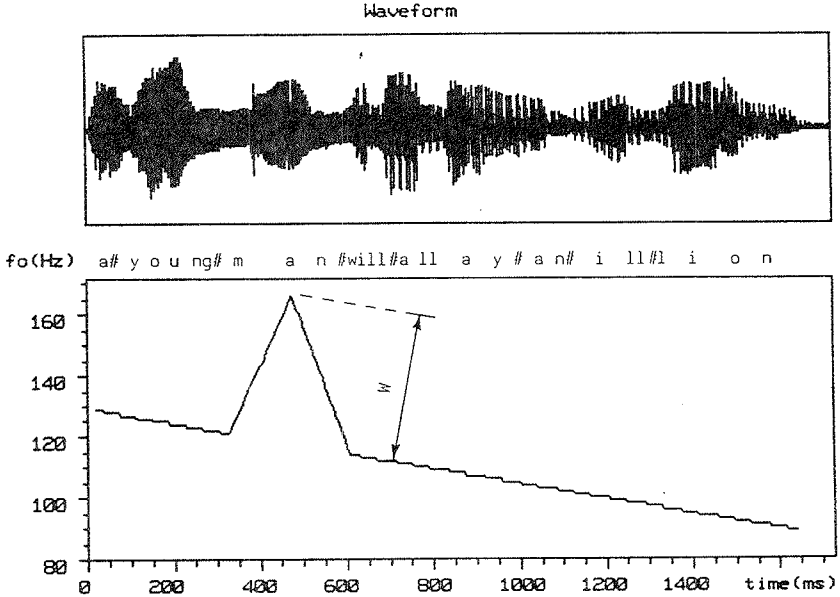


FIGURE 7. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE SUBJECT

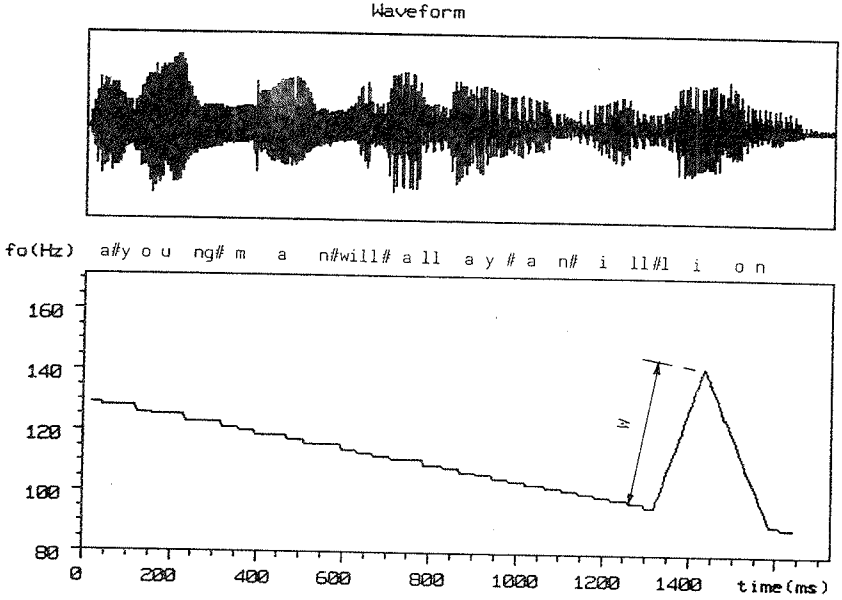


FIGURE 8. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE OBJECT

associated with focus, i.e., we were taking the strong position that prominences associated with other grammatical features, for example, phrase boundaries, would, if perceptually important, be sufficiently signalled by other phonetic parameters, for instance, duration.

Continuing along this line of reasoning, we first hypothesized that perhaps the starting point was too high, i.e., that the declination was too extreme for there just being one focussed constituent in the sentence and that the starting point was perhaps determined by the number of focussed constituents, say 10 Hz for each focussed constituent. Consequently, we lowered the starting point to 110 Hz instead of 130 Hz and resynthesized the curve but the output still sounded peculiar. Another unacceptable output was obtained when we kept the starting point at 130 Hz, rose on the subject to a height of 25% from the baseline and then continued with a very slight declination to the focal object, following Ladd's (1986) "overall contour shape" approach (see Figure 9). Again, the long stretch without any F_0 movement sounded unnatural. It was subsequently hypothesized (Thore Pettersson, personal communication) that what was needed in this deviant case was an early peak or peaks that would function as reference points for the late focal obtrusion. As mentioned above, such prefocal F_0 disturbances are what are commonly observed in real language data when focal accents come relatively late in an utterance, in contrast to what happens when a focal accent comes early in the utterance (cf. Figure 7); in such cases, F_0 is flat on the baseline after the pitch obtrusion (see Eady et al. 1986 for experimental support for the existence of prefocal "anticipatory" F_0 movements).

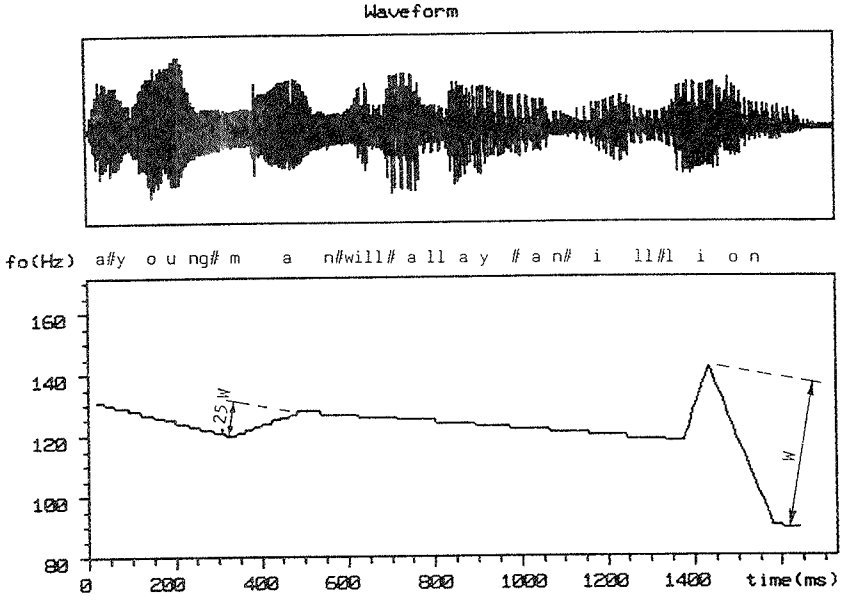


FIGURE 9. F_0 CURVE SYNTHESIZED ACCORDING TO LADD'S (1986) "OVERALL CONTOUR SHAPE APPROACH" WITH FOCUS ONLY ON THE OBJECT OF SENTENCE (4)

We subsequently decided to experiment and add F_0 obtrusions extending 25% of the way from the baseline to the topline of the grid on all lexical ('content') words (see Figure 10). This solution, however, sounded more Swedish than English; there were just too many pitch movements to be acceptable. Finally, we synthesized a version with prefocal obtrusions only on the lexical heads and this produced a very good result (see Figure 11). In subsequent syntheses, we consistently added these prefocal pitch obtrusions on lexical heads. Figure 12, for example, displays the synthesis of the same sentence with focus on the subject and object, a contour that would be generated when the sentence functions for instance as an answer to a question such as "Who will allay what?".

c) Phrase accents. The finding concerning these additional pitch movements led us to include a Phrase component in our description that would automatically assign 25% prominence to all lexical heads (see flow diagram in Figure 13). Among the Intermediary Phonological Rules in Figure 3, moreover, would then be the one which would delete all phrase accents after the last focal accent in a given (component) sentence (see Gårding 1981:152). (The environment for this rule would appear not to be the full sentence. We synthesized a version of sentence (5d) (see below) leaving a phrase accent on <u>money</u> in the first component sentence of this compound sentence and it sounded inferior to the version without this accent (see Figure 17)).

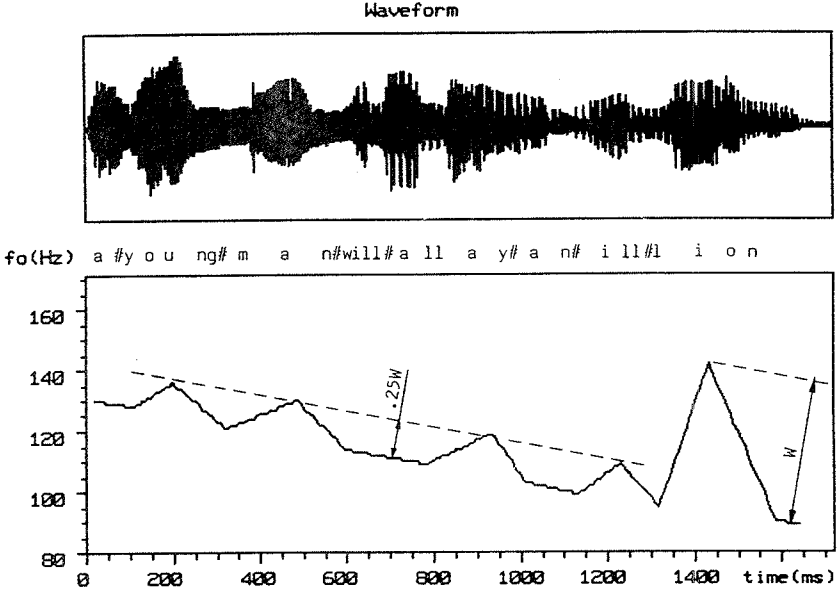


FIGURE 10. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE OBJECT AND PHRASE ACCENTS ON ALL PREFOCAL LEXICAL WORDS

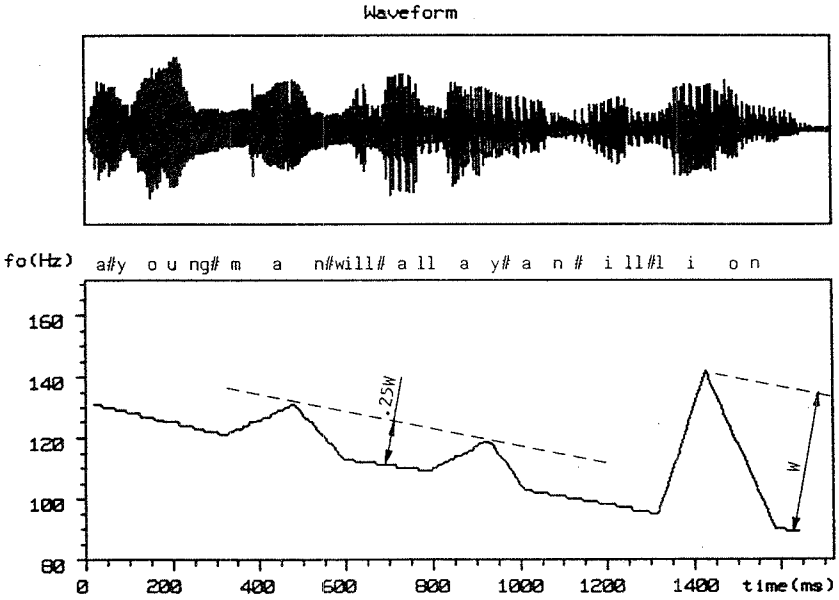


FIGURE 11. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE OBJECT AND PHRASE ACCENTS ON ALL PREFOCAL LEXICAL HEADS

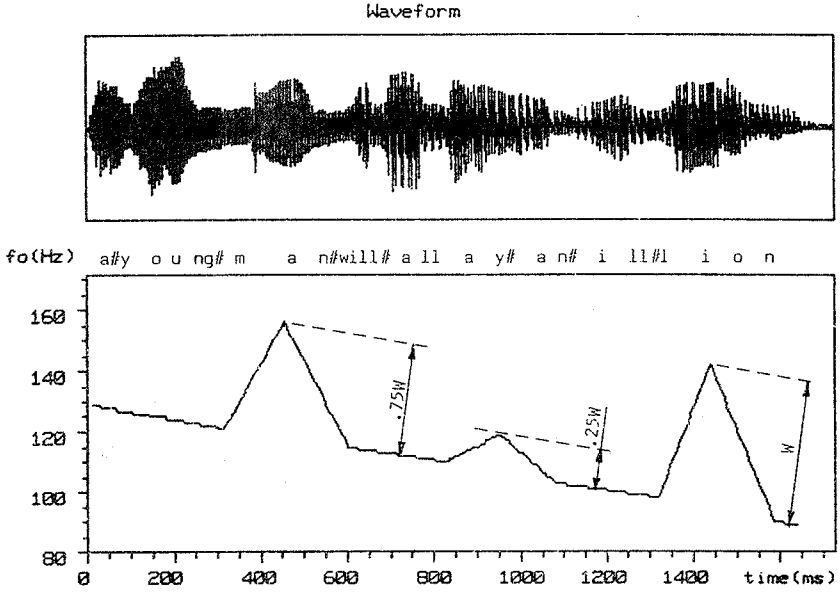


FIGURE 12. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE SUBJECT AND OBJECT AND PHRASE ACCENT ON PREFOCAL VERB

Testing the Rules on a Fragment of Discourse

After we felt confident that the rules arrived at during the preliminary syntheses described above produced acceptable results, we proceeded to test them on a set of sentences that, when connected together formed a fragment of a grammatically coherent discourse. We used words composed of sonorant segments as much as possible in order to make the pitch editing easier. The sentences were recorded in random order three times by the same speaker used in previous studies. Subsequently, the recordings were edited and the most neutral-sounding reading of each sentence was chosen for pitch editing. This was done in order to test whether, for example, we could obtain natural sounding focal prominences by just editing F_0 and leaving segment duration untouched, even in cases where the originally focussed word was extremely long in relation to the word receiving the new synthesized F_0 movements realizing focus. These recorded utterances had, in fact, prominences that would not be appropriate had the sentences been grouped together in a discourse. In (5), below, we have reproduced the sentences in the order that they would appear in a connected fragment of discourse. Subscripts indicate contextual coreference relations. We have indicated the sentences whose original intonation sounded inappropriate with a star (*) and writing the word with the deviant pitch obtrusion in bold letters. According to the focus assigning component, none of these words should receive prominence since they are contextually coreferent. For instance, the cash$_1$, it$_1$, and my money$_1$ are assumed to refer to the same referent, introduced by alimony$_1$. Cash and money are to be regarded as

hyponyms of alimony (see Granville 1984 and Fraurud 1986, for example, for a discussion of how superordinate hierarchies are built into computer text generating and interpretation systems). Moreover, the second and third occurrences of million can be replaced by such with reasonable acceptability, which proves they are coreferential. The NP the creep, would be construed by its definiteness to be coreferential with some preceding animate noun (according to Sidner's (1983) model for determining coreferents, it is the nearest preceding focussed animate NP that would be construed as the antecedent, in this case, lawyer):

(5)

a) My_i husband's $lawyer_j$ $mailed_k$ me_i my_i $alimony_l$ yesterday

b) *I_i really $needed_m$ the $CASH_l$

c) I_i $needed_m$ it_l immediately

d) *I_i'd given away all my_i $MONEY_l$ and demanded some more from the $CREEP_j$

e) He_j unwillingly $sent_k$ me_i a $million_n$

f) *Nine $MILLION_n$ is still $owing_o$ me_i

g) *No, ten $MILLION_n$ is still $OWING_o$ me_i

We then took each of these sentences and resynthesized the F_0 contour in accordance with the procedures used in the preliminary syntheses described above. That is to say, we used the same grid design as in Figure 4. Following the focus assigning model in Figure 1, the first focus assigned was given a pitch level extending over 100% of the width of the grid, the second reached 75% of the way from the baseline to

the topline, and the third, 50% of the way. Furthermore, all prefocal lexical heads in a given sentence were assigned a 'phrase accent' corresponding to a level of prominence extending 25% of the perpendicular distance from the baseline to the topline.

Scope of F_0 obtrusion. A new problem arose, however, when we followed the earlier practice of letting the focal pitch obtrusions extend over just the lexically stressed syllable. In cases where the rate of speech was relatively fast, a very unnatural sounding result was obtained by just placing the obtrusion over the stressed syllable. This was particularly evident in the case of sentence (5d), where, for example, the stressed syllable of more was so short that a rise and a fall over it was deemed unacceptable. On subsequent examination of F_0 contours produced by the speaker, however, it was observed that the minimal F_0 focal obtrusion in the data extended over a stretch of segments covering about 40 'frames' (=40X6.4ms). The obtrusions were, moreover, seen to be symmetrical around the peak, which occurred towards the end of the stressed vowel. We therefore decided to modify the rule for generating the pitch obtrusions so as to read:

From a point 2/3 of the way into the stressed vowel, define points 20 frames (= 20X6.4ms) to the left and right of this point. Connect the peak with these points. In cases of overlapping F_0 movements, join the peak with the point where the F_0 movements would potentially intersect (see, e.g. Figure 19).

Elaborated prosody model

Following in Figure 13 is a flow-chart elaborating on Figure 3 and containing all the information necessary in order to synthesize the F_0 contours for the sentences in (5). In Figures 14-20, we have presented the synthesized F_0 of all sentences in (5). Sample derivations are given in Figures 17 and 19 for sentences (d) and (f), respectively.

As regards the actual way the synthesis (point 14 in Figure 13) of overlapping contours would be accomplished in a computerized program, it has been pointed out (Lars Eriksson, personal communication) that one method would be to first derive intermediary curves, one for each F_0 movement and subsequently make a synthesis of all these, connecting all the highest points in all cases (see Figure 19 for an illustration of how this would be effected).

Discussion and conclusion

The syntheses (Figures 14-20) resulting from the rules in Figure 13 sounded very good[1]. Contrary to what has often been reported, the declining contours on all sentences did not sound monotonous. This reported monotony of synthesized speech is perhaps due to some other factors such as assigning the same pattern of F_0 peaks to all sentences, disregarding relative levels of focal and phrasal prominence.

Assigning a phonetic reality to the baseline had the positive consequence that one did not have to formulate separate transition rules for connecting one pitch obtrusion to another. The baseline took the place of these transitions, since the pitch movements were defined with respect to this

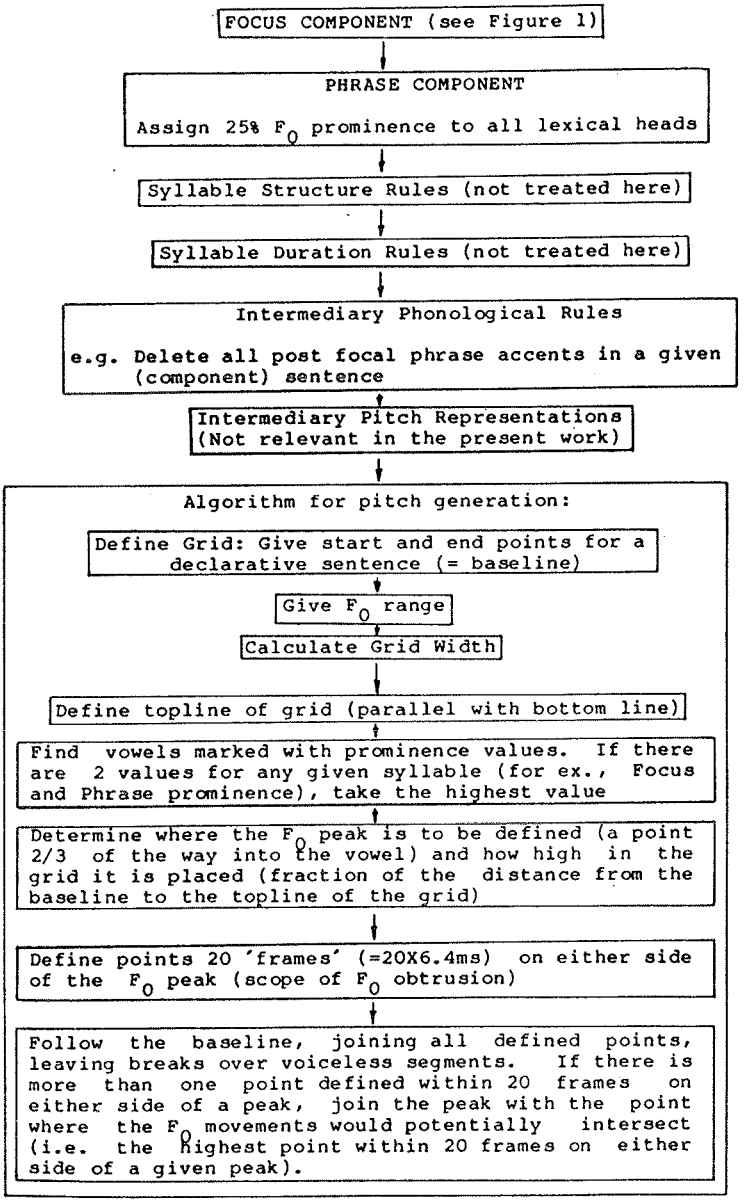

FOCUS COMPONENT (see Figure 1)
1
PHRASE COMPONENT
Assign 25% F_0 prominence to all lexical heads
2
Syllable Structure Rules (not treated here)
3
Syllable Duration Rules (not treated here)
4
Intermediary Phonological Rules
e.g. Delete all post focal phrase accents in a given (component) sentence
5
Intermediary Pitch Representations (Not relevant in the present work)
6
Algorithm for pitch generation:
Define Grid: Give start and end points for a declarative sentence (= baseline)
7
Give F_0 range
8
Calculate Grid Width
9
Define topline of grid (parallel with bottom line)
10
Find vowels marked with prominence values. If there are 2 values for any given syllable (for ex., Focus and Phrase prominence), take the highest value
11
Determine where the F_0 peak is to be defined (a point 2/3 of the way into the vowel) and how high in the grid it is placed (fraction of the distance from the baseline to the topline of the grid)
12
Define points 20 'frames' (=20X6.4ms) on either side of the F_0 peak (scope of F_0 obtrusion)
13
Follow the baseline, joining all defined points, leaving breaks over voiceless segments. If there is more than one point defined within 20 frames on either side of a peak, join the peak with the point where the F_0 movements would potentially intersect (i.e. the highest point within 20 frames on either side of a given peak).
14

Figure 13

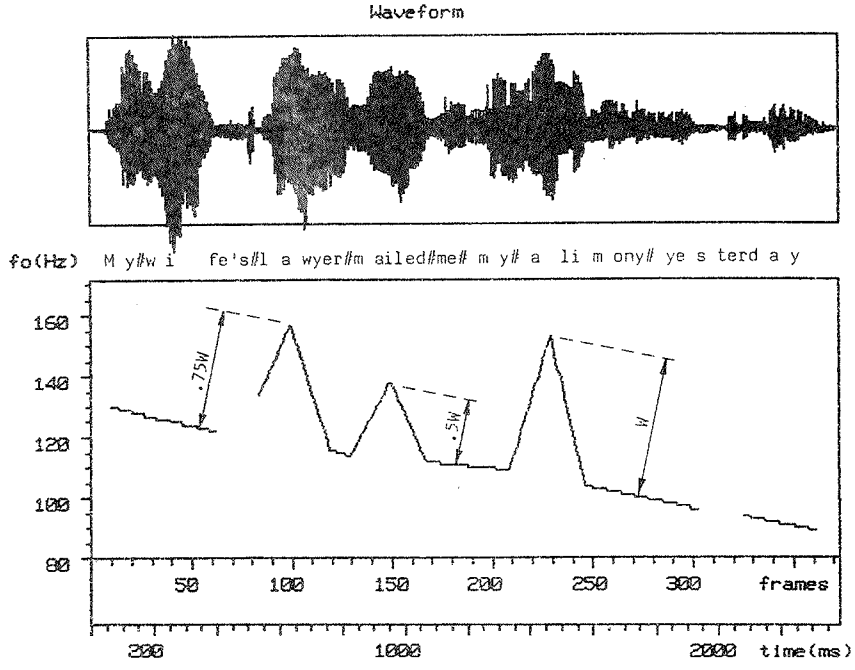


FIGURE 14. SYNTHESIZED F_0 CURVE OF SENTENCE (5a) WITH FOCUS ON SUBJECT, PREDICATE AND OBJECT

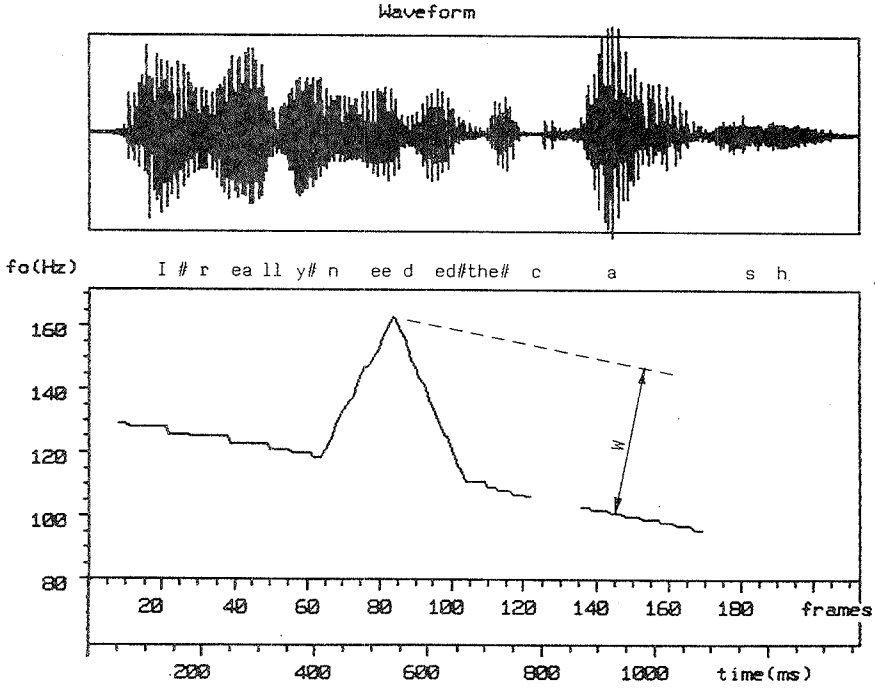


FIGURE 15. SYNTHESIZED F_0 CURVE OF SENTENCE (5b) WITH FOCUS ON PREDICATE

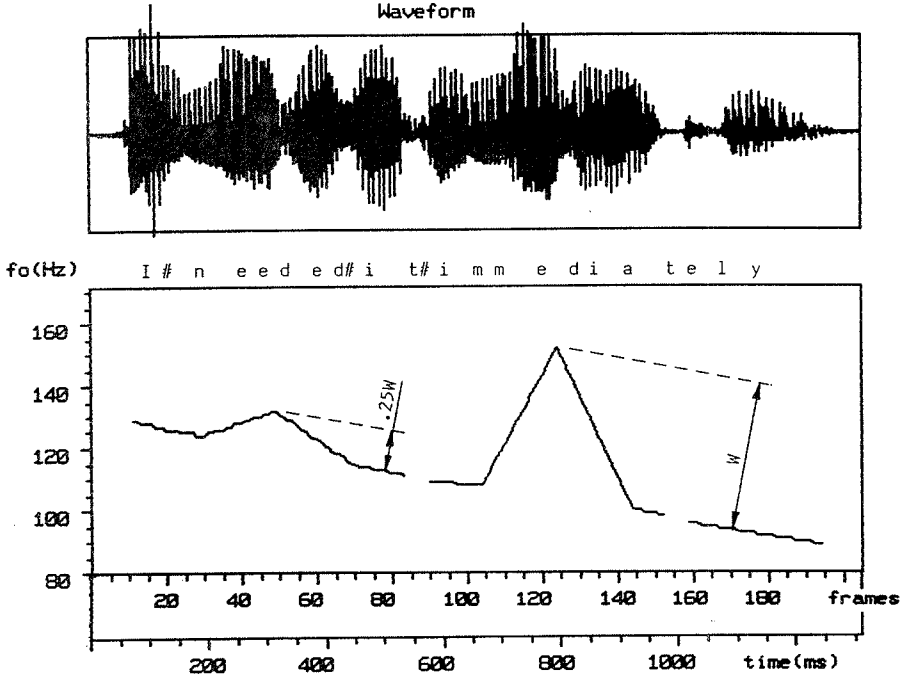


FIGURE 16. SYNTHESIZED F_0 CURVE OF SENTENCE (5c) WITH FOCUS ON PREDICATE COMPLEMENT AND PHRASE ACCENT ON PREFOCAL PREDICATE (HEAD)

Derivation of sentence (5d) (see Figure 17 below) following prosody model in Figure 13

F=W F=.75W F=W
1) I'd given away all my money and demanded some more from the creep.

P=.25W P=.25W P=.25W P=.25W P=.25W
F=W F=.75W F=W
2) I'd given away all my money and demanded some more from the creep.

P=.25W P=.25W P=.25W
F=W F=.75W F=W
5) I'd given away all my money and demanded some more from the creep.

7) Baseline defined in Figure 17a

8) F_0 range defined in Figure 17a

9) Grid width (W) calculated in Figure 17a

10) Topline of grid defined in Figure 17a

F=W F=.75W F=W
11) I'd given away all my money and demanded some more from the creep

12) Define F_0 peaks in grid (X's in Figure 17a)

13) Define scope of F_0 obtrusion (*'s in Figure 17a)

14) Generate F_0 contour (Figure 17b)

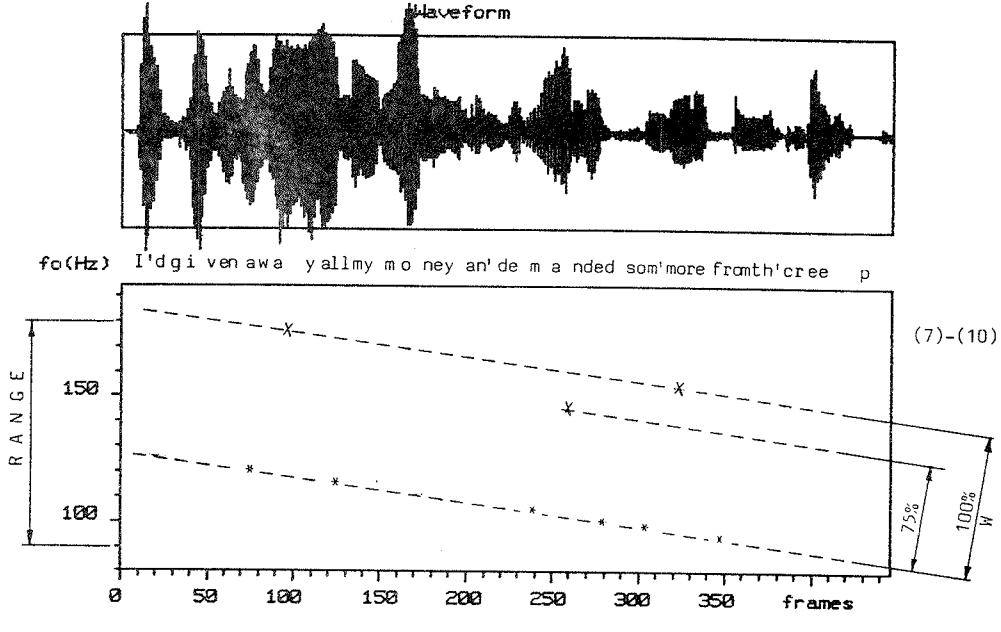


FIGURE 17a. PARTIAL DERIVATION OF F_0 CURVE OF SENTENCE (5d) (AFTER POINT 13 IN FLOW-DIAGRAM). SEE ABOVE FOR A CLARIFICATION OF THE FIGURE.

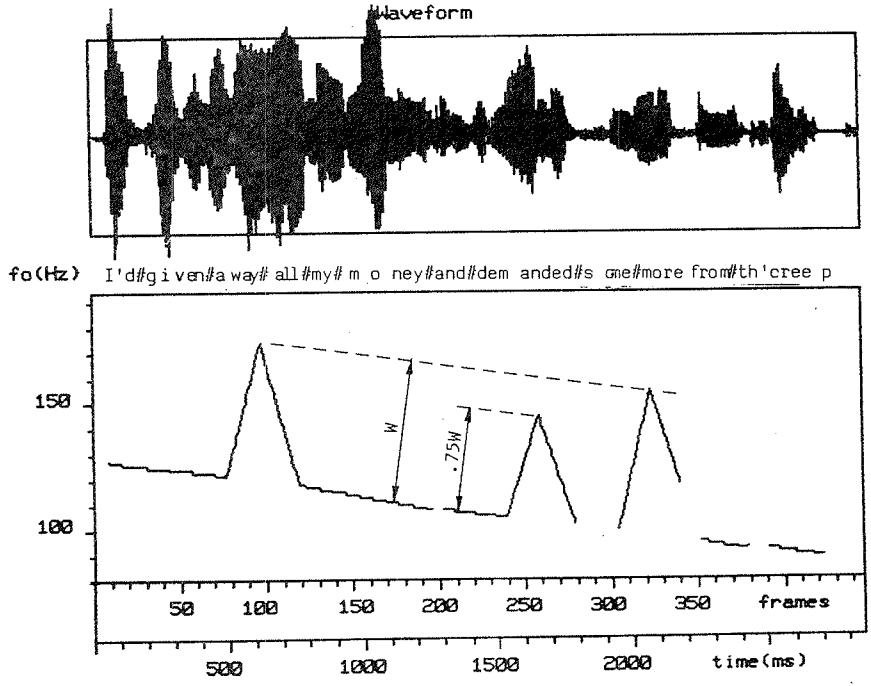


FIGURE 17b. OUTPUT OF THE PITCH GENERATING COMPONENT WITH FOCUS ON PREDICATE IN FIRST COMPONENT SENTENCE AND FOCUS ON PREDICATE AND PREDICATE COMPLEMENT IN SECOND COMPONENT SENTENCE. NOTE THAT DECLINATION IS NOT 'RESET' AT BEGINNING OF SECOND CLAUSE.

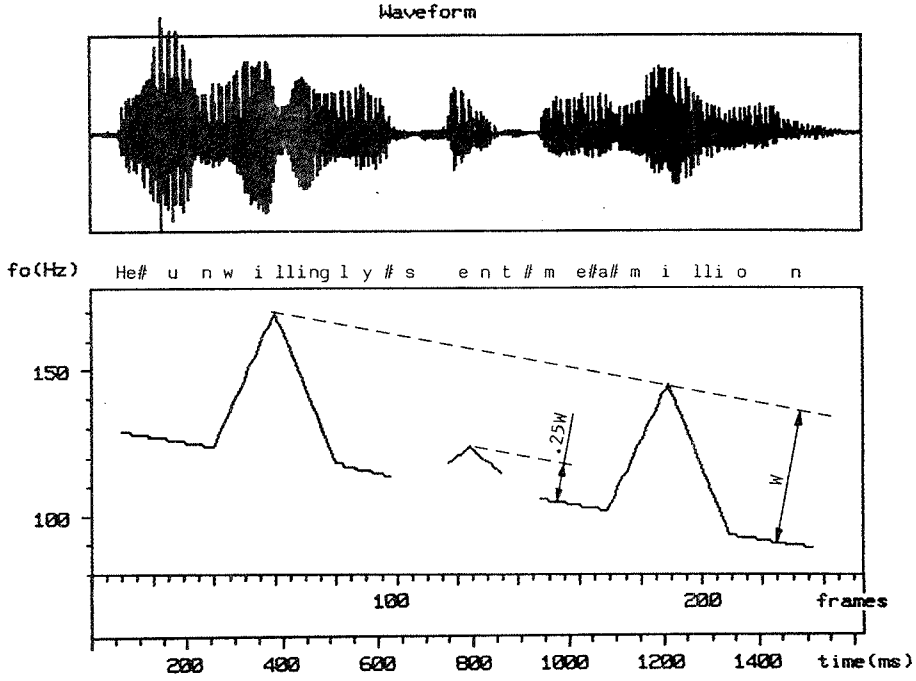


FIGURE 18. SYNTHESIZED F_0 CURVE OF SENTENCE (5e) WITH FOCUS ON PREDICATE COMPLEMENTS AND PHRASE ACCENT ON PREDICATE (HEAD)

Derivation of sentence (5f) (see Figure 19 below) following prosody model in Figure 13

```
   F=W                    F=.75W
1) Nine million is still owing me

        P=.25W            P=.25W
   F=W                    F=.75W
2) Nine million is still owing me
```

5) Not applicable

7) Baseline defined in Figure 19a

8) F_0 range defined in Figure 19a

9) Grid width (W) calculated in Figure 19a

10) Topline of grid defined in Figure 19a

```
        P=.25W
   F=W                    F=.75W
11) Nine million is still owing me
```

12) Define F_0 peaks in grid (X's in Figure 19a)

13) Define scope of F_0 obtrusion (*'s in Figure 19a)

14) Generate F_0 contour (Figure 19b)

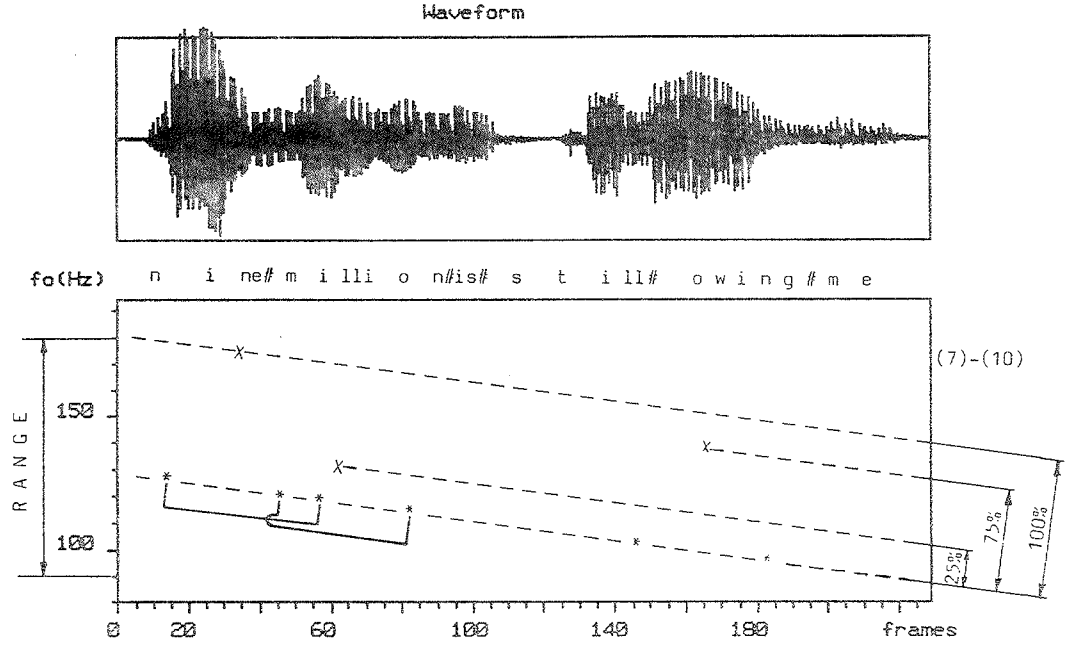


FIGURE 19a. PARTIAL DERIVATION OF F_0 CURVE OF SENTENCE (5f) (AFTER POINT 13 IN FLOW DIAGRAM IN FIGURE 13). SEE ABOVE FOR A CLARIFICATION OF THE FIGURE. DERIVATION CONTINUED IN FIGURE 19b.

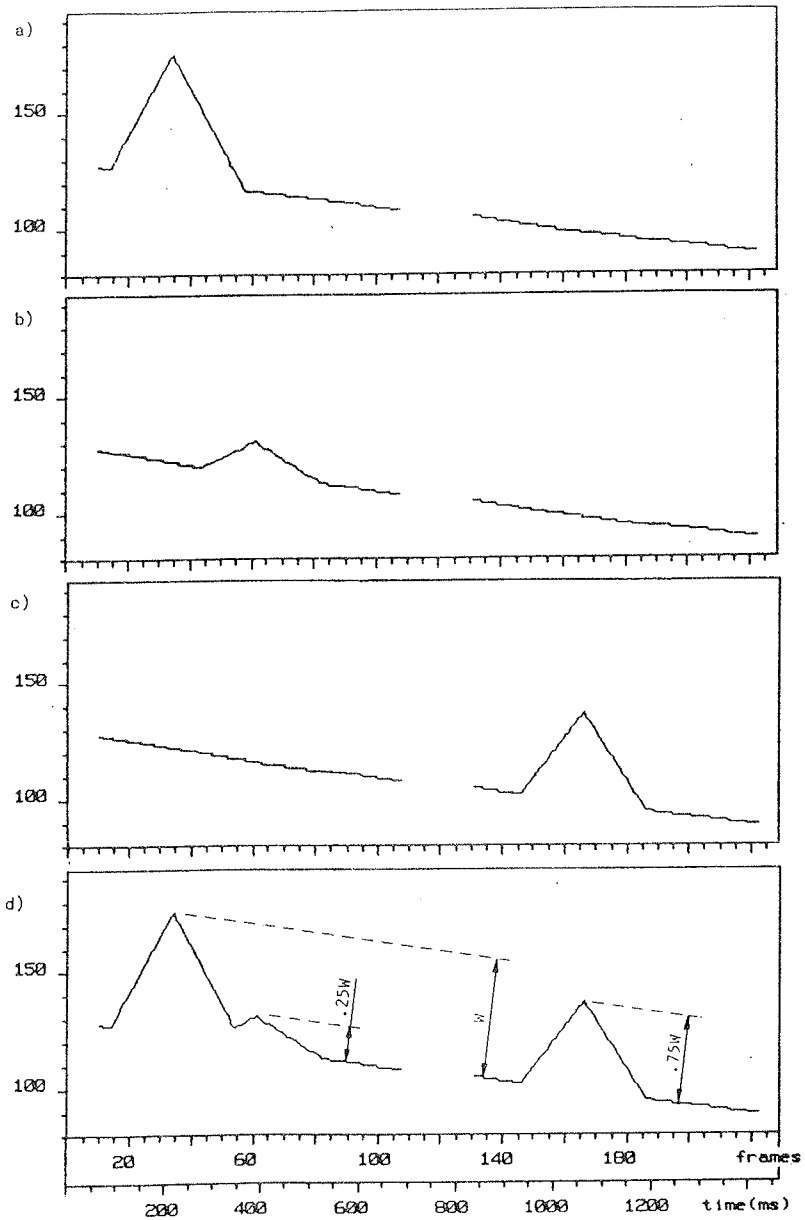


FIGURE 19b. POTENTIAL STAGES IN THE SYNTHESIS OF THE F_0 CURVE WHERE THE FIRST TWO PITCH OBTRUSIONS OVERLAP. THE FINAL OUTPUT IN (d) IS OBTAINED BY CONNECTING THE HIGHEST POINTS IN THE INTERMEDIARY CURVES (a-c).

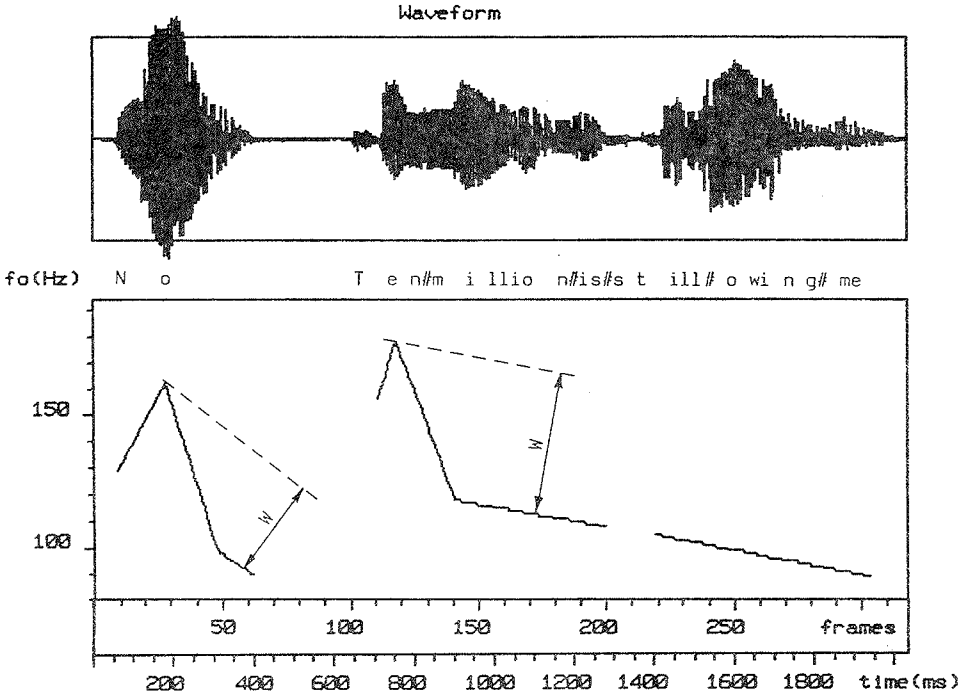


FIGURE 20. SYNTHESIZED F_0 CURVE FOR SENTENCE (5g) WITH FOCUS ON SUBJECT MODIFIER 'TEN'. PROMINENCE ON 'NO' NOT ACCOUNTED FOR BY THE MODEL PRESENTED HERE BUT RATHER ASSUMED TO BE ASSIGNED BY OTHER RULES.

reference line; their theoretical beginning and end points lay on this line. It is perhaps the case, however, that for certain speech styles or rates, one would have to define special rules that connected pitch obtrusions with transitions that lie higher or lower than the baseline. More research is needed in order to clarify this point.

The analyses done here with synthesized F_0 supported the well-known fact that pitch constitutes a more important indicator of focal prominence than duration in English. For example, we could 'deaccent' the very long word cash in sentence (5b) and move the focus to the relatively short word needed by just adding an F_0 obtrusion (see Figure 15). Duration is, however, an important concomitant feature of focal prominence (see e.g. Bannert 1986, Eady et al. 1986). House & Horne (1987) also found that the duration of the stressed vowel in a focussed word was essentially constant for a given speaker regardless of the rate of speech.

An interesting side-result concerning the segmental content of the data studied here, was that in the synthesis of sentence (5d), the movement of focal prominence from creep to more left creep sounding rather peculiar due to the strong aspiration of p after the 'deaccented' vowel. Heavy aspiration is obviously an unacceptable feature in this environment and something that should be ruled out in segment synthesis programs.

The Lund model of prosody revealed itself to be very useful in synthesizing F_0 contours in English, easily lending itself to quantification. The concept of the phonological grid to express sentence intonation proved to be most

appropriate for representing the F_0 movements realizing focal prominences and phrase boundaries. We can expect, however, that our application of the model to English will differ from its quantification for Swedish but this is mainly due to the different prosodic natures of the two languages. Put in a nutshell, we have analysed English sentence intonation as being built up around focal accents; Swedish sentence intonation, on the other hand is built up on the lexical word accents, nonexistant in English. This fundamental difference between the two languages has important consequences when one attempts to formulate rule systems to account for the intonational patterning in each language. It is, as pointed out, focus which lies at the basis of our analysis of English and empirical observations of focal prominence, moreover, which determined the design of the grid. In Swedish, on the other hand, it is (at least in the analyses discussed in this work) the distinctive word accents which form the basis of the prosodic analysis and upon which the description is built up. In the phonological description of Swedish, words come from the lexicon with pitch accents. Other prominences signalling focus and phrase boundaries are then assumed to be added, or superimposed on these already existing word accents. Our goal has been to show how certain generalizations about English declarative sentence prosody can be structured into a rule system to synthesize appropriate F_0 contours for a fragment of discourse. We feel that an approach based on focal prominence constitutes an insightful way to account for the patterning of sentence intonation in this language. More research is of course needed in order to expand the rule

system so as to be able to synthesize other patterns of sentence prosody.

ACKNOWLEDGEMENTS

I am grateful to Gösta Bruce, Eva Gårding, Thore Pettersson, and Bengt Sigurd for comments on earlier versions of this paper. All remaining shortcomings are, however, my own responsability.

FOOTNOTES

1. A casette tape containing copies of all sentences with synthesized F_0 curves discussed in this paper can be supplied by the author upon request.

REFERENCES

Bannert, R. 1986. Independence and interdependence of prosodic features. Working papers 29 (Dept. of Linguistics, U. of Lund), 31-60.

Bruce, G., 1977. Swedish word accents in sentence perspective. Lund:Gleerups.

_____, 1982. Developing the Swedish intonation model. Working papers 22 (Dept. of Linguistics, U. of Lund), 51-116.

_____ and Gårding, E., 1978. A prosodic typology for Swedish dialects. In E. Gårding, G. Bruce and R. Bannert (eds.),

Nordic prosody (Travaux de l'Institut de linguistique de Lund, 13), 219-28.

Cohen, A., Collier, R., and t'Hart, J. 1982. Declination: construct or intrinsic feature of speech pitch? Phonetica 39, 254-73.

Eady, S., Cooper, W., Klouda, G., Mueller, P., and Lotts, D., 1986. Acoustical characterization of sentential focus: narrow vs. broad and single vs. dual focus environments. Language and speech 29, 233-50.

Fraurud, K. 1986. The introduction and maintenence of discourse referents. In: Ö. Dahl (ed.), Papers from the ninth Scandinavian conference of linguistics, Stockholm, Jan. 9-10, 1986, 111-22. University of Stockholm: Inst. of Linguistics.

Fujisaki, H., and Hirose, K., 1982. Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In: Preprints, Working group on intonation, XIII Int. Cong. of Linguists, Aug. 31, 1982, 57-70. Tokyo.

Gårding, E., 1977. The importance of turning points for the pitch patterns of Swedish accents. In L. Hyman (ed.), Studies in stress and accent, 27-35. Los Angeles: University of Southern California.

_____, 1981. Contrastive prosody: a model and its applications. Studia linguistica 35, 146-65.

_____, 1983. A generative model of intonation. In: A. Cutler and D. Ladd (eds.), Prosody: models and measurements, 11-25. Berlin:Springer.

Granville, R. 1984. Controlling lexical substitution in

computer text generation. In: Proceedings Coling 84, 2-6 July 1984, Stanford Univ., California, 381-4. Association for Computational Linguistics.

t'Hart, J. 1982. The stylization method applied to British English intonation. In: Preprints, Working group on intonation, XIII Int. Cong. of Linguists, Aug. 31, 1982, 23-33. Tokyo.

Horne, M., 1985. English sentence stress, grammatical functions and contextual coreference. Studia linguistica 39, 51-66.

_____, 1986a. Information focus: assignment and phonological implications. In: L. Evensen (ed.), Nordic research in text linguistics and discourse analysis, 155-70. University of Trondheim:TAPIR.

_____, 1986b. Focal prominence and the 'phonological phrase' within some recent theories. Studia linguistica 40, 101-21.

House, D. and Horne, M., 1987. Focus, reduction and dialect: a study of fast speech phenomena. Paper presented at the Miniconference on reduction and elaboration phenomena in speech production, U. of Stockholm, May 4-5, 1987.

Huber, D., 1985. Swedish intonation contours in text-to-speech synthesis. Working Papers 28 (Dept. of Linguistics, U. of Lund), 109-25.

Ladd, R., 1983. Phonological features of intonational peaks. Language 59, 721-59.

_____, 1986. Intonational phrasing: the case for recursive prosodic structure. Phonology yearbook 3:311-40.

Lindau, M., 1986. Testing a model of intonation in a tone language. J. Acoust. Soc. Am. 80, 757-64.

Olive, J., and Liberman, M., 1979. A set of concatenative units for speech synthesis. In J. Wolf and D. Klatt (eds.), Speech communication papers presented at the 97th meeting of the Acoustical Society of America, 515-8. New York: Acoustical Society of America.

Pierrehumbert, J., 1981. Synthesizing intonation. J. Acoust. Soc. Am. 70, 985-95.

Sidner, C., 1983. Focusing in the comprehension of definite anaphora. In: A. Brady and R.C. Berwich (eds.), Computational models of discourse, 267-330. Cambridge, Mass.: MIT Press.

Sigurd, B., 1984. Computer simulation of spontaneous speech production. In: Proceedings Coling 84, 2-6 July 1984, Stanford Univ., California, 79-83. Association for computational linguistics.

_____, 1987. Referent grammar (RG) in computer comprehension, generation and translation of text (SWETRA). Working Papers, Dept. of Ling., U. of Lund (this volume).

Thorsen, N. 1980. A study of the perception of sentence intonation: evidence from Danish. J. Acoust. Soc. Am. 67, 1014-30.

V

'CONTRASTIVE PROMINENCE' AND SYNTACTIC PARALLELISM

Merle Horne
Dept. of Linguistics
Lund University
1987

Introduction

In Horne (1985-1987), a discourse-based model was developed for describing and synthesizing English declarative, 'non-expressive' sentence intonation. As the model stands, however, it does not account for a number of F_0 patterns that occur quite frequently in non-expressive contexts.[1] These are usually grouped under the term 'contrastive stress' determined by syntactic parallelism. (Here, we will use the term 'contrastive prominence' instead of 'contrastive stress', reserving 'stress', like Bolinger (1958), to refer solely to underlying, 'abstract', word stress, potentially realizable as pitch.)

Just what this syntactic parallelism is has never been explicitly defined in detail, but it is clear that an adequate prosody model must be able to handle it (see e.g. Vanderslice 1968). Not only are the 'contrastive prominences' conditioned by other parameters than the 'normal' focal prominences assigned by our algorithm, but they also have implications for intonational phrasing. They are also realized by other F_0 contours than the 'normal' HL contour associated with focal

prominences in noncontrastive declarative sentences. The purpose of this paper is to provide a better understanding of the parallel structures conditioning contrastive F_0 patterns and to indicate how they can be included in a rule system for synthesizing English intonation.

Contrastive prominence in parallel sentences

A typical example of contrastive prominence is presented in (1) (see also Horne 1985:65, 1986a:162):

(1) Peter_i kicked_j Ulrika_k and then SHE_k kicked_j HIM_i

(where identical subscripts identify coreferential relations)

The prominence on SHE and HIM has also been termed, for example, 'reciprocal contrastive stress' (Lakoff 1971), and the pattern as a whole has been referred to as 'paired narrow foci' (Ladd 1978), or 'bi-focal structures' (Enkvist 1979).

As indicated by the indices, SHE and HIM in the second component sentence are coreferential with Ulrika and Peter, respectively, in the first component sentence and consequently would not be expected to be assigned prominence associated with information focus according to our model.[2] Nevertheless, the pronouns do receive prominence. According to Enkvist (1979:141), this prominence functions to indicate a "shift in semantic roles"; that is to say, what was the 'actor' in the first sentence becomes the 'patient' in the second and what was the 'patient' in the first sentence becomes the 'actor' in the second. One could also, within the framework of a referent grammar such as that developed in Sigurd (1987a,b), relate the

intonation pattern to a change, in the second component sentence, of referents associated with the grammatical functions in the first component sentence, i.e. the referent that functioned as the 'subject' in the first sentence becomes the 'object' in the second, and what was the 'object' in the first sentence becomes the 'subject' in the second.

The formal specification of the conditions under which this contrastive intonation pattern occurs would, at first glance, seem quite easy to define. According to Lakoff (1971: 63), the rule is to be expressed (in generative semantics terms) as in (2):

(2) In sentences of the form f(a,b) and f(b,a), where f stands for a phrase-marker minus the elements a and b, stress a and b

However, as Lakoff himself points out (1971:63), (2) is not general enough, since

"the rule applies not only in conjunctions but whenever disjoint clauses of this form occur in a sentence. For example:
The fact that John insulted Mary indicated that SHE would soon insult HIM.
Note that the time adverb soon must be considered semantically external to its clause if one is to state the identity condition uniformly".

In addition to Lakoff's own comments on the form of (2), we can specify further conditions under which the rule applies. Not only does the contrastive pattern occur when the same nominal referents occur in both predications as in (2), but also when only one of the nominal referents that occurred in the first sentence also occurs in the second, provided of course that it does not fill the same grammatical function as it did in the

first, e.g.:

(3) Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$ and then CHRISTINA$_{(d)Subj}$ kicked$_{(b)Pred}$ HIM$_{(a)Obj}$

(where the letters in brackets are referent indices characteristic of referent grammar)

In (3), only one of the two nominal referents in the second sentence is identical to those in the first, i.e. HIM, but the fact that it fills a different grammatical function than it did in the first provides a sufficient environment for the contrastive intonation pattern to occur.

Furthermore, not only is it possible for only one of the nominal referents from the first sentence to occur in the second sentence, but, in fact, it is the case that neither of the nominal referents from the first sentence need reappear in the second. To obtain a contrastive reading, it is sufficient that the verbal referents be identical in the two sentences, as in (4):

(4) Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$ and then FREDRIK$_{(d)Subj}$ kicked$_{(b)Pred}$ CHRISTINA$_{(e)Obj}$

The identity of the predicates, in fact, allows one to delete the second occurrence, as illustrated by the following, elliptical version:

(4′) Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$ and then FREDRIK$_{(d)Subj}$, CHRISTINA$_{(e)Obj}$

Actually, other cases of contrastive prominence can be accounted for by generalizing (2) even further. It is not the case, for example, that it is the predicates in the two parallel sentences that must be identical, but rather, it is sufficient that the lexical material realizing any grammatical function whatsoever be identical in the pair (actually, n-tuple) of sentences for the contrastive pattern to occur. For example, all that the sentence pairs in (5) have in common is one nominal referent and yet they exhibit a contrastive intonation pattern analogous to those discussed above with identical verbal referents:

(5) a. Peter$_{(a)\text{Subj}}$ kicked$_{(b)\text{Pred}}$ Ulrika$_{(c)\text{Obj}}$ and then he$_{(a)\text{Subj}}$ KISSED$_{(d)\text{Pred}}$ SUE$_{(e)\text{Obj}}$

b. Peter$_{(a)\text{Subj}}$ kicked$_{(b)\text{Pred}}$ Ulrika$_{(c)\text{Obj}}$ and then BILL$_{(d)\text{Subj}}$ KISSED$_{(e)\text{Pred}}$ her$_{(c)\text{Obj}}$

In (5a), it is the subjects that are identical, and in (5b), the objects.

It should be pointed out here that parallel sentences in which only one of the referents is different have sometimes also been considered to have 'contrastive stress' (see Bing 1980:147), as in (6):

(6) a. Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$

and then he$_{(a)Subj}$ kissed$_{(d)Pred}$ her$_{(c)Obj}$

b. Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$

and then Sue$_{(d)Subj}$ kicked$_{(b)Pred}$ her$_{(c)Obj}$

These cases, however, can be adequately handled by our algorithm[2]. Moreover, they are not characterized by the same kind of F_0 contour ('fall-rise') associated with the contrasted elements in the parallel sentences with two nonidentical referents discussed above. This gives support to the model which includes them under the 'normal' (HL) intonation pattern.

In light of the above discussion, therefore, we can generalize (2), rewriting it as (2'):

(2') Contrastive Prominence in Parallel Sentences:

In sentences of the form:

Subj(a_i), Pred(b_j), PredComp(c_k)...

Subj(d_l), Pred(e_m), PredComp(f_n)

where a-f are constituents with associated referents (i-n) realizing a particular grammatical function, if i=l, or j=m or k=n, assign contrastive prominence to the nonidentical constituents (prefinal contrasted items will receive a HL+H% F_0 contour; the final contour will be HL. See below for phonetic detail).

In other words, (2') says that in a pair of sentences, if there is one referent that is identical in the two sentences which

is also associated with the same grammatical function in the two sentences, each lexical item associated with the nonidentical referents in the second sentence is assigned contrastive prominence. The following sentences illustrate a case where contrasting genitive modifiers receive contrastive prominence:

(7) (John's$_{(a)}$ dachshund$_{(b)}$)$_{Subj}$ bit$_{(c)}$ Pred (Peter's$_{(d)}$ son$_{(e)}$)$_{Obj}$ and then (MY$_{(f)}$ dog$_{(b)}$)$_{Subj}$ bit$_{(c)}$Pred (YOUR$_{(g)}$ little boy$_{(j)}$)$_{Obj}$

In (7), it is the modifiers, MY and YOUR that receive contrastive prominence and not the heads of their respective constituents which are coreferential.

Note that, unlike Lakoff's rule (2), we restrict the assigning of prominences in (2') to words only in the second of two parallel sentences. This is because the intonation pattern of the first sentence corresponds to what one would predict from our model[2]. It is not necessarily the case that, for example, a and b in the first sentence in (2) would be assigned "stress" (focal prominence). This depends on their coreferential status with respect to preceding parts of the discourse. See, for example, the fragment of text in (8). In this example, both nominal referents in the first of the two parallel sentences in (III) are coreferential with identical occurrences in the first sentence of the fragment (I) and therefore receive no prominence whatsoever. Nevertheless, in the second of the two parallel sentences in (III), both SHE and

HIM receive prominence as in (1).

(8) (I) Peter$_i$ and Ulrika$_j$ were playing in the sandbox. (II) They seemed to be having a good time. (III) All of a sudden, however, Peter$_i$ kicked$_k$ Ulrika$_j$ and then, a few seconds later, SHE$_j$ kicked$_k$ HIM$_i$

Note, also, that the condition on identical referents in (2'), for example, involves, as with our notion of 'contextual coreference', establishing referent identity on the basis of e.g., rules of pronominal anaphora, as well as lexical implicational rules that account for such relations as synonomy and hyponomy. Notice, for instance, that j and m in (2') can be realized as synonyms of each other (9a), or the first verbal predicate can e.g. be a hyponym (subordinate term) of the second (9b), but not vice versa (9c):

(9) a. Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$ and then SHE$_{(c)Subj}$ booted$_{(b)Pred}$ HIM$_{(a)Obj}$

(kick, booted are synonyms)

b. Peter$_{(a)Subj}$ kicked$_{(b)Pred}$ Ulrika$_{(c)Obj}$ and then SHE$_{(c)Subj}$ hurt$_{(b)Pred}$ HIM$_{(a)Obj}$

(kicked is a hyponym of hurt)

c. * Peter$_{(a)Subj}$ hurt$_{(b)Pred}$ Ulrika$_{(c)Obj}$ and then SHE$_{(c)Subj}$ kicked$_{(d)Pred}$ HIM$_{(a)Obj}$

(hurt is not a hyponym of kicked)

Inference Rules

There are some cases exhibiting the intonation pattern associated with parallel sentences which we believe can be explained by assuming a more abstract analysis involving inference. For example, the following pair of sentences in (10) (taken from Werth 1979:243), are characterized by the same intonation pattern as those above, yet they do not exhibit any 'surface' parallelism:

(10) $\text{John}_{(a)\text{Subj}}$ $\text{married}_{(b)\text{Pred}}$ $\text{a blonde}_{(c)\text{Obj}}$.

$(\text{MY}_{(d)}\ \text{wife}_{(e)})_{\text{Subj}}$ $\text{has}_{(f)\text{Pred}}$ $(\text{DARK}_{(g)}\ \text{hair}_{(h)})_{\text{Obj}}$.

One way, we think, of understanding the intonation pattern in the second sentence of (10) is to assume that it constitutes the second part of a parallel sentence construction, the first part being an inferred sentence which forms the environment of the intonation pattern, as for instance in (10'):

(10') $\text{John}_{\text{Subj}}$ $\text{married}_{\text{Pred}}$ $\text{a blonde}_{\text{Obj}}$

<Inferred sentence:

$(\text{John's}_{(l)}\text{wife}_{(m)})_{\text{Subj}}$ $\text{has}_{(n)\text{Pred}}$ $(\text{blond}_{(o)}\ \text{hair}_{(p)})_{\text{Obj}}$>

$(\text{MY}_{(d)}\ \text{wife}_{(m)})_{\text{Subj}}$ $\text{has}_{(n)\text{Pred}}$ $(\text{DARK}_{(g)}\ \text{hair}_{(p)})_{\text{Obj}}$

where d and l, and g and o contrast.

According to Werth (1979:243), the examples in (10) are to be explained by the concept of 'lexical decomposition'. For example, (10) is analyzed as in (11):

(11) John <took as wife> a < girl with blond hair>.

MY wife has DARK hair.

With such a representation, however, it is difficult, if not impossible, to express the conditions under which the intonation pattern in (10) occurs. The verbal referents <took as wife> and <has> do not exhibit identity, for example. We feel that it is inference rules that best can account for the parallel sentence intonation pattern in (10). Just how inference rules would be applied in text analysis is not immediately evident. However, their presence in a complete text-to-speech system is an obvious necessity. At least as far as assigning sentence intonation is concerned, it is possible to speculate on how these inference rules would operate. It is clear that it is important to be able to infer from a given sentence, referents that are identical to those in a preceding sentence. This would be done, for instance, in sentence sequences like that in (10), where the second sentence, with respect to the surface text, introduces a new topic, i.e., it does not comment on anything said in the previous sentence, the usual situation within a given paragraph, for example. The inference rules would then try to create a context in which the second sentence could constitute a well-formed sequence to the first sentence. This would be the case, of course, if at least one of the referents in the second sentence was identical to that in the first, or an inferred sentence of the first. This is seen to be the case in the example discussed in (10).

Phonetic Realization of Contrastive Prominence in parallel sentences

Phonetically, the prominences associated with the parallel

sentence intonation pattern are associated with an accent pattern that has been variously termed 'fall-rise' (Ladd 1978), H* L- H% (Pierrehumbert 1980), HL+H% (Ladd 1983), or 'A-contour-rise' (Bing 1980). In the sentences we informally examined exhibiting this F_0 pattern, the tops of the F_0 obtrusions on the 'contrasted' elements all received an equal level of prominence (corresponding approximately to .75 W (W=grid width)). Furthermore, the contrasted elements in the second (or, noninitial) component sentence fell within separate 'tone units' (Crystal 1969), 'speech chunks' (Sigurd 1984) or 'intonation phrases' (Selkirk 1984) (we assume that a speech chunk- or intonation phrase boundary is inserted after each constituent containing a contrasted word). In the data we have informally analyzed (parallel sentences produced by the author), the contrastive F_0 contour was realized phonetically by three signals: first of all, the F_0 level dropped to the bottom of the speaker range after the rise on the nominal words. Second, there was a rise after this fall up to the end of the constituent (about 20 frames = 20 X 6.4 ms); this rise reached a level corresponding to approximately the normal sentence initial starting point for this speaker. Thirdly, there was a pause after the first contrasted constituent (subject) in the second component sentence and then at the beginning of the following constituent (predicate), the F_0 began at the same level at which it stopped in the preceding phrase. Following in Figure 1 is an example of an occurring and synthezized F_0 curve for the sentence in (12):

(12) The boys$_{(a)\text{Subj}}$ eyed$_{(b)\text{Pred}}$ the girls$_{(c)\text{Obj}}$

and then THEY$_{(c)\text{Subj}}$ eyed$_{(b)\text{Pred}}$ THEM$_{(a)\text{Obj}}$

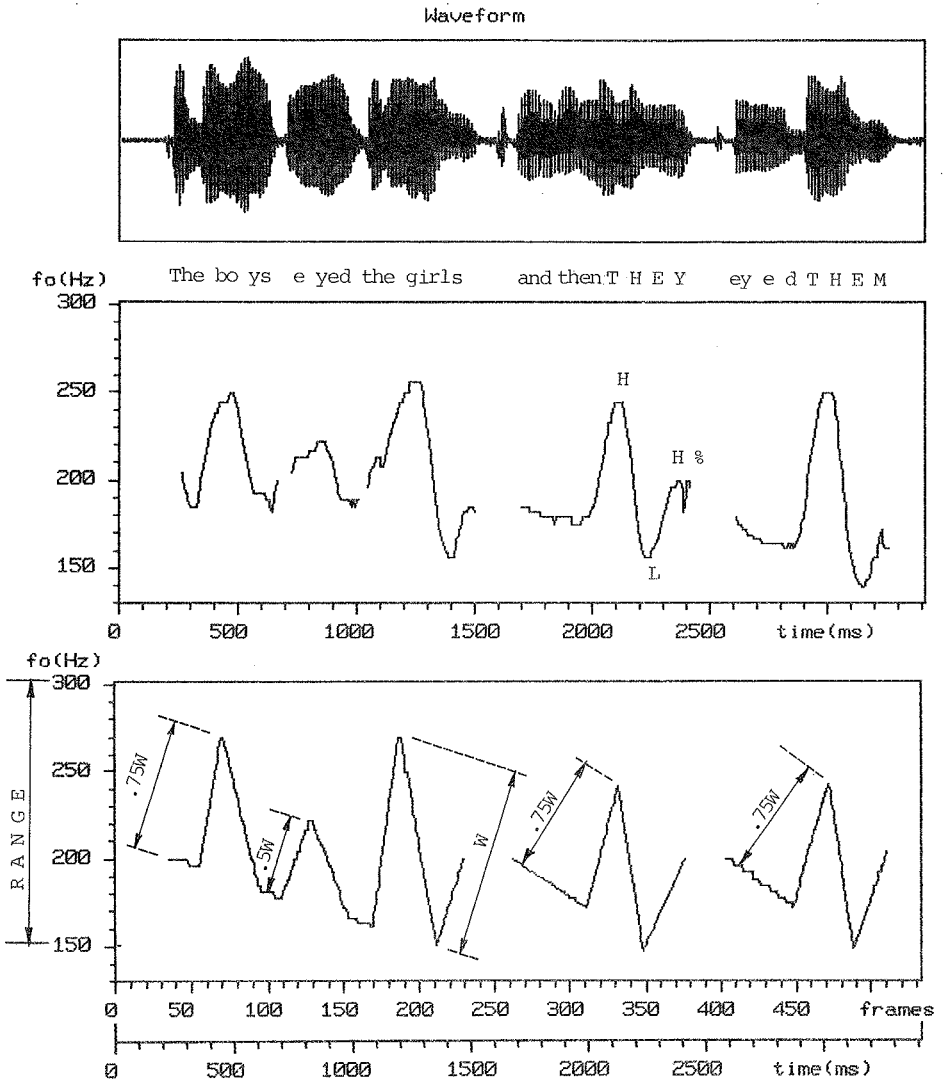


Figure 1. Occurring (a) and synthesized (b) F_0 contours of sentence 12. Notice the HL+H% prepausal contours. The synthesized version was made following a number of generalizations: The speaker range was fixed (as in Horne 1987) at 1 octave, with low point at 150Hz and high point at 300Hz. The speaker's normal starting point was set at 200Hz. Resetting of F_0 at the beginning of a new intonation phrase was also fixed at 200Hz, as was the second H in the HL+H% boundary contour. Note that the 'continuation' F_0 rise at the end of the sentence is not obligatory. HL contours on the contrasted elements have been provisionally set at a level corresponding to .75W (W = grid width). The scope of the F_0 obtrusions has not been fixed, but it can be seen that, as discussed in Horne 1987, they extend over about 40 frames (= 40x6.4ms)

Contraction of the Copula in Comparative Sentences

The prosodic boundary conditioned by the parallel sentences as defined above can perhaps be used to explain a problematic issue involving contraction of forms of the copula 'be' in English. It has been noted by Selkirk (1984), for instance, that in comparative constructions such as those in (13), the copula does not contract in the second 'compared' sentence:

(13) a. Jane {is / 's} a more brilliant doctor than Mary {is / *'s}
a promising lawyer

b. Willie {is / 's} as English as Malin { is / *'s} Norwegian

The impossibility of contraction in the second clause can be explained, we feel, as due to the same kind of intonation phrase boundary (%) as in the parallel sentences discussed above, since both sentences contain an identical verb, 'to be'. In order for contraction to take place, the form of the verb 'to be' must, of course be enclitic to the preceding word. This, in turn, implies absence of any strong boundary between the verb and the preceding word. Thus the fact that is does not contract in the second component sentences in (13) can be explained by the presence of an intonation phrase boundary after the first nominal phrase, conditioned by the parallel sentence structure as described in detail above. In Figure 2, we have presented an occurring F_0 contour for sentence (13b) where one can clearly see the intonation phrase boundary after Malin:

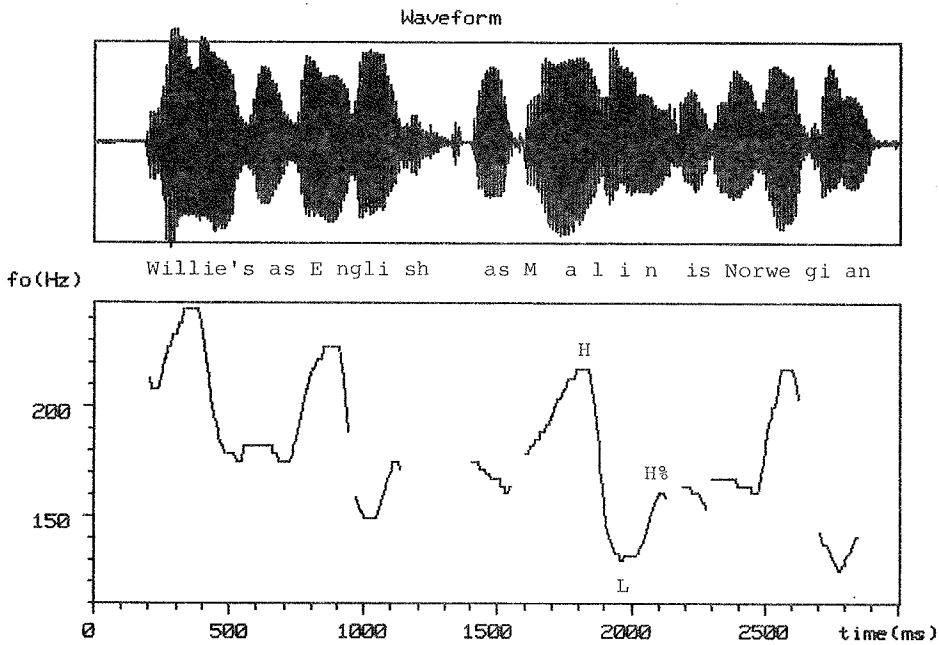


Figure 2. F_0 contour of sentence (13b) showing intonational phrase boundary (%) after the HL+H contrastive contour on Malin. It is this boundary which can account for the noncontraction of the copula is by blocking cliticization of is to Malin.

Selkirk, on the other hand, attempts to explain the absence of contraction by an ad hoc rhythm rule which adds a 'silent demibeat', a rhythmic disjuncture, between a constituent and the second member of a 'focus pair'. Thus as regards sentence (13b), the demibeats would be placed between is and Norwegian, as well as between as and Malin. These demibeats then are assumed to function to prevent destressing of a preceding copula by Monosyllabic Destressing through the effect of the 'Grid Culmination Condition'. This would produce the correct output in the case of is but, as Selkirk herself notes (1984:379), the rule is "still too crude: it would insert a silent demibeat before Mary in (13a) thus wrongly preventing the destressing of than". We feel that the

contraction phenomena in comparatives can better be explained by relating them to the more general category of parallel sentences. The presence of an intonational phrase boundary before the copula would seem to be a more likely reason for the noncontraction than the postulation of a silent rhythmic demibeat after the copula.

Enumerative Contrastive Prominence

In addition to the cases of contrastive prominence discussed above, there are two others that should be mentioned here. One of these is contrastive prominence associated with lists of words having the same syntactic category in a given constituent filling a given grammatical function, e.g. (14) (examples taken from Chomsky 1971 and Erteschik-Shir & Lappin 1983):

(14) a. Mary is neither WILLING to please, nor EAGER to please, nor INCLINED to please

b. I dusted UNDER the table, on TOP of it, and BESIDE it

c. I went to the store and bought SUGAR, EGGS, FLOUR, and HONEY.

d. You can buy curtains at MACY'S, SEAR'S, or GIMBLE's.

These sentences are characterized by the fact that the F_0 contours on all but the last constituent can typically be characterized as either 'low rises' (LH) (Ladd 1978), L* H- H% contours 'Pierrehumbert 1980) or 'fall rises' (HL+H), whereas the last contour has a 'fall' (HL) pattern. The choice of the

LH or HL+H contour on the contrasted elements would seem to depend on whether the contrasted items can be grouped into the same superordinate category. If they can, as in (14c), where the listed words belong to the class of FOODSTUFFS, or in (14d), where the enumerated items belong to the category of DEPARTMENT STORES, then they are assigned a HL+H contour. Assignment of enumerated phrases to the same superordinate category constitutes an instance of the meaning of the HL+H tone ("focus within a given set") discussed in Ladd (1978:153ff). On the other hand, if the listed phrases cannot be grouped into the same category, as in (14a,b), then they are assigned a LH contour. The LH tone could thus perhaps be associated with a meaning such as "focus within different sets". The fall on the last contrasted item signals the end of the list and also the end of the constituent associated with a particular grammatical function. The rule for iterative assignment of F_0 contours within a particular grammatical function can then be informally expressed as in (15):

(15) Contrastive Prominence in Parallel Phrases:

If the constituent realizing a given grammatical function contains iterative enumeration of a given type of syntactic phrase (NP, VP, AdjP, PrepP, AdvP), assign all words realizing the nonidentical referents of the categories making up the prefinal phrases a low-rise (LH) F_0 contour if the contrasted words do not belong to the same superordinate category and a HL+H contour if they do. Assign the final phrase noncoreferential elements a falling (HL) contour.

Following in Figure 3 is an example of how the contrastive F_0 contour on sentence (14a) is realized, naturally (a), and in a synthesized version (b). The low rise contour on the enumerated prefinal constituents was synthesized so that the low on the stressed syllable of 'willing' and 'easy' was placed at the bottom of the speaker range (150Hz). The rise following the low was synthesized as a straight line covering the identical parts of the repeated constituents; it reached a constant height of about 260Hz. Resetting of F_0 at the beginning of each intonation phrase was fixed at the usual sentence initial starting point for the speaker, i.e. 200Hz. More data is needed in order to make any conclusive generalizations about the realization of these contrastive contours. However, it would seem to be a fairly straightforward matter to expand the rule system of our prosody model in order to allow their synthesis.

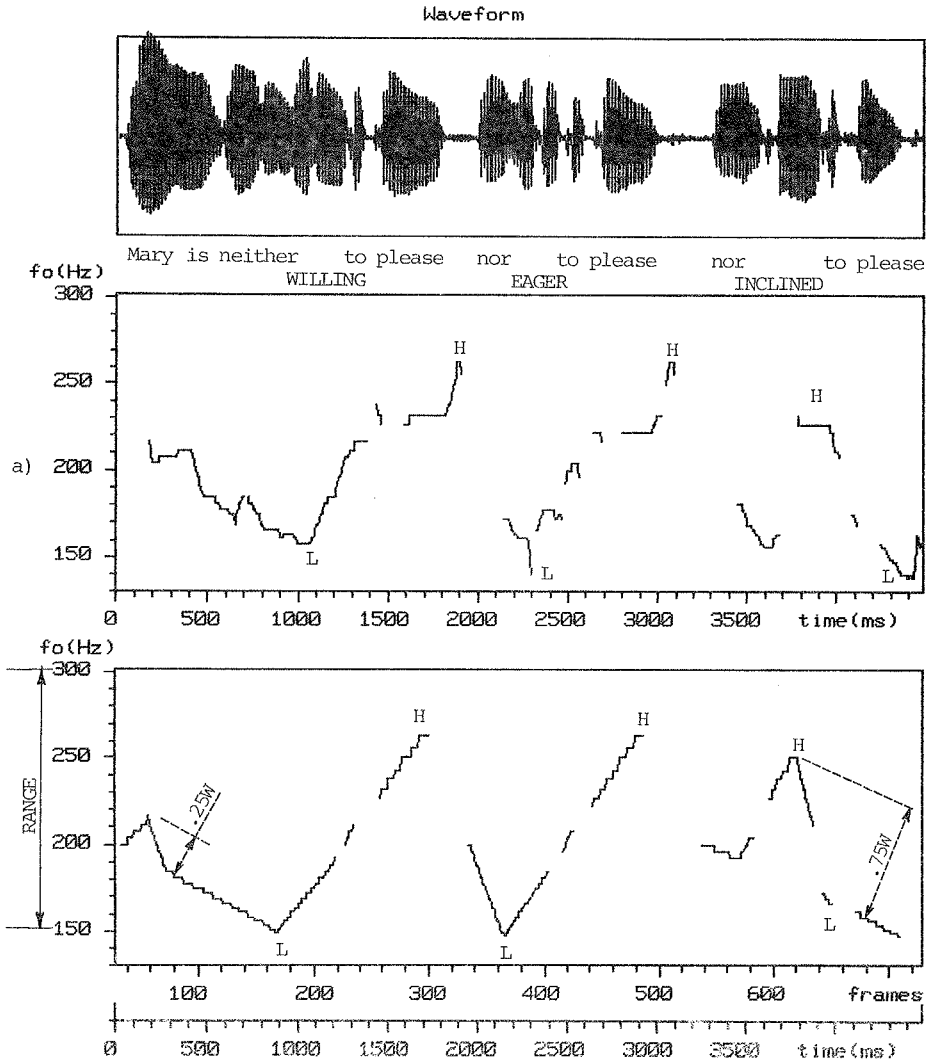


Figure 3. Occurring (a) F_0 contour of sentence (14a) and synthesized (b) contour. Notice the LH contours on the prefinal contrastively enumerated constituents and HL on the final constituent. In the synthesized version, the L's on the stressed syllables were placed at the bottom of the speaker range (150Hz). Resetting of F_0 at the beginning of an intonation phrase was set at 200Hz (normal starting point). Prominence corresponding to .25W was placed on Mary in accordance with the Phrasal rule discussed in Horne 1987.

Contrastive Prominence Conditioned by a Change in Polarity

The final case of contrastive prominence we will discuss here is that conditioned by a change in polarity, marked by words such as not, rather, or instead as in (16) (examples taken from Bolinger 1961, Chomsky & Halle 1968 Erteschik-Shir & Lappin 1983, Malmberg 1952):

(16) a. John gave MARY, not SALLY or ESMERALDA, the book

b. John RAN, rather then WALKED or JOGGED, up the hill

c. We went to England AFTER the war, not DURING the war

d. Instead of ENcouraging her, they actually DIScourage her

e. This whiskey was not EXported from Scotland; it was DEported

f. I said REarm, not UNarm.

In these examples, it is seen that the material within the scope of the polarity marker is contrasted with that in other occurrences of the same repeated grammatical category within a given syntactic phrase, both within a given sentence and intersententially, where the sentences are otherwise syntactically and referentially identical. Intrasententially, this case of contrast differs from those discussed above in (14) in that the polarity change triggers a somewhat different F_0 pattern. Here, the nonidentical material within the scope of the polarity marker in the repeated category receives a HL+H contour, that not in the scope of the marker, a HL contour. Thus, it is not necessarily the case that the HL contour comes

at the end of the last contrasted word as in (14). In (16f), the HL is on the prefinal contrasted word. The rule for assigning this 'polarity change' contrastive prominence can be provisionally formulated as in (17):

(17) Polarity Change Contrastive Prominence:

In the environment of referentially different realizations of the same grammatical phrase marked for opposing values of polarity and filling a given grammatical function, assign a HL+H contour to the words realizing the contrasting referents within the scope of the negative polarity marker, and a HL contour to those outside the scope of the polarity marker.

Notice that, in order for this rule to apply properly, words such as import and deport would have to have phonological representations with a word boundary between the prefix and the root, e.g. $(\underline{ex}\#\underline{port}_{(i)})_{(j)}$, $(\underline{de}\#\underline{port}_{(i)})_{(k)}$ so that the contrastive prominence will get placed on the affix instead of the root. This would seem motivated, since these affixes are usually classified as independent lexical entries.

In conclusion, it can be said that even these cases of contrastive prominence seem to be based on syntactic parallelism, although it is only words or phrases which, on the surface, are treated as parallel. We assume, moreover, that all the rules for assigning contrastive prominence apply before those which assign prominence associated with information focus (see Footnote 2).

Conclusion

The goal of this paper was to shed some light on the factors that condition patterns of contrastive prominence. It has been seen that the notion of syntactic parallelism interacting with referent identity can be used to explain three different instances of contrastive prominence, each associated with a particular intonation contour. Preliminary rules making reference to grammatical information and identity relations were proposed for assigning contrastive prominence in the different cases. These rules are assumed to apply before the regular focus-assigning rules apply. Suggestions were also made as to how the associated F_0 contours could be synthesized. Furthermore, by assuming the insertion of an intonation phrase boundary at the end of a given constituent containing a contrasted word, data involving noncontraction of the copula in English were also explained.

Footnotes

1. We have not discussed prominence associated with 'focus governing words' either. These are words like also, too, indeed, that fall outside the grammatical categories of subject, predicate, predicate complement which lie at the basis of our focus-assigning model, but yet attract prominence. See Gussenhoven 1983:381 and Altenberg 1987 for a discussion. These words could perhaps be assumed to be lexically marked for receiving a certain degree of prominence.

2. Our model for assigning information focus as well as its projection onto constituents and its realization as F_0 can be summarized in the following flow-diagram:

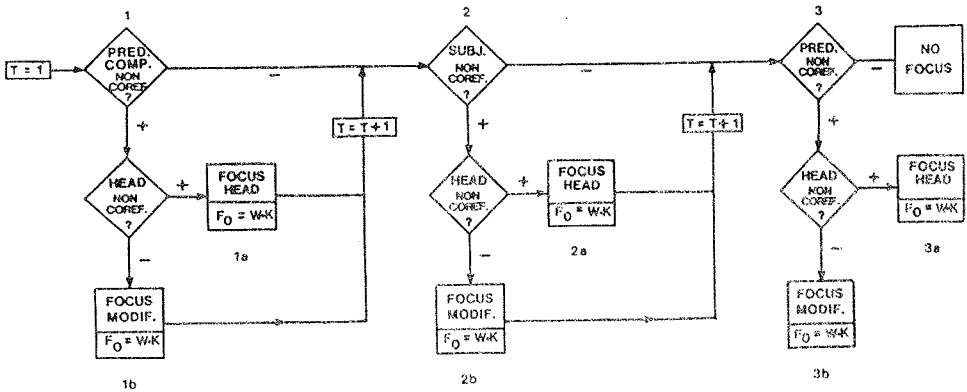


Figure 4. Model (flowchart) for assigning information focus to constituents on the basis of grammatical functions and the coreferential status of the lexical material realizing a particular function. The input to the model is a given sentence (S). Focus is realized as pitch (F_0) according to the equation $F_0 = W.K$ where F_0 here refers to the relative height of a given pitch obtrusion, W designates the width of the grid within which F_0 moves and K is a variable ranging over a number of prominence levels defined as fractions of the distance from the baseline to the topline of the grid. In Fig. 2, K assumes the values 1 (for the first focussed constituent), 0.8 (for the second focussed constituent), and 0.4 (for the third focussed constituent). For the syntheses done in Paper IV, however, the values were 1, 0.75, and 0.5, respectively. The box, T=T+1, is a counter which adds 1 each time focus is assigned. T is used in determining the coefficient, K: If T=1, -> K=1, if T=2 -> K=0.75, and if T=3 -> K=0.5. The diagram is to be read as follows: 1.: check to determine if there is a predicate complement that is non-coreferential with something in the preceding part of the text. If there is one, check whether it is the head that is non-coreferential. If this condition is met, focus the head, assigning it a level of prominence where $F_0 = W.K$ (1a). If the head is coreferential, assign the modifier focal prominence instead (1b). Go to the subject (2) and repeat the same routine, and then go to predicate (3), again repeating the same routine.

References

Altenberg, B. 1987. Prosodic patterns in spoken English. Lund Studies in English 76. Lund: Lund University Press.

Bing, J. 1980. Aspects of English prosody. Bloomington: Indiana Univ. Linguistics Club.

Bolinger, D. 1958. A theory of pitch accent in English. Word 14: 109-49.

_____ 1961. Contrastive accent and contrastive stress. Language 37: 83-96.

Chomsky, N. 1971. Deep structure, surface structure and semantic interpretation. In: D. Steinberg and L. Jakobovitz, Semantics: an interdisciplinary reader in philosophy, linguistics and psychology, 183-216. Cambridge: Cambridge University Press.

_____ and Halle, M. 1968. The sound pattern of English. New York: Harper & Row.

Crystal, D. 1969. Prosodic systems and intonation in English. Cambridge: Cambridge University Press.

Enkvist, N. 1979. Marked focus: functions and constraints. In S. Greenbaum, G. Leech, and J. Svartvik (eds.), Studies in English linguistics for Randolph Quirk, 134-52. London: Longmans.

Erteschik-Shir, N. and Lappin, S. 1983. Under stress: a functional explanation of English sentence stress. J. of linguistics 19:419-53.

Gussenhoven, C. 1983. Focus, mode and the nucleus. J. of linguistics 19:377-417.

Horne, M. 1985. English sentence stress, grammatical functions and contextual coreference. Studia linguistica 39: 51-66.

_____. 1986a. Information focus: assignment and phonological implications. In L. Evensen (ed.), Nordic research in text linguistics and discourse analysis, 155-70. University of Trondheim: TAPIR.

_____. 1986b. Focal prominence and the 'phonological phrase' within some recent theories. Studia linguistica 40: 101-21.

_____. 1987. Towards a quantified, focus-based model for synthesizing sentence intonation in English. Working papers (Lund Univ., Dept. of Linguistics) 31: 41-86 (also to appear in Lingua).

Ladd, D.R. 1978. The structure of intonational meaning. Bloomington: Indiana Univ. Press.

_____. 1983. Phonological features of phonological peaks. Language 59: 721-59.

Lakoff, G. 1971. The role of deduction in grammar. In C. Fillmore and D. Langendoen (eds.), Studies in linguistic semantics, 62-70. New York: Holt, Rinehart and Winston.

Malmberg, B. 1952. Engelsk ljudlära. Stockholm: Almqvist & Wiksell.

Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. Ph.D. Diss., M.I.T.

Selkirk, E. 1984. Phonology and syntax: the relation between sound and structure. Cambridge, Mass.: MIT Press.

Sigurd, B. 1984. Computer simulation of spontaneous speech production. In Proceedings of COLING 1984, 79-83. Assoc. for Computational Linguistics.

_____. 1987a. Referent grammar (RG) in computer comprehension, generation and translation of text (SWETRA). Working papers (Dept. of Ling., Univ. of Lund) 31: 137-67.

_____. 1987b. Referent grammar(RG)--a generalized phrase-structure grammar with built-in referents. Studia linguistica 41:2 (to appear).

Vanderslice, R. 1968. Synthetic elocution. Working papers in phonetics 8, UCLA.

Werth, P. 1979. If linear order isn't in the base, then where is it? The role of focus in word order. In J. Meisel and M. Pam (eds.), Linear order and generative theory, 187-251. Amsterdam: John Benjamins B.V.