LUND UNIVERSITY

DEPARTMENT OF LINGUISTICS

General Linguistics
Phonetics

WORKING PAPERS

31. 1987

FOR EVA GÅRDING

LUND UNIVERSITY

DEPARTMENT OF LINGUISTICS

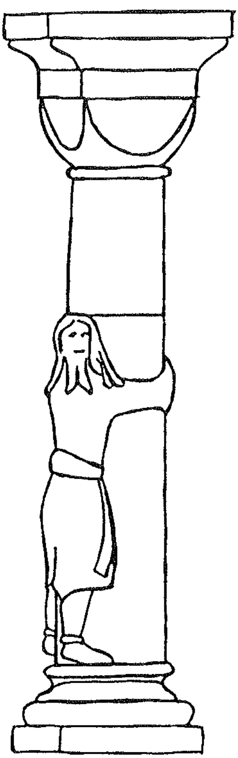
General Linguistics
Phonetics

WORKING PAPERS

31. 1987

FOR EVA GÅRDING

ISSN-0280-526X



Working Papers
Department of Linguistics and Phonetics
Helgonabacken 12
S-223 62 LUND

This issue has been edited by Gösta Bruce,
Bengt Sigurd and Janna Geggus

PREFACE

Volume 31 of Working Papers from the Department of Linguistics and Phonetics of Lund University honours EVA GÅRDING, Professor Emeritus since 1986. It includes contributions from her as well, which provide evidence that she is still very active in research.

Working Papers 30 was Kjell Norlin's PhD thesis: A Phonetic Study of Emphasis and Vowels in Egyptian Arabic, which he defended on June 3. That thesis illustrates cooperation between our department and the Department of Middle East Languages. The present volume (31) includes further documentation of the research carried out at our department.

The annual national Summer School in Speech, Sound and Hearing was arranged by the Department of Linguistics and Phonetics and the Department of Logopedics and Phoniatrics in 1986 and focused on computers in speech, sound and hearing research. It was held at Bollerup Castle and was organized by Christina Dravins and David House.

The Fourth International Congress for the Study of Child Language was organized in July 1987 by Ragnhild Söderbergh and the Child Language Research section of the department. Its theme was: The Active and Creative Child. The congress was held in Lund and attracted around 300 participants, including the leading scholars in the field.

Members of the department have participated in conferences and congresses in many different countries. At the same time, several researchers have visited the department and presented their work, e.g. Wim de Geest (Brussels), Julie Feilberg (Trondheim), Jan Katlev (Copenhagen), Wolfgang Schulze (Bonn), Björn Lindblom (Stockholm), Nina Thorsen (Copenhagen), Ailbhe Ni Chasaide (Dublin), Östen Dahl (Stockholm), Kamal Prakash Malla (Nepal), Suzanne Romaine (Oxford), Arendse Bernth (Copenhagen), Eva Hajicova (Prague) and Geoffrey Leech (Lancaster).

Thanks to a grant from the University a group of PhD students and teachers from the department visited London and Edinburgh in April 1987. Papers were presented by our group and the hosts at the School of Oriental and African Studies and the Phonetics Department at London. In Edinburgh, where two colleagues, Robin Cooper and Elisabet Engdahl, have received new appointments, the group met with members of the Linguistics Department and studied the Alvey research program at the new Centre for Speech Technology Research.

The year's teaching included a new course in Computational linguistics taught by Eeg-Olofsson, newly appointed lecturer in computational linguistics, and Sigurd. The laboratory has bought more computers (PC and Macintosh). Lars Eriksson has succeeded Mandersson as research engineer, and is presently engaged in developing the signal processing, communication and printing facilities.

The department has received support for some new projects: Computer comprehension, generation and translation of text (SWETRA) (HSFR: Sigurd, Eeg-Olofsson, Gustafsson), Linguistic aspects of reading and spelling development (HSFR: Magnusson, Nauclér), Languages of the world survey (Lundbergska IDO-fonden: Svantesson, Vamling, Ohlsson). Ongoing projects are Automatic prosodic analysis for Swedish speech recognition (STU: Bruce, House and in cooperation with the Department of Linguistics, University of Stockholm), Text linguistic analysis of computer manuals (HSFR: Tengberg) and From text to prosody (RJ: Bannert). These projects illustrate some of the directions of the current research at the department.

Gösta Bruce   Bengt Sigurd

CONTENTS

Eva Gårding	How many intonation models are there in Lund?	1
Eva Gårding and Jialu Zhang	Tempo and prosodic pattern in Chinese and Swedish	11
Merle Horne	Towards a quantified, focus-based model for synthesizing sentence intonation in English	41
David House, Gösta Bruce, Francisco Lacerda and Björn Lindblom	Automatic prosodic analysis for Swedish speech recognition	87
Yasuko Nagano-Madsen	Effects of tempo and tonal context on fundamental frequency contours in Japanese	103
Bengt Sigurd	Multilanguage translation of numerals by Prolog	117
Bengt Sigurd	Referent grammar (RG) in computer comprehension, generation and translation of text (SWETRA)	137
J-O Svantesson	A note on the duration of Chinese questions	169
J-O Svantesson	Some data on the duration of Chinese stops and affricates	171
Madeleine Wulffson	Discourse intonation in Swedish: an exploratory study in communication games	175

HOW MANY INTONATION MODELS ARE THERE IN LUND?

Eva Gårding

(Paper presented at a phonetics conference in Uppsala, October 1986)

In this review of works on intonation which have emanated from Lund during the last two decades, my joint paper with Gösta Bruce from 1978 will serve as a vantage point from which I am going to look backwards and forwards according to the sketch presented in Figure 1. My review will end with an answer to the question given in the title.

Our paper of 78 is entitled "A prosodic typology for Swedish dialects". The title would have been more appropriate, had it shown that the typology was expressed in terms of a general, generative intonation model. Figure 2 illustrates how this model works for a neutral statement. The input is a sentence for which the segments and their durations are given and the intonation is missing. It is equipped with phonological markings for word prosody and sentence prosody and has in addition a label for dialect according to the four categories established earlier (Gårding 1970, Meyer 1954). This input sentence passes through an ordered set of generative rules, of which some differ according to dialect. The result is an intonation curve characteristic of the dialect marked in the input.

My comments about the rules will follow the order given by

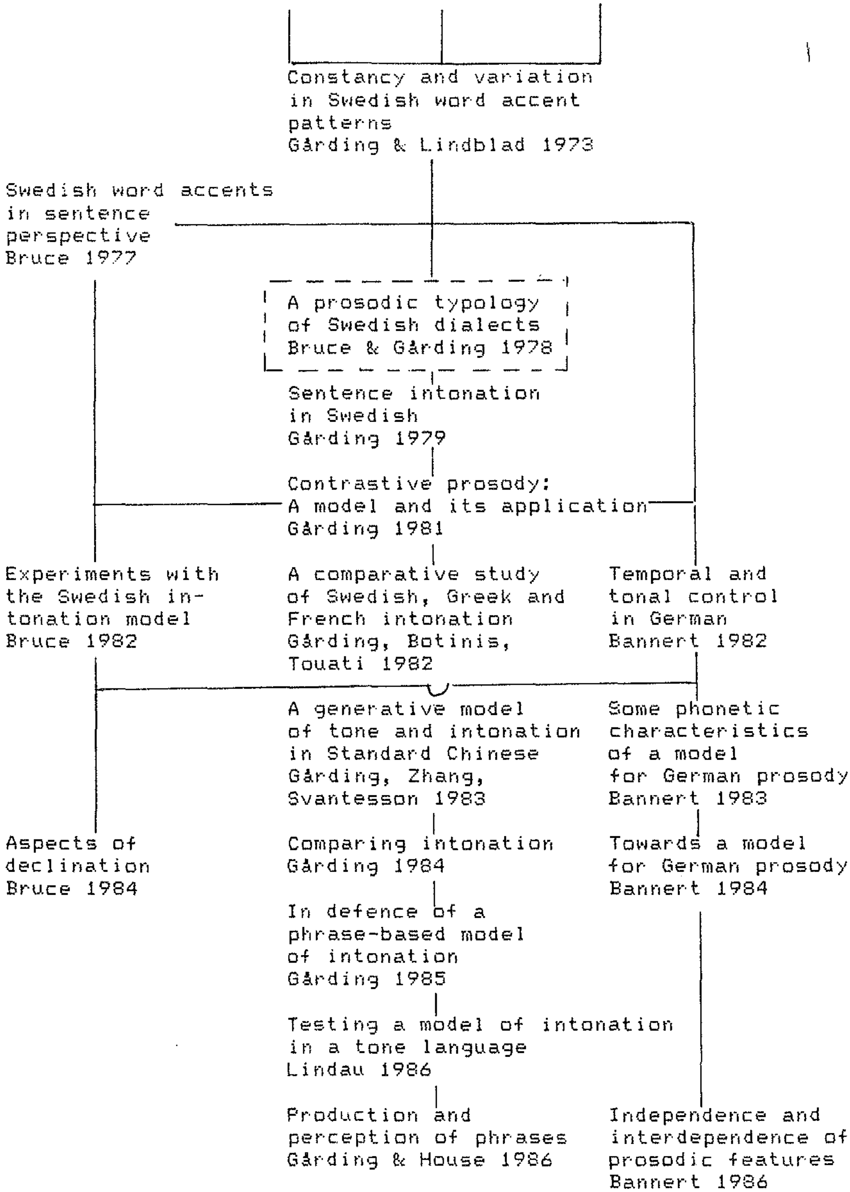


Fig.1 Interrelations between intonation analyses from Lund

For a complete bibliograhy see Gårding: Swedish Prosody, Summary of a project, Phonetica 39 (1982) and recent issues of Working Papers from the Department of Linguistics of Lund University.

Figure 2. In the paper of 78, the sentence intonation of rule 1 has for the first time received an autonomous global expression. (It will later be called tonal grid.) In Gårding and Lindblad 1973 there was no global expression for sentence intonation, only a local phrase-final one, and pitch levels which specified the rises and falls required by the accents. Nor was sentence intonation a variable in Bruce's 1977 dissertation about Stockholm Swedish. The auxiliary lines of his generative rules do not express sentence intonation, they are four pitch levels similar to the intonation levels of earlier works. To these levels he relates the pitch points that are relevant for a certain accent in a certain context in a certain position in a sentence which is all the time a statement.

The word accents of rule 2 in the model of 78 are analysed as combinations of highs and lows. This representation is taken over from Bruce's analysis and gives the model an important connection with phonological analyses of African and Asian tone languages. The pitch values of the highs and lows are obtained with the aid of the auxiliary lines. They are in other words determined by the sentence intonation. Bruce had transformed his highs and lows into pitch values by means of context rules.

For the sentence accent of rule 3 we used the same elicitation procedure as Bruce had done and the same term, focus, for its phonetic manifestation. Earlier, Lindblad and I had used contrast/emphasis.

The fourth rule concerns interpolation and this rule has correspondences in both the preceding works.

I shall now give an overview of work presented after 78 and start in alphabetic order with Bannert.

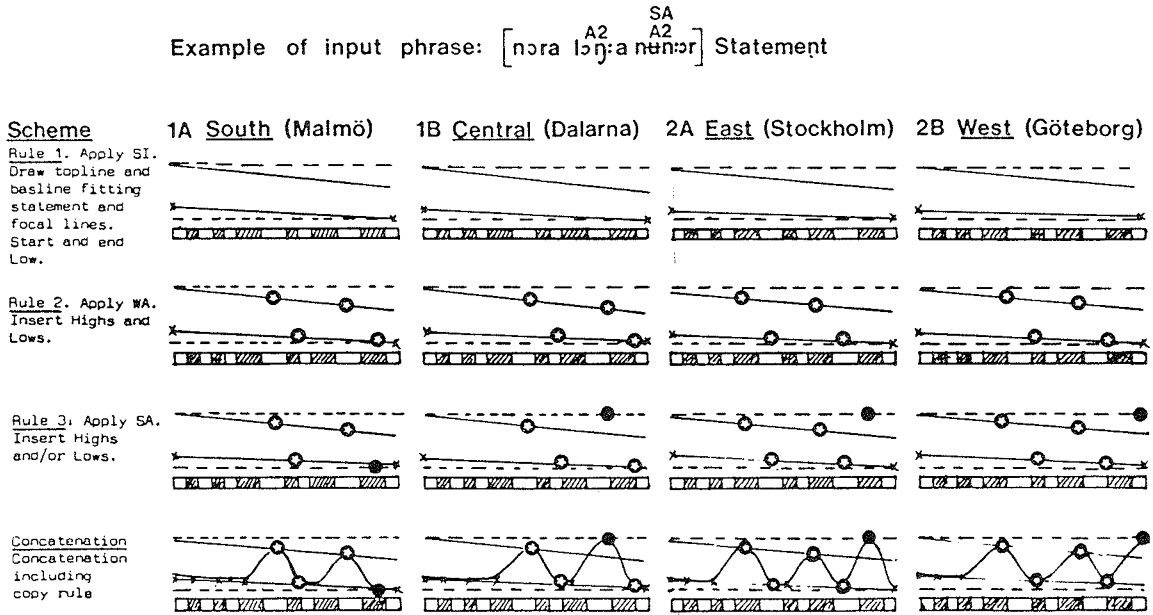


Fig.2 Intonation model for Swedish dialects. Bruce & Gårding 1978

In a paper from 1983 Bannert sketches a model of German intonation for statements and questions. As in the model of 78, the turning points of the intonation curve are generated in several steps. The first step generates the four characteristic levels of intonation (Fig.3). For statement intonation the curve is generated in the following way. All the accents are analysed as low points except the next to the last which is high. These low points, which are to become minima in the final curve, are placed on the pitch scale according to rules common to all the three intonation types analysed by Bannert. As in Gårding and Lindblad , minima and maxima of the intonation curve are fixed to certain segments and the curve is first generated as rises from minima and falls from maxima and in a final step by interpolation. Attitudes like emphasis are generated by enlarged rises. For questions the last part of the curve is modified. In a recent paper Bannert has sketched a comprehensive model which incorporates all the factors which are known to influence intonation and their mutual dependencies (Bannert 1986).

Bruce, in a paper from 1982, reports on a series of experiments in which he examines some of the predictions of the model of 78, in particular the influence of focus on the general direction of the intonation and the shape of the accents. Figure 4 shows a schematic figure summing up his results. Accents before focus are of the type rise-fall and after focus of the type plateau- fall. According to Bruce an intonation curve has both global and local features. The global aspect is apparent in the fact that the starting frequency of the first accent is dependent on the number of

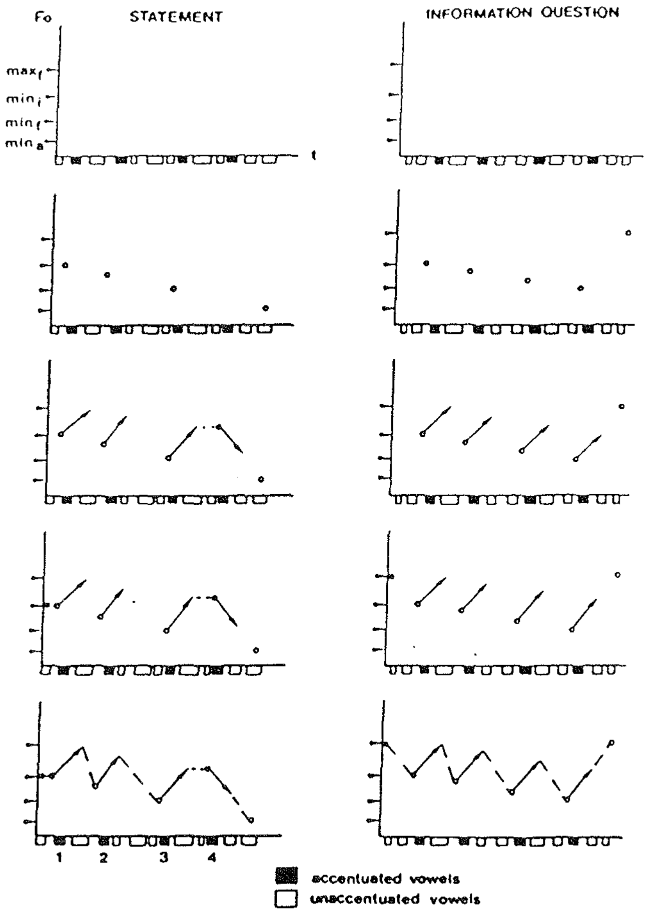


Fig.3 Intonation model for German. Bannert 1983

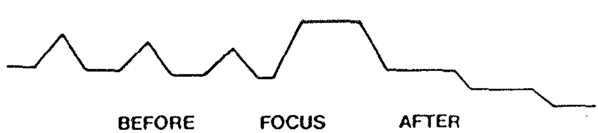
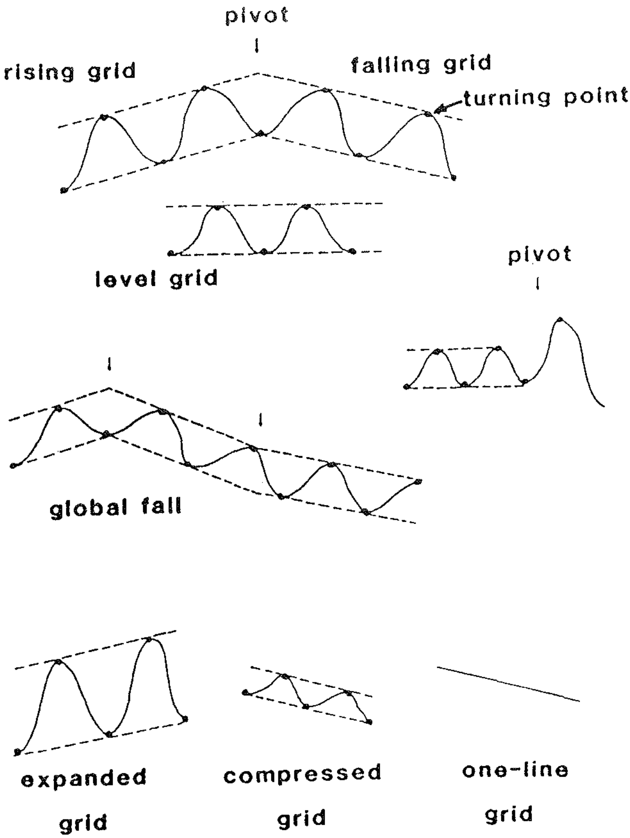


Fig.4 Downdrift in Swedish. Bruce 1982

upcoming accents and the local aspect is revealed by the constant frequency relations between neighbouring accents. He makes no attempt to incorporate these findings in a generative scheme. The generative aspect of the model of 78 has been abandoned in favor of a cautious and more descriptive attitude.

After 78 my main interest has been the interaction of sentence intonation and word intonation in different prosodic systems. Our material from the project "Swedish Prosody" showed that questions could be analysed in the same way as statements, i.e. by means of a set of auxiliary lines with a characteristic distribution for the particular sentence intonation and with the word accent rules unchanged (Gårding 1978).

The same method of analysis was tried in collaboration with Antonis Botinis and Paul Touati for Greek and French and in connection with the project "Phonetic descriptions of some non-European languages" for Hausa, Arabic and Standard Chinese in collaboration with Mona Lindau, Kjell Norlin and Jan-Olof Svantesson. Special attention has been paid to the exploration and definition of the tonal grid. A grid encompasses normally stressed accents and tones and shows the general direction of phrase and sentence intonation. Another concept connected with the grid is the pivot (Gårding 1982, Gårding, Botinis and Touati 1983) which denotes a place where the grid changes direction or width. The location of pivots is correlated to the syntactic structure or the information structure of the sentence. An important notion in my analysis which has been further examined is the turning point fixation which means that certain turning points in an intonation curve are fixed



Intonation parameters	Function Semantic	Syntactic
turning points	words, morphemes	d:o
pivots	constituents (theme/rheme)	d:o (subject/predicate)
grid:direction	speech act type	sentence type
grid width, position	information weight (focus)	clause type

Fig 5. Concepts of the model and their communicative functions. Gårding 1984

to the segments and the grid in a regular way. Figure 5 shows the main concepts of the model and their communicative functions. Our analysis of different prosodic systems with the same method brings out the similarity across languages of the manifestation of speech act and makes it possible to compare the local manifestations bearing on tone and accent.

Mona Lindau applied the model to Hausa, a two-tone language, as part of our project work. In a paper from 1986 she presented an algorithm which generates statements and questions according to the main principles of the model: Sentence intonation is separated from word tones by means of a tonal grid. Tonal movements, rises or falls according to the tone, are assigned to the syllables via tonal points inserted into the grid.

Let me now finally return to the title of the paper: How many intonation models are there in Lund? All of them make use of turning points for the generation of intonation curves. In this respect there is one model. However, only in the model that I advocate are the local features of intonation subordinated to the global ones. For this reason it might be justified to say that there are two models, one with a grid which brings out the intonation planned and executed over phrase and sentence and one in which intonation is seen as a result of local rules or decisions.

TEMPO AND PROSODIC PATTERN IN CHINESE AND SWEDISH

Eva Gårding and Jialu Zhang

Abstract

This paper is a study of how speech rate influences temporal and tonal patterns in Standard Chinese. It includes a comparison with an earlier, similar study of Swedish. The material, sentences with tonal (accentual) patterns in focus-free statement intonation, was produced by several speakers under three rates of speech, labelled Normal, Fast and Slow. The main effect of a shift of tempo from slow to fast is a compression and an upward movement of the overall pitch range while the number of turning points and their positions relative to the segments are well retained. In the time domain there is a lengthening of about 50% from normal to slow and a shortening of about 25% from normal to fast speech. This compression is not uniform. For the higher tempo the last constituent increases in relative duration and prominence at the expense of the segments of the first constituent. The comparison with Swedish shows basic similarities. Observed differences can be explained by reference to the larger number of pitch-determined syllables in a Chinese sentence as compared to a Swedish one.

Introduction

The main purpose of this study is to show how speech rate affects prosodic patterns in Chinese and Swedish.

Speech rate as a variable was used by Kozhevnikov and Chistovich with a view to investigating the organisation of articulatory movements on a neurological level (1965). By measuring durations in Russian sentences between sharply defined articulatory events at different rates of speech, they found that the segments of certain sequences, e.g. the word and the syllable, have constant relative durations. They concluded that speech rate must be below the articulatory program of such sequences.

The view of the syllable as a complex articulatory gesture was also adopted by Gårding who studied the manifestations of various boundary signals in different styles and rates of speech (Gårding 1967 p.23 ff.).

Recently a group of researchers at Haskins Laboratories have used 'gestures', i.e. a limited set of specified coordinated articulatory movements, as primitive units in their development of a speech synthesis model (Browman et al. 1986). Within this framework dynamic properties in speech caused by shifts of tempo and style are being analysed.

The suprasegmental level has received much less attention. One previous study dealt with the effect of speech rate on temporal and tonal patterns in different Swedish dialects (Gårding 1975). With support from an electromyographic study (Gårding et al. 1975) the notion of laryngeal gestures and their timing was used to explain the effect of tempo on the pitch curve. Our present paper deals with Standard Chinese and a comparison wih Swedish makes it possible to sort out general tempo effects in the speech signal from phonological peculiarities caused by tones or accents. It also shows that the notion of laryngeal gestures and their overlap in fast speech is fruitful to explain tonal targets and tempo effects in Chinese as well as Swedish.

After a presentation of the material and the method which are similar for both analyses the main section will be devoted to our recent analysis of Standard Chinese. The results are compared with observations in the earlier Swedish investigation.

1. Material, method and terminology

For Swedish the material consisted of various noun phrases in which the noun and the preceding adjective had four combinations of accent 1 and accent 2 (see Gårding 1975). For Chinese the following two sentences were used:

(1) Sùng Yán mài niúròu 'Sung Yan sells beef' with alternating falling and rising tones

(2) Sùng Yàn mài mòli 'Sung Yan sells jasmin' with falling tones only, the last syllable being atonic.

Phonologically the falling tone is analysed as HL (high low), and the rising tone as LH (low high)(see Gårding et al. 1983). The abbreviations may be regarded as the targets of a particular tone.

Behind the tonal movements there is a laryngeal gesture, i.e. a coherent cycle of activity of some laryngeal muscles in particular the cricothyroid (for an EMG investigation of Swedish accents, see Gårding et al. 1975).

The Chinese informants represented Standard Chinese and the Swedish informants three well-defined dialects. The test sentences were read from cards in focus-free statement intonation at three rates of speech, labelled Normal, Fast and Slow. The instructions were: Start with normal. For Slow think of a situation where you have to dictate slowly without ruining the coherence of the sentence. For Fast imagine that we have very little tape left. The test sentences were read three times by the informants.

In the acoustic analysis, segment durations and fundamental frequency were measured from broad- and narrow-band spectrograms. The Chinese material was also analysed by speech wave, intensity and fundamental frequency records obtained from the digitized speech wave by means of the ILS program.

Special attention was paid to parameters and concepts that had proved useful in cross-linguistic analyses of intonation (Gårding 1981 and 1984): the turning points of the F0 curve (i.e. the local maxima and minima of the curve) with their positions relative to segments and frequency scale, the notion of turning point fixation (which means that certain turning points have rather fixed locations relative to segments regardless of spech act), and the tonal grid as an expression of sentence intonation. A tonal grid is a set of near-parallel lines that enclose normal-size tones or accents.

In our analysis we shall make a distinction between 'rhythmic' which refers to the subjective impression of the time relations of a pattern and is transcribed by using marks for short (᾿) and long (-), and 'temporal' which refers to objective measurements.

The term sentence prominence denotes the phrase-level prominence of Chinese and sentence accent denotes the corresponding phenomenon in Swedish.

2. Chinese

Temporal pattern

Figures 1 and 2 show examples of the raw data obtained by the ILS program. With a shift of tempo there is an overall lengthening of 50% for slow speech and a shortening of 25% for the fast speech of both speakers. There is more variation in lengthening than in shortening.

In Figure 3 the segments of the utterances of the three tempi have been brought to comparable durations by normalized scaling. Since the lines connecting the segments are not vertical, expansion and compression of the normal-rate utterance have not been uniform. In other words, the time scale has been warped.

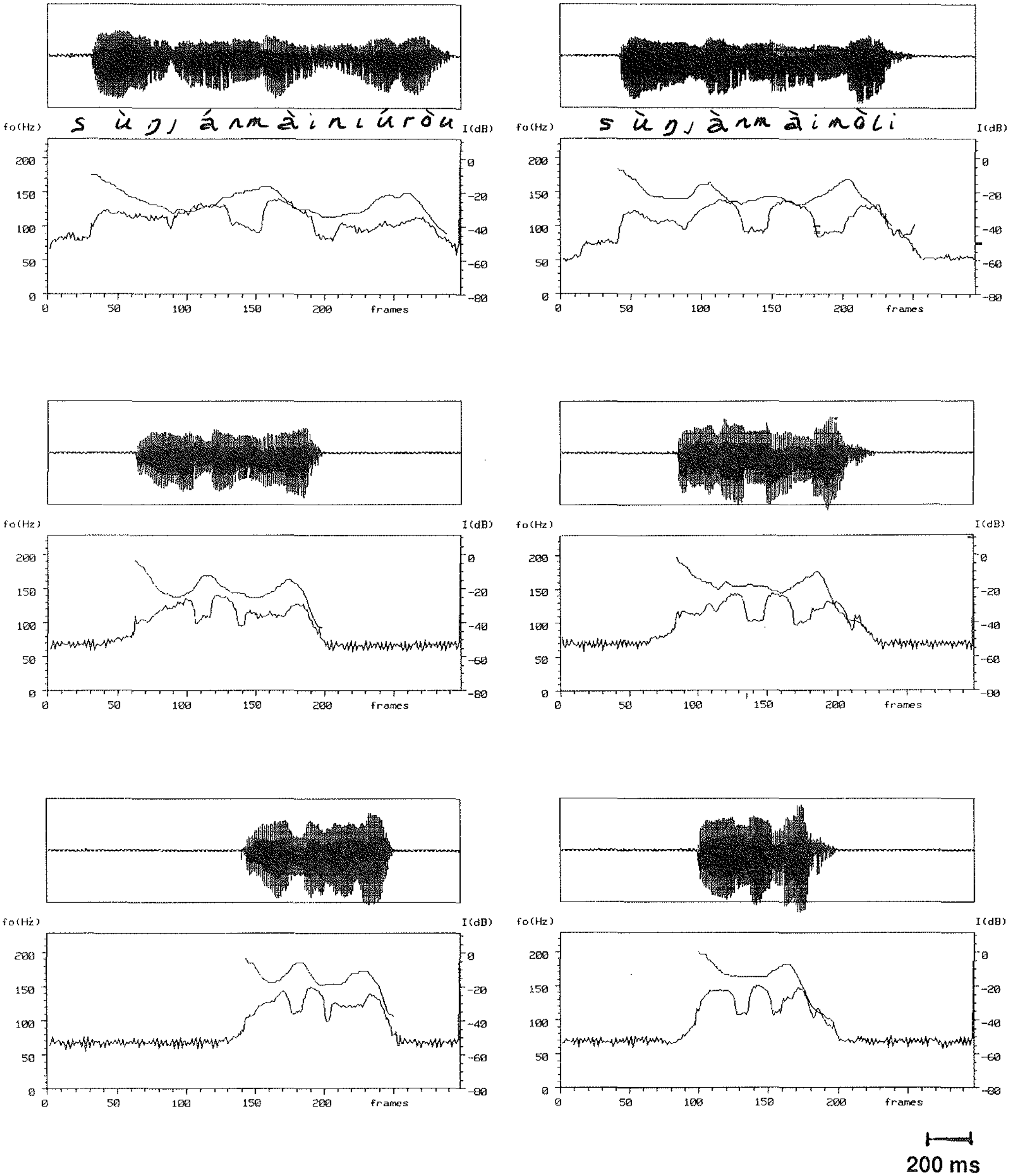


Fig.1. Waveform, FO and intensity tracings of Sùng Yán mài niúròu left and Sùng Yàn mài mòli right in three tempi, from top to bottom, slow, normal, fast. Speaker 1.

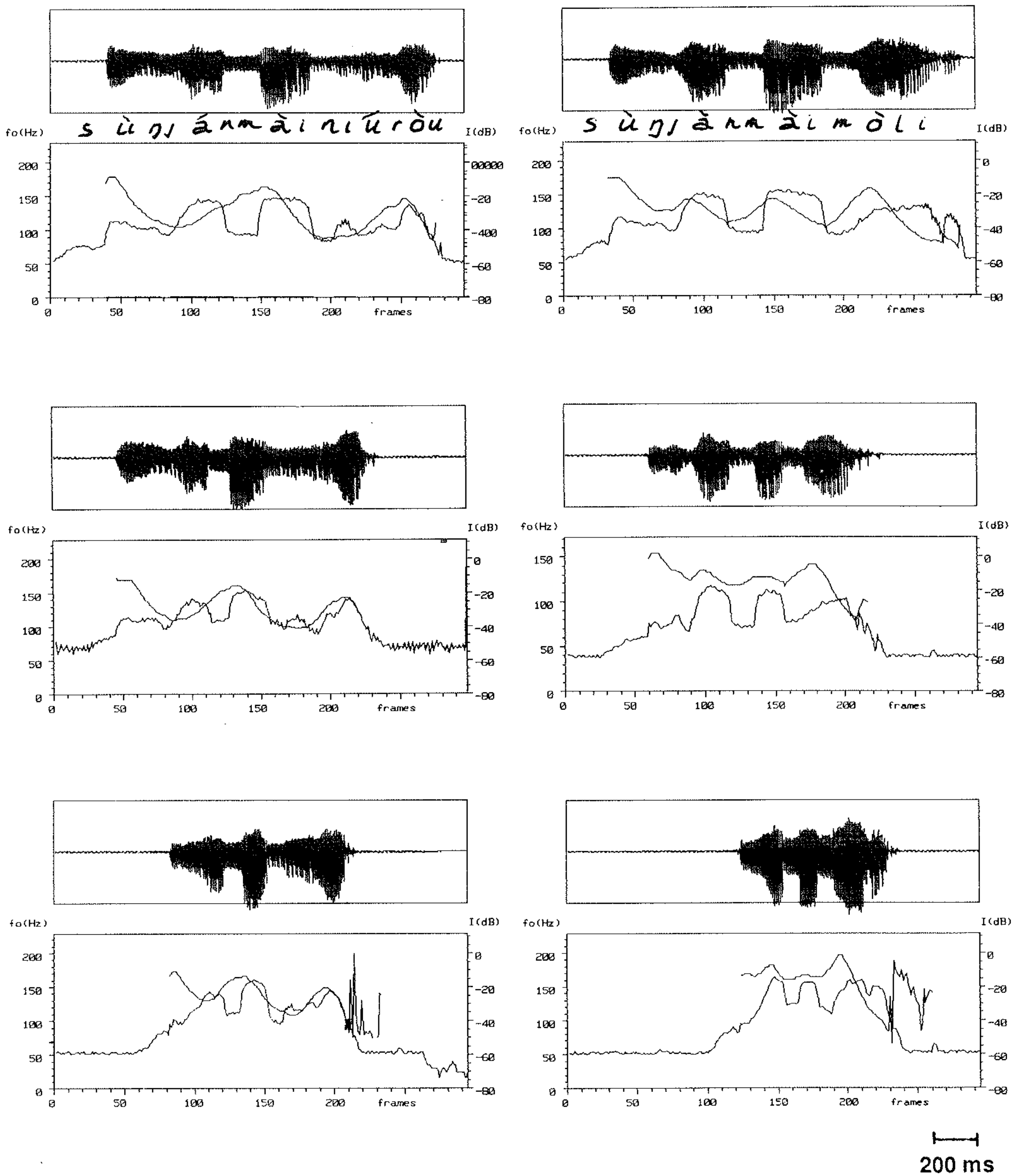


Fig.2. Waveform, F0 and intensity tracings of <u>Sùng Yán mài niúròu</u> left and <u>Sùng Yàn mài mòli</u> right in three tempi, from top to bottom slow, normal, fast. Speaker 2.

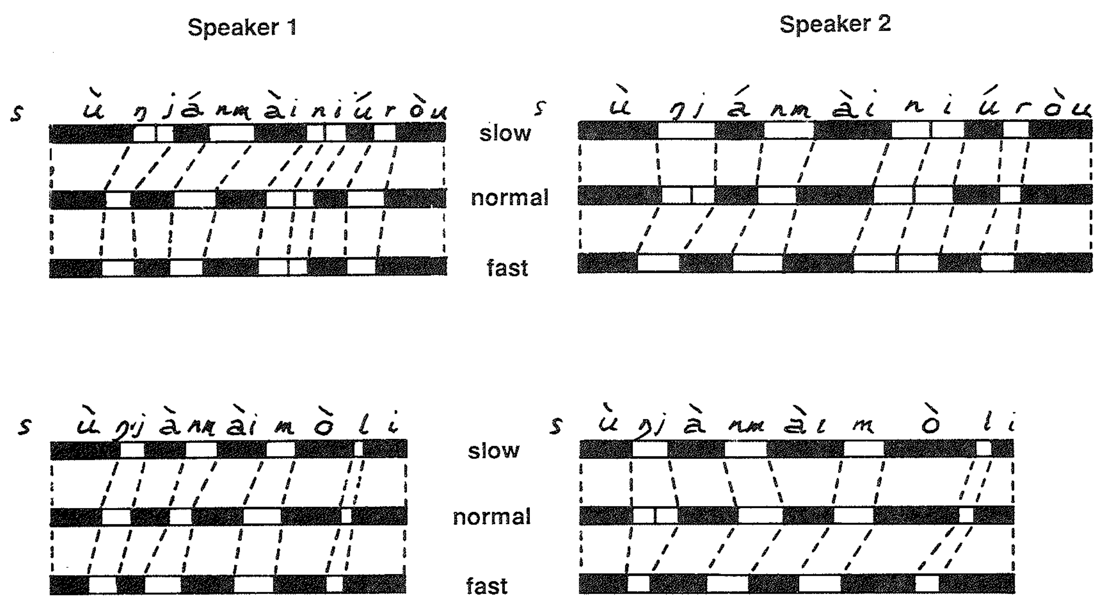


Fig.3. Relative durations of segments in three tempi, slow, normal fast. Sentence 1 is above and Sentence 2 below. Speaker 1 is to the left and Speaker 2 to the right.

With the arrangement chosen in the presentation most lines are tilted to the right, which means that with increased tempo the early part of the sentence gets more compressed than the latter part which has a corresponding expansion. In this process the first constituent of the phrase, the subject, goes down from about half of the utterance to about one third, and the last constituent, consisting of the compounds *niúròu* and *mòli*, changes correspondingly. In fast speech the compounds occupy half of the duration of the utterance. Mainly responsible for this change is probably the sentence prominence (comparable to the sentence accent in Swedish and other accent languages) which has a more important role in fast speech and is carried by the last tonal morpheme, i.e. by *-ròu* and *mò-*, according to the rules of this dialect. The predicate verb, *mài*, has a uniform part of the utterance, regardless of tempo.

For the first sentence the auditive rhythmic pattern may be represented as:

Sūng Yǎn māi niǔrōu

and for the second

Sūng Yǎn mǎi mōlǐ

The acoustic record suggests that the impression of prominence is a combined effect of relative duration, intensity and pitch range. There seem to be trading effects in neighbouring syllables belonging to the same phrase. The lengthening of *mò-* under sentence prominence may be the cause of the short duration of the preceding *mài*. (Compare the corresponding *mài* of the other sentence which is further away from the sentence prominence and longer.) The interaction of temporal and tonal patterns needs a special investigation.

Tonal pattern

In Figure 4 the normal-rate temporal pattern was used as a base and the tonal patterns of the three tempi were transposed to their respective segments via the turning points. The fast renderings are generally a bit higher and have a smaller range, at least in the falling-rising pattern.

In one case, Sùng Yàn mài mòli, Fast, the number of turning points has changed with tempo. The two falls of the name Sùng Yàn are exchanged for a single fall which makes it comparable to the compound mòli appearing in the same sentence. However, mòli has one fall regardless of tempo.

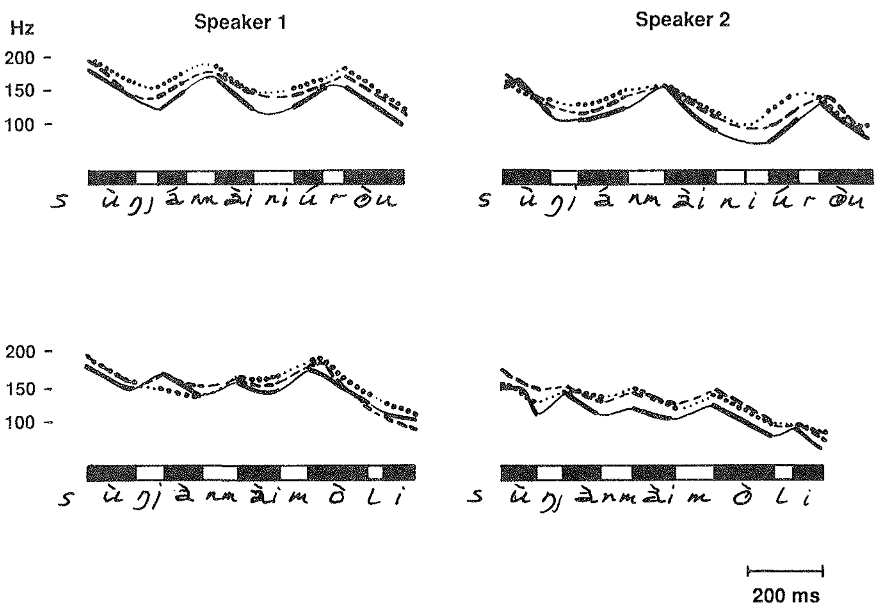


Fig.4. Tonal patterns of three tempi transposed to the normal-rate temporal pattern. Unbroken line is for slow performance, broken line for normal and dotted for fast. Sentence 1 is above and Sentence 2 below. Speaker 1 is to the left and Speaker 2 to the right.

In all other cases the number of turning points is retained and they have rather fixed locations relative to the segments. In most cases the first target of a tone is reached at the beginning of the vocalic segment in close connection with the C/V boundary. However, in fast speech there is a tendency for the first target point to be reached in the preceding consonant rather than at the beginning of the vocalic segment. See Figures 1 and 2.

There is covariation in the positions of the turning-points. This means that a contour is retained in shape even if it is shifted up or down in pitch.

In sentence 1 the falling and rising tones have about the same ranges for a particular tempo. However, in sentence 2, the sequence of falls, the middle tones have compressed ranges in accordance with a sandhi rule, stating that a falling tone becomes reduced to half size before a subsequent falling tone. This compression may be due to undershoot (see below), which can be explained by the physiological difficulty of producing two contour tones of he same kind after each other in a short time span. Since it is the vocalic segment that carries the perceptually pertinent tone pattern (Kratochvil 1968, Howie 1974, Gårding et al. 1987), the laryngeal muscles have to make a very quick manoeuvre at best in the consonant(s), in this case from the low target of one tone to the high target of the next. From this point of view, sentence 1 with its falling and rising tones gives the speaker a much easier tonal task since the second target of the first tone coincides with the first target of the next.

In Sùng Yàn of Sentence 2 (Speaker 1, Fast) there is no individual trace of the second tone and before sentence prominent mòli there is a levelling of the falling tone which may be caused by gesture overlap. Speaker 2 is not comparable in this respect since he changes the rhythmic pattern of his fast utterance.

Cut out from its context, this and similar fast utterances would be ambiguous. In real life, however, such utterances hardly exist in isolation. Coarticulation does not endanger communication in general.

Laryngeal pattern

The turning points are associated with changes in the activity pattern of the pitch-controlling muscles, in particular the cricothyroid muscles. Hence the distances between the high turning points, which have a more precise location than the low ones, indicate how tempo influences the train of laryngeal gestures. A list of these measurements (not given here) shows that the relative distance between the two peaks of Sùng Yán in Sentence 1 decreases with increased speed. In other words the gestures behind the semantically close-knit names come closer together than those controlling mài niúròu i.e. the predicate verb and the object. These results are of course just another aspect of the turning point fixation and the observed changes in the temporal pattern.

Sentence intonation (disregarding tempo)

Since the speakers have used the same declarative intonation in both sentences it is possible to abstract it from the curves.

The similarity of the six renderings of each sentence and the covariation permits us to exhibit them simultaneously in one diagram. In Figure 5, the left part of which refers to Sentence 1 produced by Speaker 1, the intonation curves are shown together with an average normal-rate temporal pattern as a base. All renderings lie within a band limited by the small vertical intervals. A curve has been interpolated through the means of each interval. We regard this curve as an expression of the target sentence intonation, i.e. the intonation that the speaker aimed at in all his productions of this sentence. This intonation is clearly falling and since the first and the last intervals are not broader than the others we can infer that the slope is determined by the length of the utterance. The combined effect of the sentence intonation and the tonal pattern can be given a concrete representation by a common tonal grid, which is illustrated by the broken lines in the figure.

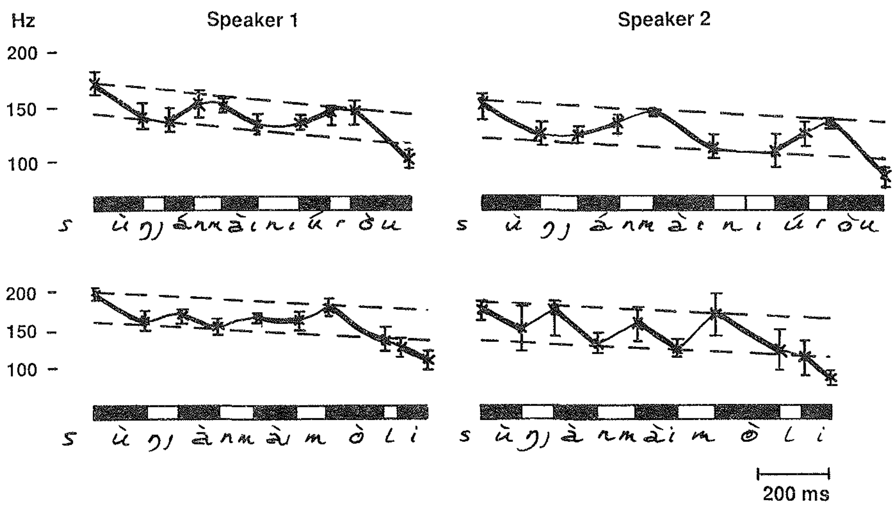


Fig.5. Intonation ranges and means of six renderings of each sentence imposed on a common normal-rate temporal pattern. All renderings lie within a band limited by the small vertical intervals. A curve has been interpolated through the means of each interval. Broken lines illustrate the tonal grid.

In the right-hand part of Figure 5, the renderings of sentence 2 have been treated in the same way and the grid of sentence 1 has been imposed. The sentence prominence is in both cases rendered by a large amplitude of the tonal movement. In the second sentence the sentence prominence falls on the penultimate syllable, the last one being atonic. In spite of the different locations of the sentence prominence the grid fits rather well, and the statement intonation used for both sentences can be expressed by the same falling grid. For sentence 2, due to the tone sandhi rule mentioned earlier, only the first and the last tones reach the topline of the grid.

A similar grid for the different renderings of the two sentences have been obtained for the other speaker. One difference is that his speech shows less coarticulation. Regardless of tempo his falling tones remain falling.

Grids adjusted to tempo

In the preceding section tempo was disregarded. When tempo is taken into account, the following small changes in the intonation curves are apparent. With increased rate of speech the entire curve is shifted up in frequency by a small amount and the range is compressed. The appendix is an attempt to simulate these modifications in a numerical generative scheme.

Summary of the Chinese analysis

Judging from this material the following prosodic features stay invariant through the three tempi.

(1) The sentence prominence.

(2) The tonal pattern.

(3) The sentence intonation.

Invariance is here taken in the phonological sense. There may be modifications in the signal but there is no loss of identity. The speaker obviously intends to retain the communicative value of a sentence even if he speaks fast.

The following phonetic modifications have been observed.

(1) Although the binary rhythmic division of the sentence into strong and weak is still present in fast speech, the temporal pattern changes so that more relative duration is given to the last constituent (niùróu, mòli) which enhances its perceptual importance and may motivate a third degree of prominence in a transcription.

(2) With a shift of tempo the position of the tonal pattern relative to the segments has a certain leeway. The first target point, represented by the first high or low turning point of the syllable is more fixed than the second one. It is closely connected with the C/V boundary but has a tendency to move leftward into the preceding consonant in fast speech. The position of the second target point is somewhat adjusted to meet the requirements of the following syllable and is therefore more variable.

There may be complete tone reduction in syntactically and semantically close knit syllables in fast speech.

(3) The overall range of sentence intonation gets smaller and moves up on the frequency scale.

The observed tempo effects are illustrated by the pitch curves of Figures 6 and 7, derived from our two speakers' slow and fast renderings of the test sentences. Below each observed curve is a schematized construction which follows the model for intonation described elsewhere (e.g. Gårding et al. 1984. The main principles are given in the appendix). The schematized curves have been imposed on the digitized originals by means of the ILS

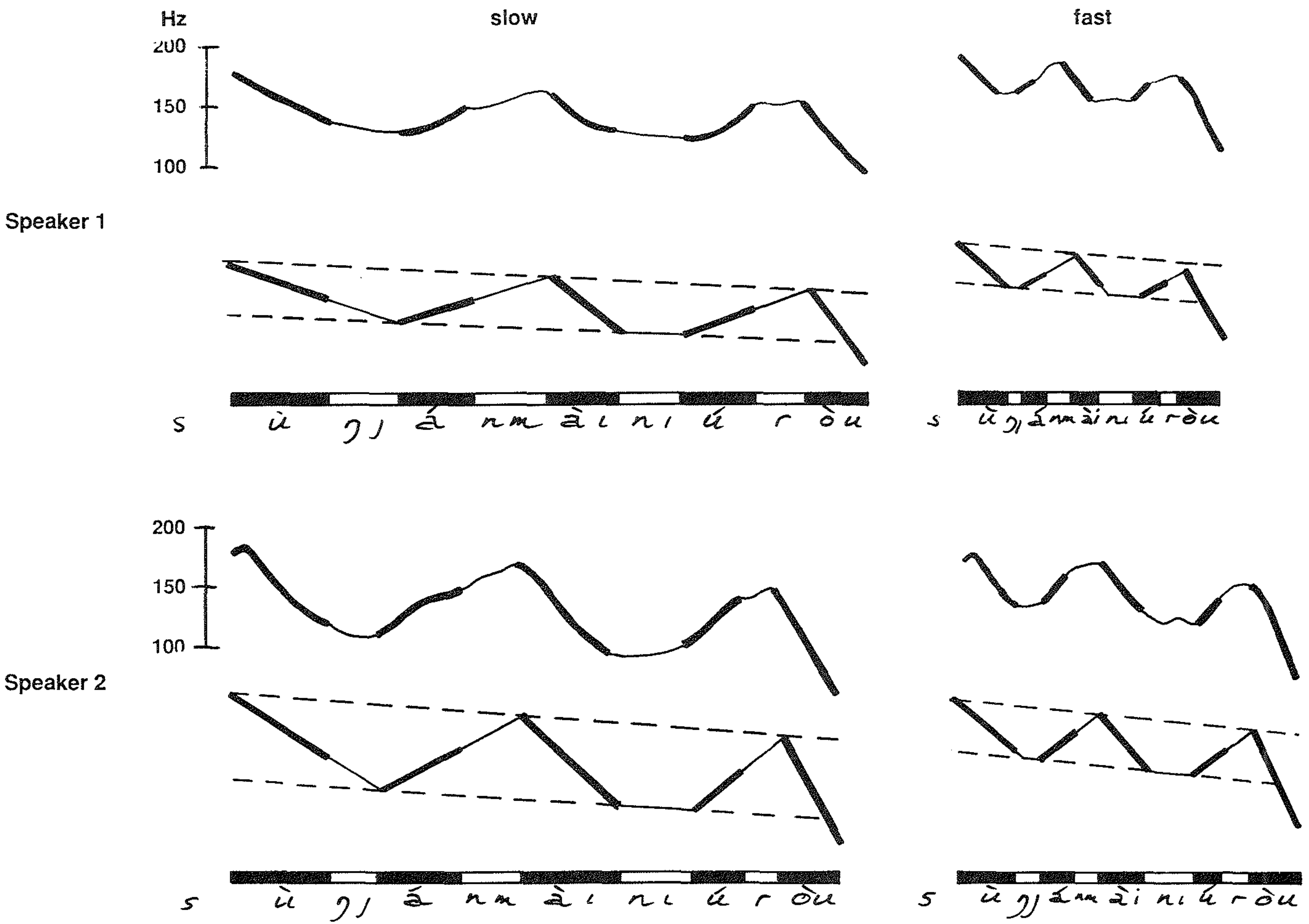


Fig.6. Observed and generated F0 curves of Sentence 1 in the slow and fast performance of Speaker 1 above and Speaker 2 below.

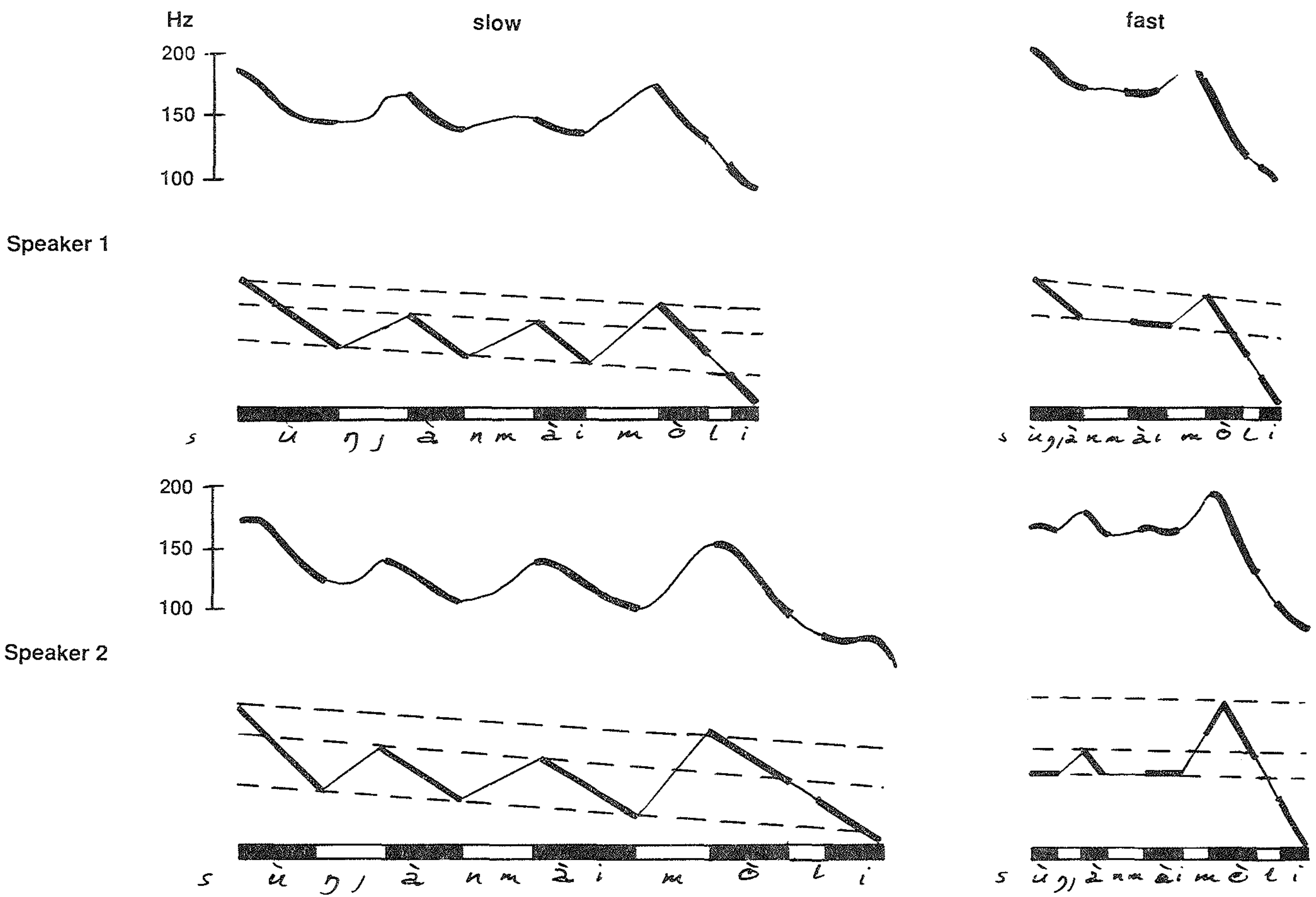


Fig.7. Observed and generated F0 curves of Sentence 2 in the slow and fast performance of Speaker 1 above and Speaker 2 below.

program. In spite of the straight-line interpolations and the other simplifications the resulting utterances are hardly distinguishable from the original ones. Another result of our ILS experiments is that the characteristics of fast speech, narrower and raised grids imposed on the normal-rate sentences, were not sufficient to create an impression of increased tempo. A change in the time domain seemed necessary.

3. Swedish

Only the most important findings in the earlier paper (Gårding 1975) will be summarized. Three dialects were considered, Southern Swedish (Helsingborg, Skåne), Central Swedish (Stockholm) and Eastern Swedish (Helsinki). Southern Swedish, being spoken on a substratum of Eastern Danish, is prosodically very different from the central dialects, e.g. the sentence accent is rising in the central dialects and falling in the south (Bruce & Gårding 1978). Lacking distinctive accents, Finland Swedish is different from both.

The sentence used for demonstration is <u>en manli(g) nunna</u> 'a masculine nun' with A2, transcribed by a grave accent mark, on both the noun and the adjective.

Figure 8 shows average durations and fundamental frequency curves of the demonstration sentence spoken at three rates of speech by representatives of three dialects.

Temporal pattern

Figure 9 shows the relative durations of the segments of the demonstration sentence in focus-free declarative intonation, spoken by three informants across three tempi. We can infer from the Central Swedish renderings that the last syllable, in this

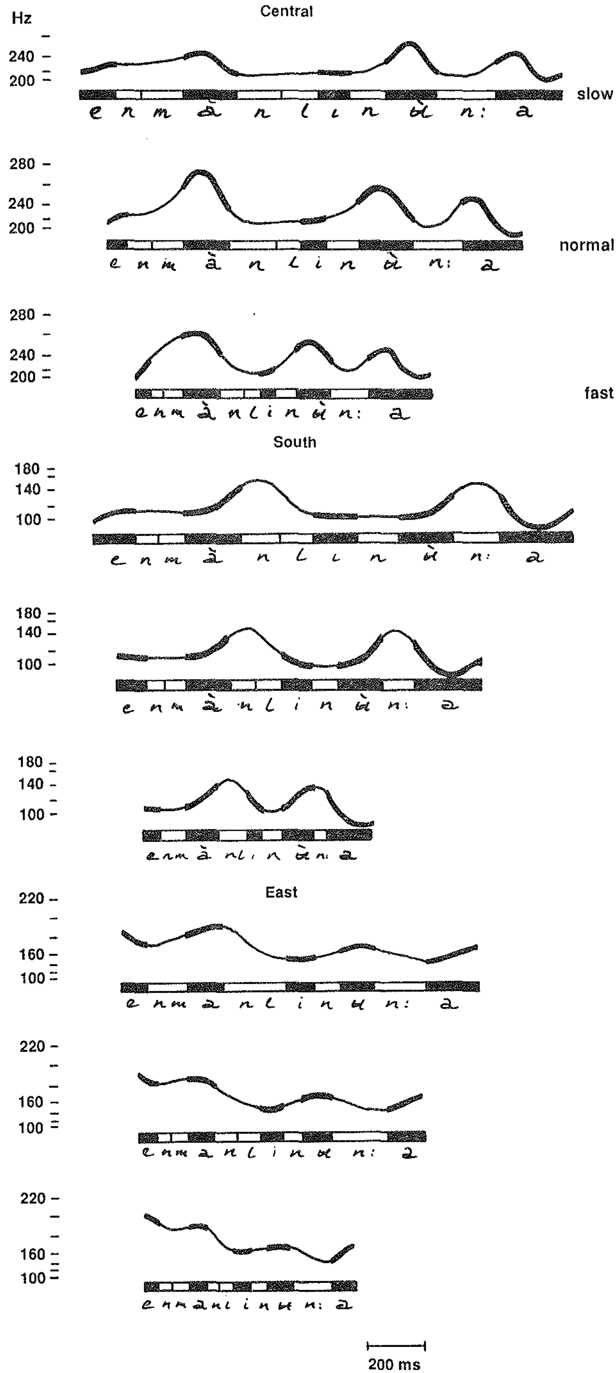


Fig.8. Average durations and F0 curves of 3 repetitions of the sentence <u>en mànli nùnna</u> spoken at three rates by representatives of three Swedish dialects, Central, South, East.

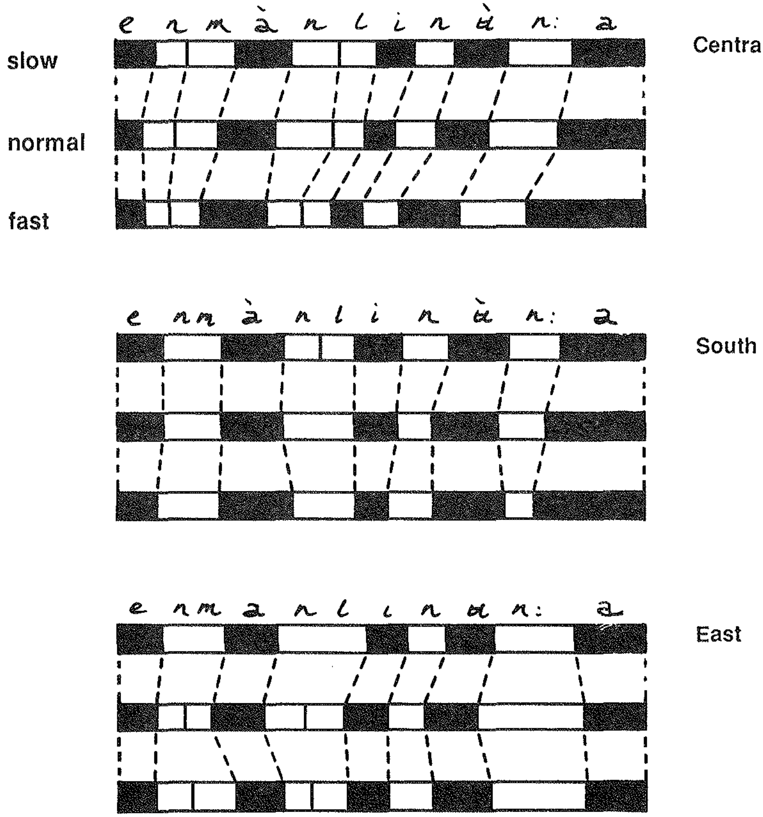


Fig.9. Relative durations of segments in three tempi. Test sentence en mànli nùnna in three Swedish dialects.

dialect the carrier of a sentence accent peak, occupies a much larger part of the fast utterance than it does for the southern speaker who has a very small relative lengthening of the final syllable. The Skåne rhythm (by outsiders judged to sound plodding) retains its characteristics all through the three tempi.

The temporal pattern of the East Swedish speaker is somewhat irregular. However, the first two syllables always occupy more room in the fast utterance than in the slow one, perhaps an indication of an early placement of sentence accent in this dialect. It is a moot question whether this is a trace of an old stage in Swedish prosody or interference from Finnish, which like all Finno-Ugric languages is characterized by an early sentence accent.

Tonal pattern

Figure 10 shows the different-rate tonal patterns imposed on a common temporal base. The already mentioned interdialectal differences are conspicuous. Central Swedish has three strong accent peaks. Southern Swedish has two strong peaks and Eastern Swedish has two weak ones. The first two peaks are connected with the main accents and the third peak of Central Swedish is due to the sentence accent which comes later than the main accent in this dialect (Bruce 1977). Changes with fast tempo can be described as follows (Fig. 11). The up- and down-going movements connected with the accents, the manifestations of the laryngeal gestures, look like triangles which retain their shapes regardless of tempo. When the segments are shortened, some turning points, which will be regarded as the first targets, are very fixed relative to the segments, some move into the neighbouring segments and syllables, while others disappear.

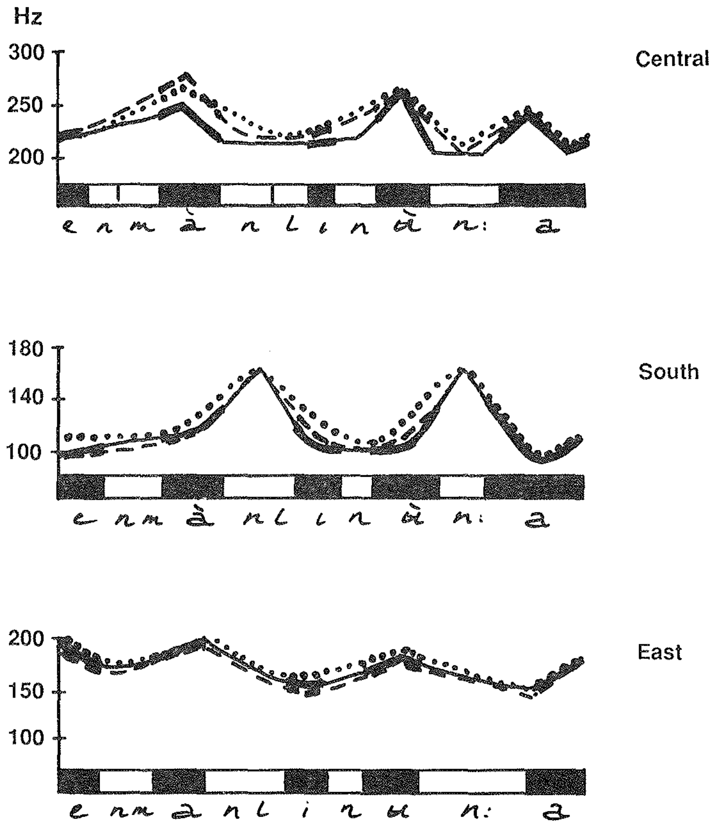


Fig.10. Different-rate tonal patterns imposed on a common temporal base. Test sentence <u>en manli(g) nunna</u> in three Swedish dialects.

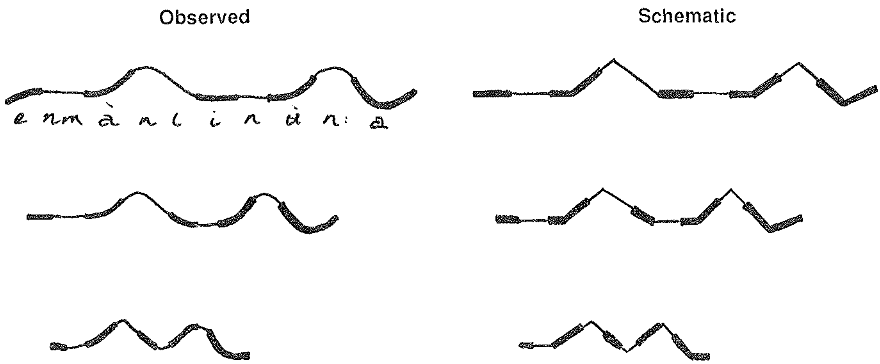


Fig.11. Southern Swedish observed and schematic F0 curves related to segments in three rates of speech. Test sentence en manli(g) nunna. Compare Fig.8.

The figure shows how the flat interval over the consonants and the unaccented syllables of slow speech disappears and is merged in the falling-rising movement of the accents in the fast renderings. It seems, then, that when time is short, it is enough to signal the general direction of the tonal movement over the accented syllable.

For Finland Swedish (see lower part of fig.10), the accent peaks are very weak and the change of the tonal pattern is uniform. The peaks are located at the end of the vocalic segment and the rise starts at the beginning of the preceding consonants regardless of tempo. The remarkable uniformity of East Swedish under tempo changes contrasts strongly with the non-uniformity of Central and South Swedish. The difference may be due to the absence of distinctive accent patterns and is perhaps also due to the influence of Finnish with its syllable-timed prosody.

In the comparison with Chinese, we will concentrate on

Central Swedish as being most similar to Chinese in the behavior of he durational patterns under changes of tempo, a similarity which is probably due to the fact that both these dialects have final sentence prominence.

Laryngeal pattern

In an EMG investigation of speech material focussing on Swedish accents (Gårding et al. 1975) it was observed that at fast rate all the EMG peaks of the vocalis and cricothyroid muscles were slightly higher corresponding to the higher peak values of the pitch curves. Also, the EMG peaks came closer together in fast speech and the cricothyroid peak which at normal rate fell to the baseline of activity turned into a rise well over the baseline. This was correlated to sharp changes from fall to rise in the fast-speech pitch curves as compared to the smooth troughs found in the normal-rate patterns. We notice a similar tendency in the Chinese pitch curves (see e.g. Figures 6 and 7). For Chinese there is also the necessity of adjusting the whole tonal movement to segments in a more precise way than in Swedish which may result in increased steepness and sharper dips in the fast-speech curves.

4. Comparison and conclusions

It follows from the preceding section on Swedish that statements 1-3 about Chinese intonation and the qualifying remarks remain valid for Swedish as well if the word 'tone' is replaced by 'accent'. The reason for the similarity is of course that with increased rate of speech the prosodic commands come closer together with overlapping and weakening of the corresponding gestures as a result. Some examples of this hypothetical process are given in Figure 12.

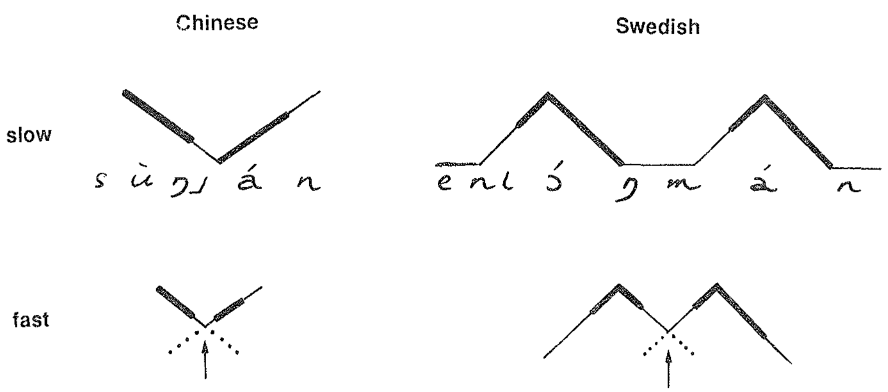


Fig.12. Encounter of laryngeal gestures. Chinese is exemplified by the encounter of a falling and a rising tone, Swedish by two acute (in this dialect falling) accents. Dotted line indicates covert parts of the gestures. Arrows point to the splices. The drawing is consistent with the observed raised F0 values of the low turning points and the narrower grids of fast speech.

What about the observed differences? In our analysis we found that the number of turning points are more numerous and their locations more stable in Chinese than in Swedish. This obviously has to do with the phonological system which results in a larger number of tones than accents in a comparable sentence. In Chinese most of the syllables are tonic and each tone has target turning points tied to the vocalic segment according to certain rules. In spite of sandhi rules and with some allowance for coarticulatory effects the communicative value of a tone will prevent too much overlap. In Swedish an accent is most often tied to a sequence of syllables (the stem) and the number of syllables for which pitch is derived by interpolation is greater than in Chinese. Although the first target point is the most fixed one in both languages there is more room in Swedish for the laryngeal gestures to extend and encompass also neighbouring syllables. As a result, the second target point has more leeway in Swedish than in Chinese.

A phonetic difference which may be worth mentioning and important for the distincion between tone and accent is that the first target point of a Swedish falling accent tends to be located near the middle of the vocalic segment whereas in Chinese a falling tone is more closely connected with the C/V boundary. Another observation is that with increased tempo a Chinese prominent falling tone becomes more similar to its Swedish counterpart in that the peak is shifted towards the middle of the vocalic segment. The larger reduction of less prominent tones in fast speech makes the Chinese fast-speech curves more like those of an accent language.

Finally, let us return to the statement of Kozsenikov and Chistovich that the relative durations of words and syllables are independent of speech rate. In the majority of our dialects, the part under sentence prominence has a longer relative duration in fast speech. Hence their durational patterns are not invariant.

It is also clear from our analysis that, in languages with distinctive tones or accents, the coordination between temporal and tonal patterns has a certain dependence on tempo.

Acknowledgement

We wish to thank the Swedish Academy of Engineering Science for continued support of our collaboration.

We are also indebted to Lars Gårding for the choice of parameters of the grid construction in the appendix.

Appendix

In this appendix we shall outline a way of generating schematic versions of our observed curves using a small number of parameters. The scheme follows in principle Gårding et al. 1983. We assume that the durations of the segments, i.e. the temporal pattern is given according to Figures 6 and 7. The construction does not take microprosody into account.

The outline has three parts, phonological rules, i.e. context-dependent changes of symbols, the tonal grid and the insertion of turning points followed by interpolation.

1. Phonological rules.

Our two sentences have the following sequences of tones in a syntactic structure indicated by parentheses (0 means atonic),

(1) (HL+LH)+((HL)+(LH+HL))#

(2) (HL+HL)+((HL)+(HL+0))#

The sign # marks the end of the sentence. An assimilation rule deletes the second element of every tone when followed by a target of the same kind. We make an exception for the sequence L+L after mai where assimilation does not seem to occur. A sandhi rule weakens a HL between two HL's. The result of these rules is the two sequences where the syntactic structure is left out in the notation

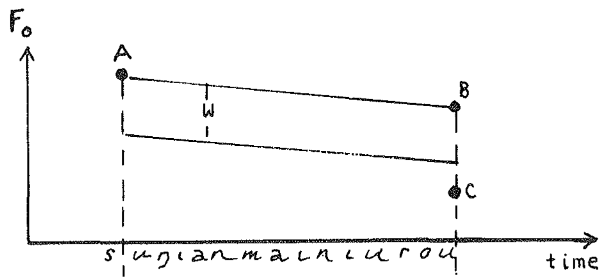
(1) H(L)+L(H)+HL+L(H)+HL#

(2) HL+HL+HL+HL+0#

2. The tonal grid

A falling grid for declarative sentence intonation is constructed using four parameters A,B,C,W. We first describe it as a geometrical figure which shows the role of the parameters. A and

B are the endpoints of a topline stretching over the whole sentence, A being higher than B. The baseline is parallel to the topline at the distance W below it. The point C marks the end of the sentence (#) and the end of the tonal curve. See the figure below.



Average values in Hz of the parameters for three tempi and the two male speakers may be given as follows

A: 195 ± 10, B: 165 ± 10, C: 90 ± 10, W: 45 ± 5.

The first figure of each pair refers to normal, + to fast and - to slow tempo. The actual values may differ by 10 or 15 Hz, but for all speakers increased tempo is accompanied by a higher grid and a smaller width. In an extended context including other tonal patterns, it is possible that relative values can be given to the parameters A,B,C and W. To account for different voice registers, the scale should be logarithmic.

3. Construction of the tonal curve

When the grid is drawn for any tempo, the highs and lows which remain after the application of the phonological rules are inserted on the tonal grid above their respective time positions. The weakened highs are inserted halfway between the grid lines. The final step is a linear interpolation between the generated points.

References

Browman C.P. & L. Goldstein. 1986. Towards an articulatory phonology. Phonology Yearbook 3:

Bruce G. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Gleerup, Lund

Cheng C.C. 1973. A synchronic phonology of Mandarin Chinese. Mouton, The Hague

Gårding E. 1967. Internal Juncture in Swedish. Travaux de l'Institut de Phonétique de Lund VI. Gleerup, Lund

Gårding E. 1975. The influence of tempo on rhythmic and tonal patterns in three Swedish dialects. Working Papers 12, Department of Linguistics, Lund University: 71-84

Gårding E. 1981. Contrastive Prosody: A model and its application. Studia Linguistica 35: 146-166

Gårding E. 1984. Comparing intonation. Working Papers 27. Department of Linguistics, Lund University: 75-100

Gårding E. 1987. Speech act and tonal pattern in Standard Chinese: Constancy and variation. Phonetica 44: 13-29

Gårding E., O. Fujimura, H. Hirose & Z. Simada. 1974. Laryngeal control of Swedish word accents. Working Papers 10. Department of Linguistics: 53-82

Gårding E., J. Zhang & J.O. Svantesson. 1983. A general model for tone and intonation in Standard Chinese. Working Papers 25. Department of Linguistics, Lund University: 53-65

Gårding E., P. Kratochvil, J.O. Svantesson & J. Zhang. 1986. Tone 4 and Tone 3 discrimination in modern standard Chinese. Language and Speech, Vol.29, Part 3: 281-293

Howie J.M. 1974. On the domain of tone in Mandarin. Phonetica 30, 129-148

Kozsevnikov V.A. & Chistovich L.A. 1965. Rech, artikulyatsiya i vospriyatiye. English translation: Speech: Articulation and perception. Published by U.S. Department of commerce, Washington D.C. 1966

Kratochvil P. 1968. The Chinese Language today. Hutchinson. London

Wu J.-J. 1982. Rules of intonation in Standard Chinese. Preprints from the 13th International Congress of Linguistics. Eds. H. Fujisaki & E. Gårding

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

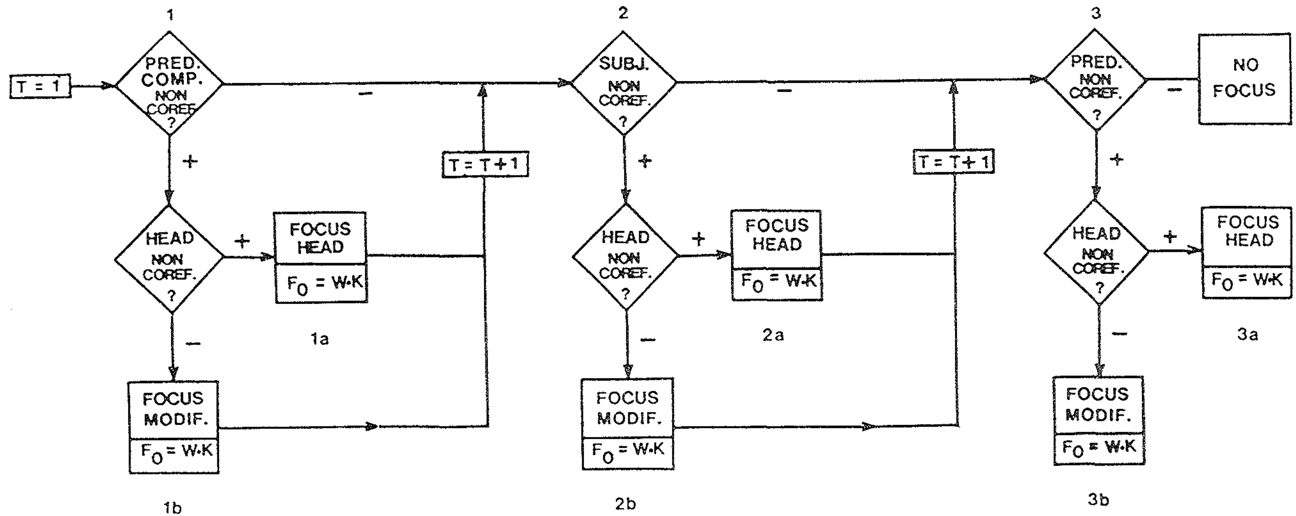


FIGURE 1. MODEL (FLOWCHART) FOR ASSIGNING INFORMATION FOCUS TO CONSTITUENTS ON THE BASIS OF GRAMMATICAL FUNCTIONS AND THE COREFERENTIAL STATUS OF THE LEXICAL MATERIAL REALIZING A PARTICULAR FUNCTION. THE INPUT TO THE MODEL IS A GIVEN CLAUSE (S). FOCUS IS REALIZED AS PITCH (F_0) ACCORDING TO THE EQUATION $F_0 = W \cdot K$ 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,

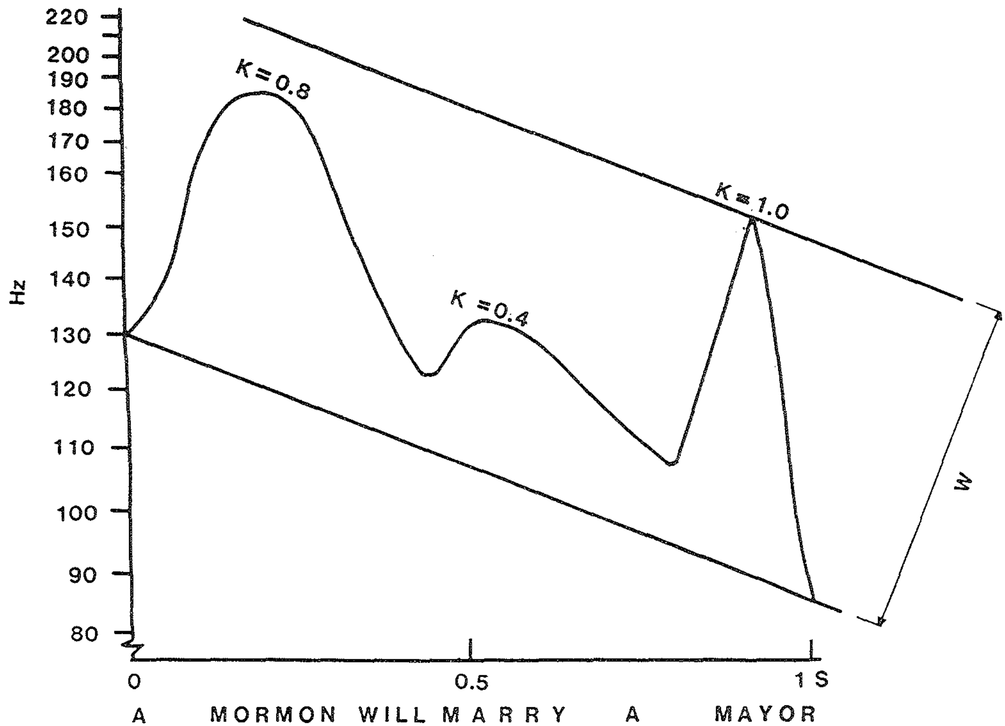


FIGURE 2a. ACTUALLY OCCURRING F_0 CURVE OBTAINED FOR A READING OF THE LAST SENTENCE IN (1) WHERE THE SUBJECT, PREDICATE AND PREDICATE COMPLEMENT ARE FOCUSSED ACCORDING TO THE MODEL IN FIGURE 1.

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

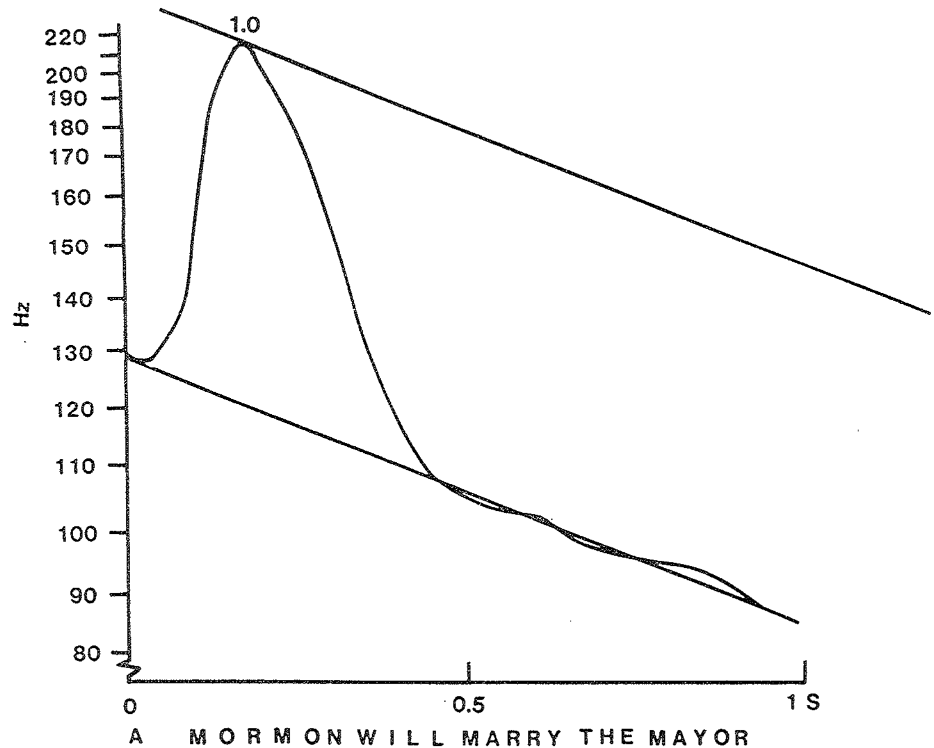
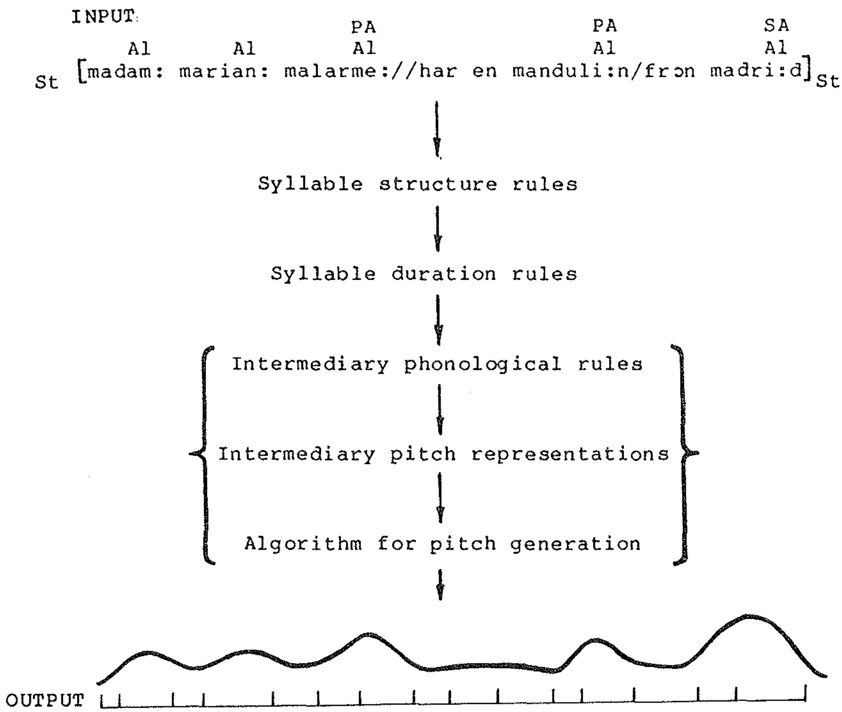


FIGURE 2b. ACTUALLY OCCURRING F_0 CURVE OBTAINED FOR A READING OF THE LAST SENTENCE IN (3) WHERE ONLY THE SUBJECT IS FOCUSSED ACCORDING TO THE MODEL IN FIGURE 1.

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 A1 = 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).

<u>Baseline vs. topline</u>. 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.

<u>Analysis by synthesis</u>. 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

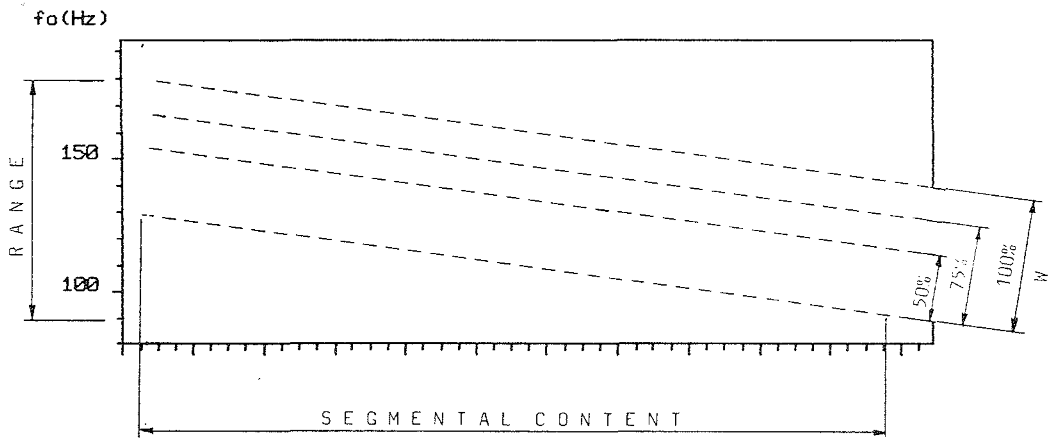


FIGURE 4. PHONOLOGICAL GRID USED FOR SYNTHESIZING F_0. THE F_0 RANGE EXTENDED BETWEEN 90 AND 180 HZ. THE BEGINNING AND END POINTS FOR A GIVEN SENTENCE WERE SET AT 130 HZ AND 90 HZ, RESPECTIVELY. ACCORDING TO FIGURE 1, THE FIRST FOCUSSED CONSTITUENT RECEIVES A LEVEL OF PROMINENCE EQUAL TO 'W', THE SECOND,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

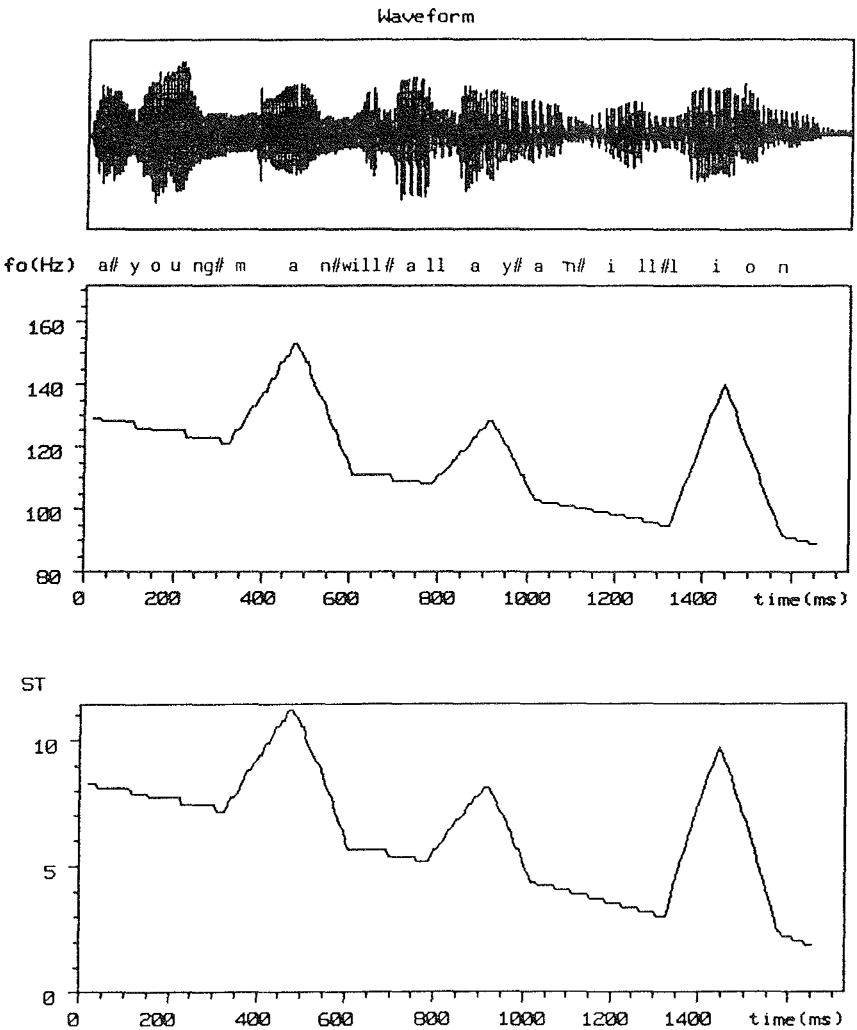


FIGURE 5. SYNTHESIZED F_0 CURVE OF SENTENCE 4 WITH FOCUS ON SUBJECT, PREDICATE, AND PREDICATE COMPLEMENT ACCORDING TO FIGURE 1. FOR SAKE OF COMPARISON, THE SYNTHESIS IS REPRESENTED USING BOTH A LINEAR SCALE (UPPER CURVE) AND A SEMITONE SCALE (LOWER CURVE). NOTE THAT THE RELATIVE PITCH LEVELS ARE ALMOST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

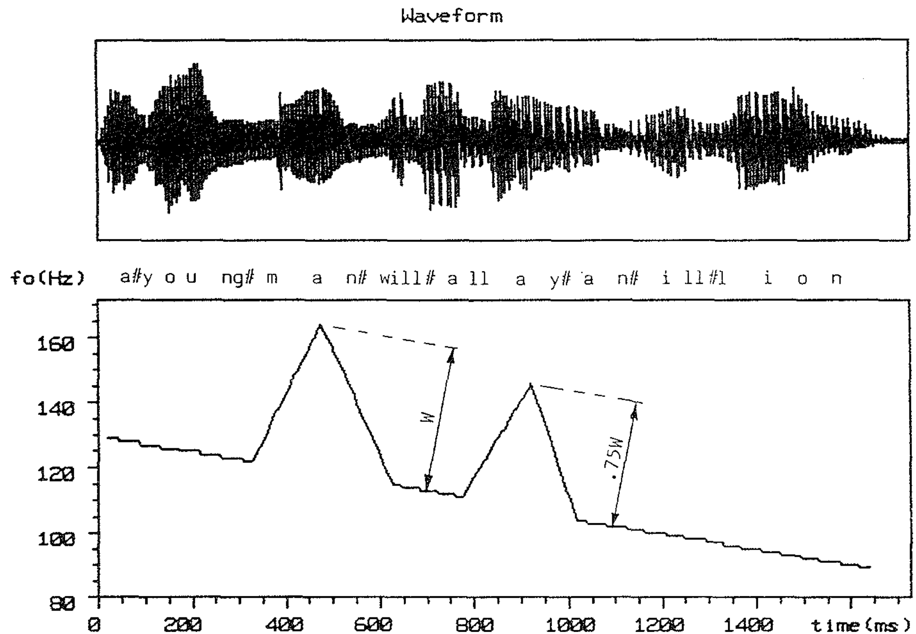


FIGURE 6. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE SUBJECT AND PREDICATE

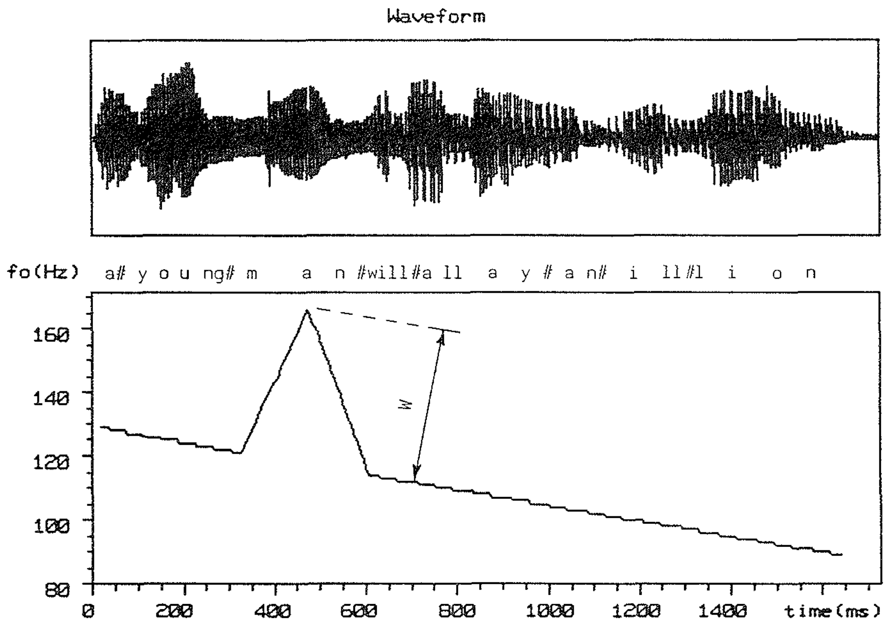


FIGURE 7. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE SUBJECT

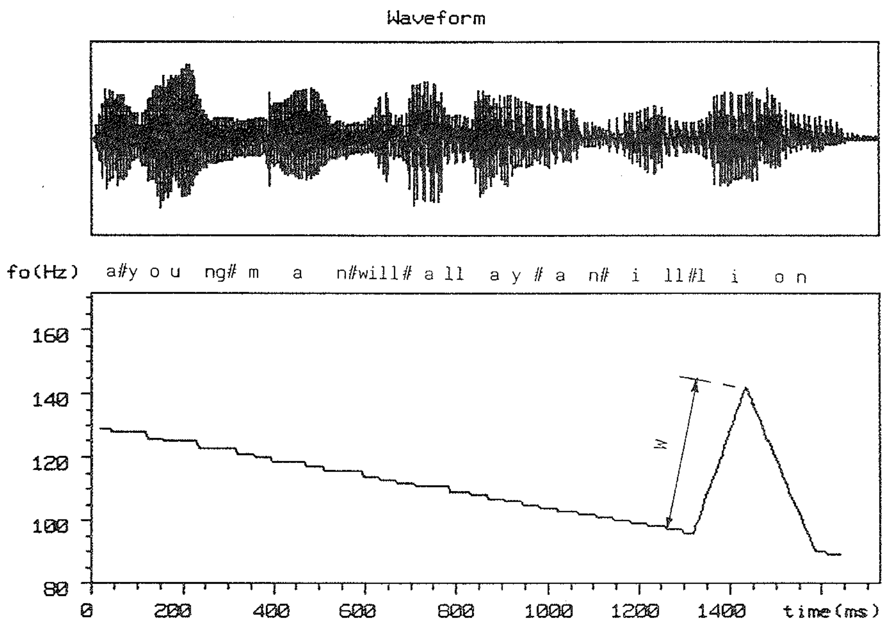


FIGURE 8. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE OBJECT

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).

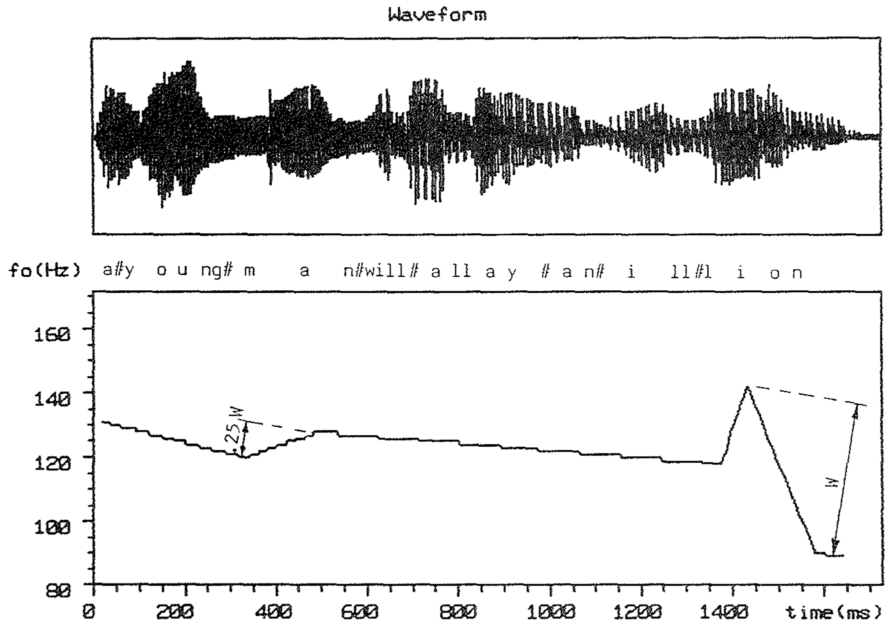


FIGURE 9. F_0 CURVE SYNTHESIZED ACCORDING TO LADD'S (1986) "OVERALL CONTOUR SHAPE APPROACH" WITH FOCUS ONLY ON THE OBJECT OF SENTENCE (4)

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)).

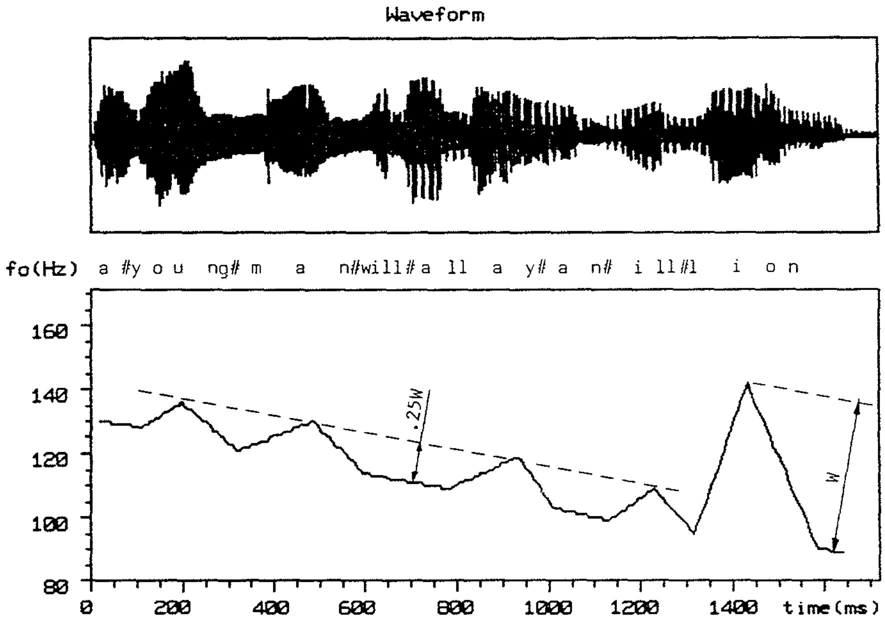


FIGURE 10. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE OBJECT AND PHRASE ACCENTS ON ALL PREFOCAL LEXICAL WORDS

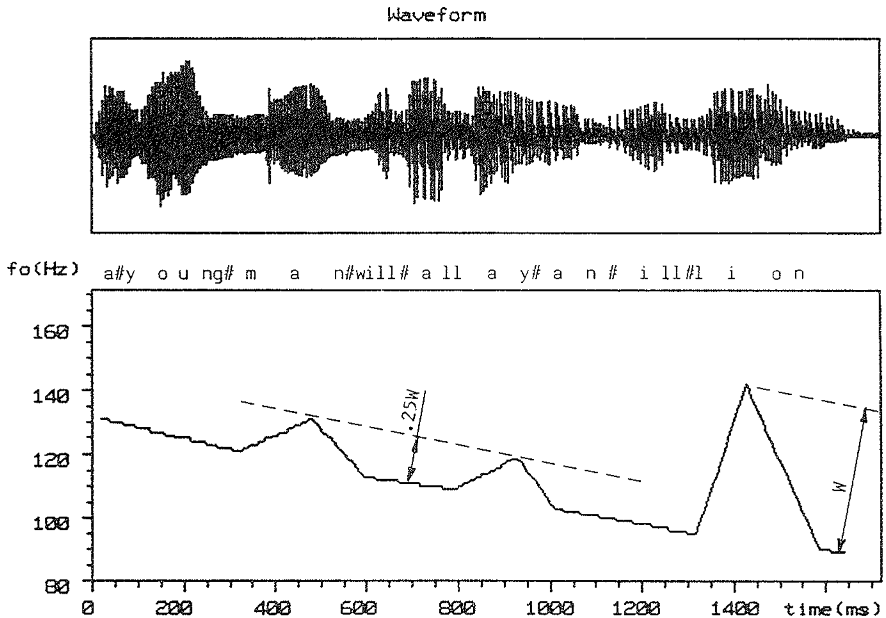


FIGURE 11. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE OBJECT AND PHRASE ACCENTS ON ALL PREFOCAL LEXICAL HEADS

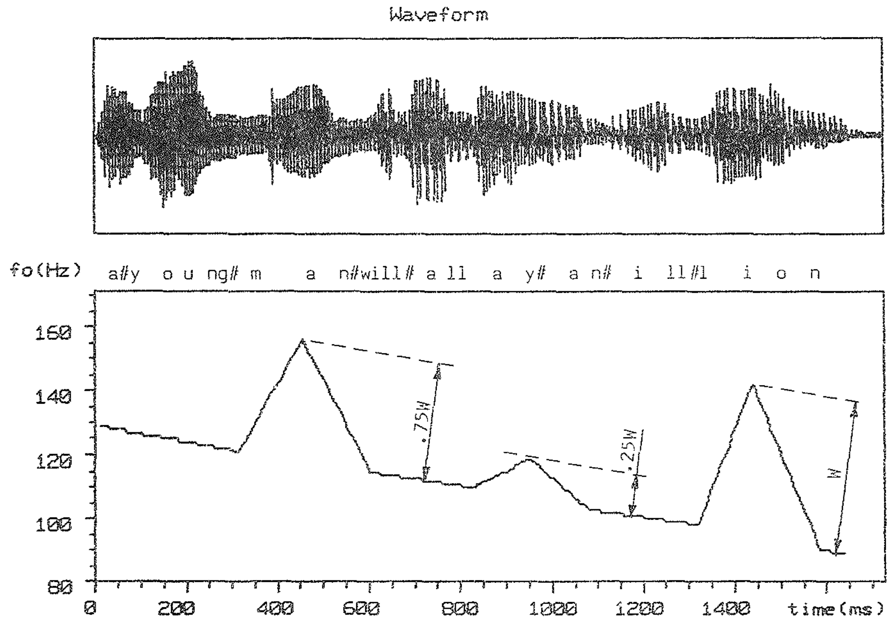


FIGURE 12. SYNTHESIZED F_0 CURVE OF SENTENCE (4) WITH FOCUS ON THE SUBJECT AND OBJECT AND PHRASE ACCENT ON PREFOCAL VERB

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, <u>the cash</u>$_1$, <u>it</u>$_1$, and <u>my money</u>$_1$ are assumed to refer to the same referent, introduced by <u>alimony</u>$_1$. <u>Cash</u> and <u>money</u> 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 5

e.g. Delete all post focal phrase accents in a given (component) sentence

↓

Intermediary Pitch Representations 6
(Not relevant in the present work)

↓

Algorithm for pitch generation:

Define Grid: Give start and end points for a 7
declarative sentence (= baseline)

↓

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 11
are 2 values for any given syllable (for ex., Focus and Phrase prominence), take the highest value

↓

Determine where the F_0 peak is to be defined (a point 12
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)

↓

Define points 20 'frames' (=20X6.4ms) on either side 13
of the F_0 peak (scope of F_0 obtrusion)

↓

Follow the baseline, joining all defined points, 14
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).

Figure 13

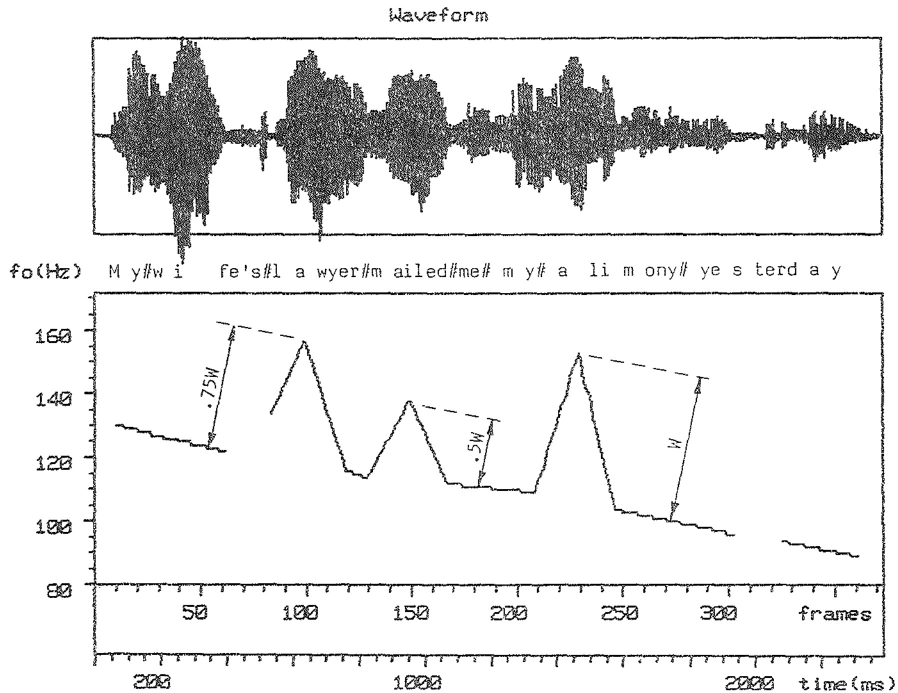


FIGURE 14. SYNTHESIZED F_0 CURVE OF SENTENCE (5a) WITH FOCUS ON SUBJECT, PREDICATE AND OBJECT

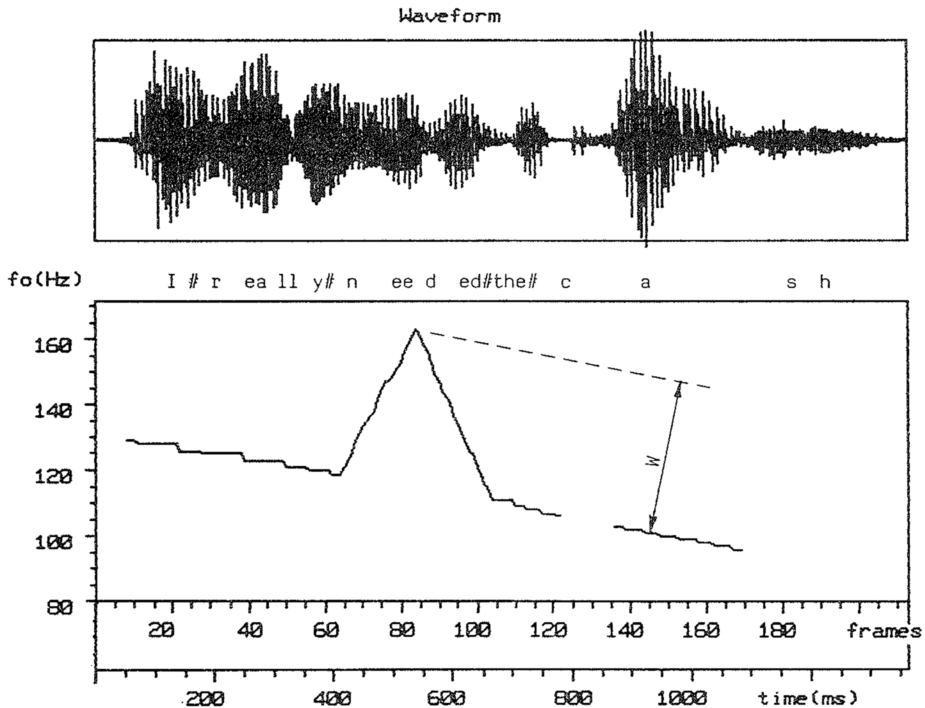


FIGURE 15. SYNTHESIZED F_0 CURVE OF SENTENCE (5b) WITH FOCUS ON PREDICATE

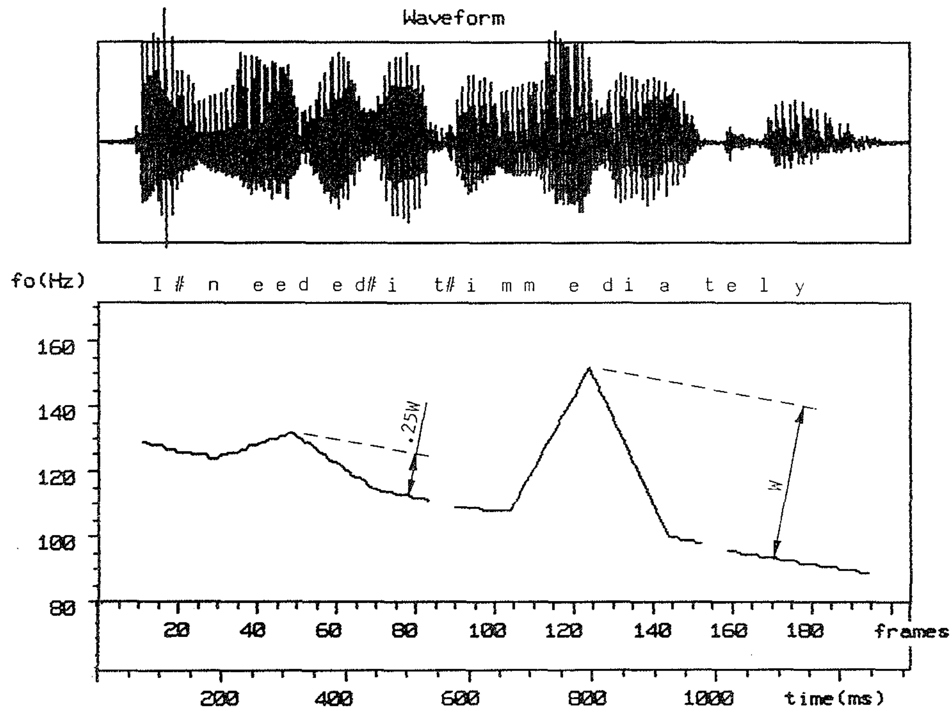


FIGURE 16. SYNTHESIZED F_0 CURVE OF SENTENCE (5c) WITH FOCUS ON PREDICATE COMPLEMENT AND PHRASE ACCENT ON PREFOCAL PREDICATE (HEAD)

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)

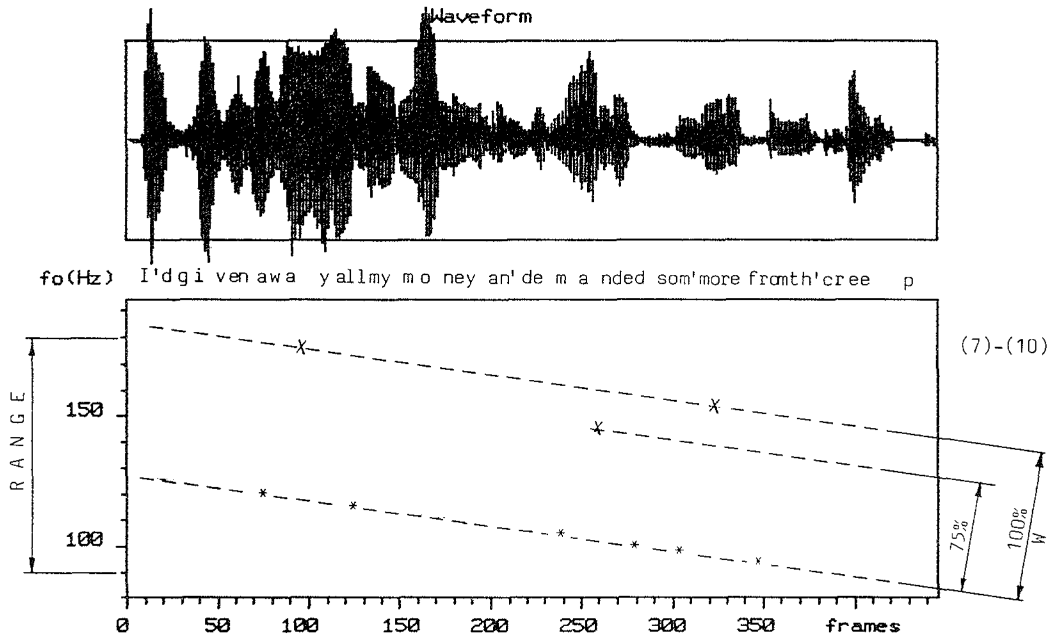


FIGURE 17a. PARTIAL DERIVATION OF F_0 CURVE OF SENTENCE (5d) (AFTER POINT 13 IN FLOW-DIAGRAM). SEE ABOVE FOR A CLARIFICATION OF THE FIGURE.

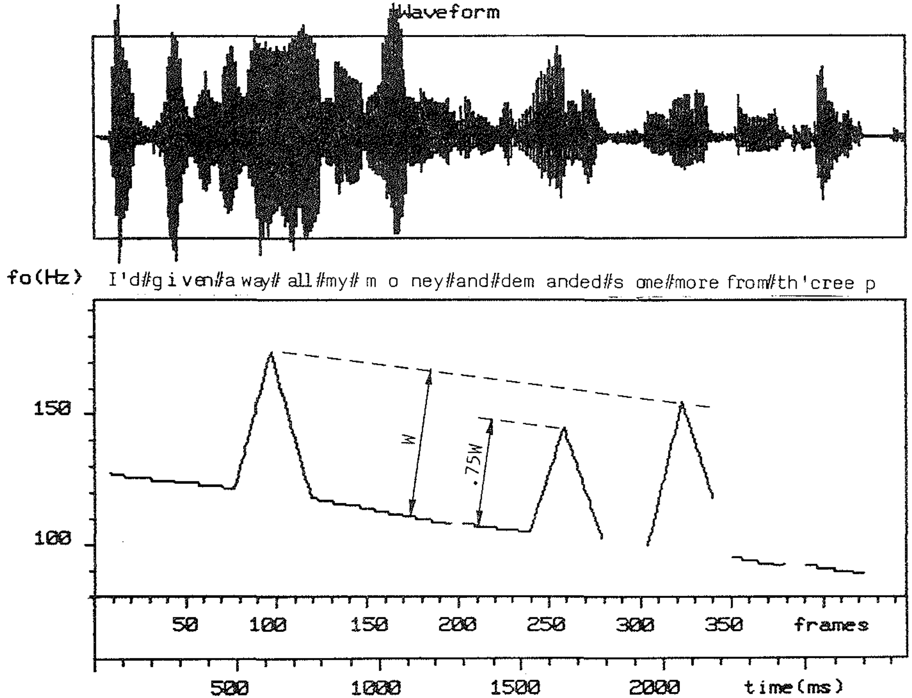


FIGURE 17b. OUTPUT OF THE PITCH GENERATING COMPONENT WITH FOCUS ON PREDICATE IN FIRST COMPONENT SENTENCE AND FOCUS ON PREDICATE AND PREDICATE COMPLEMENT IN SECOND COMPONENT SENTENCE. NOTE THAT DECLINATION IS NOT 'RESET' AT BEGINNING OF SECOND CLAUSE.

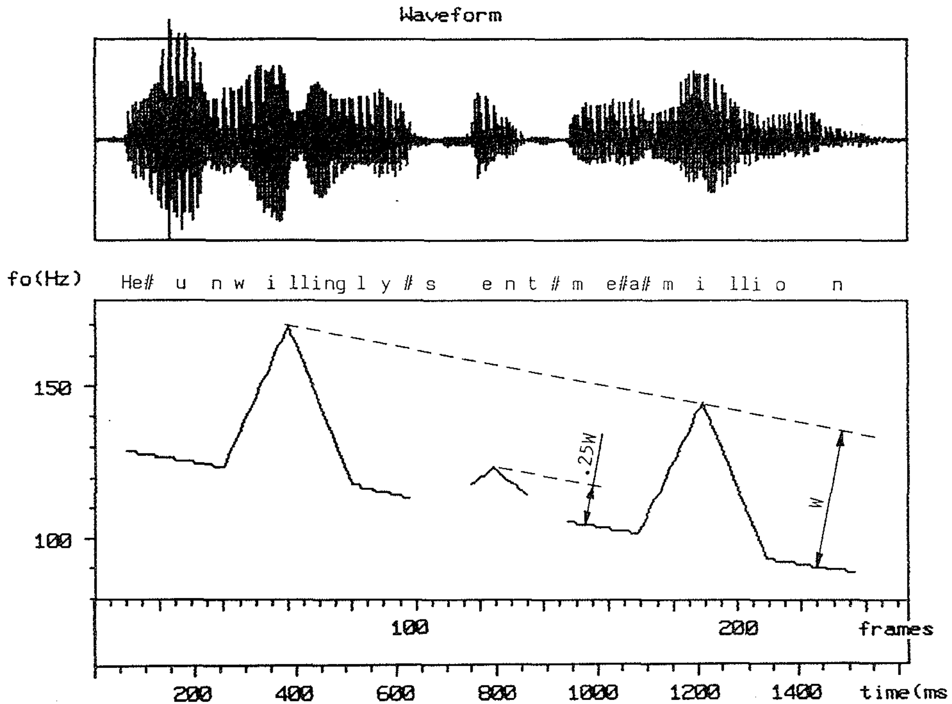


FIGURE 18. SYNTHESIZED F_0 CURVE OF SENTENCE (5e) WITH FOCUS ON PREDICATE COMPLEMENTS AND PHRASE ACCENT ON PREDICATE (HEAD)

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)

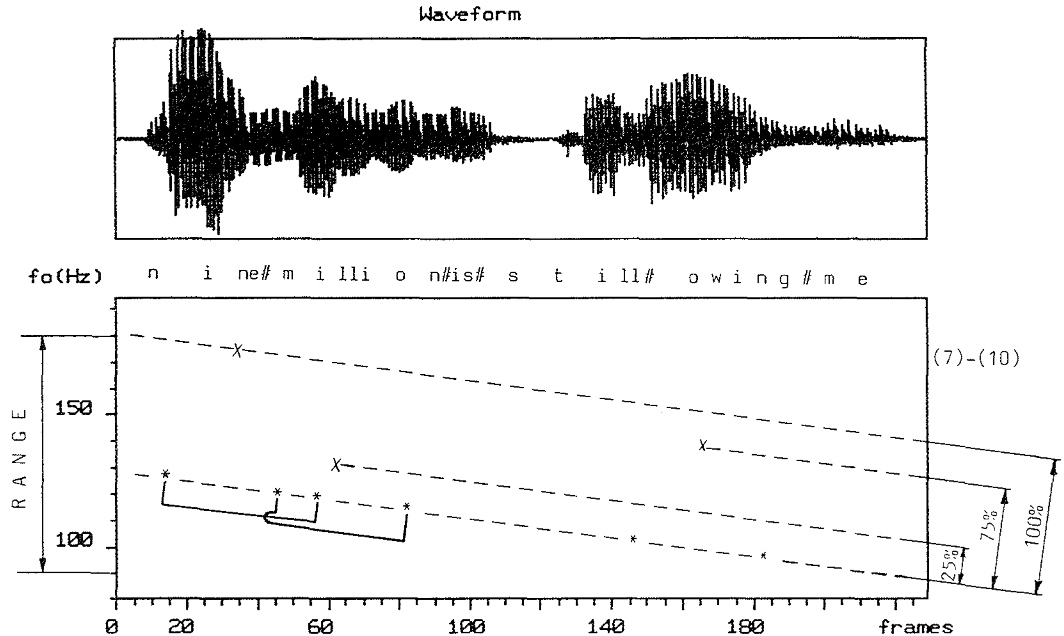


FIGURE 19a. PARTIAL DERIVATION OF F_0 CURVE OF SENTENCE (5f) (AFTER POINT 13 IN FLOW DIAGRAM IN FIGURE 13). SEE ABOVE FOR A CLARIFICATION OF THE FIGURE. DERIVATION CONTINUED IN FIGURE 19b.

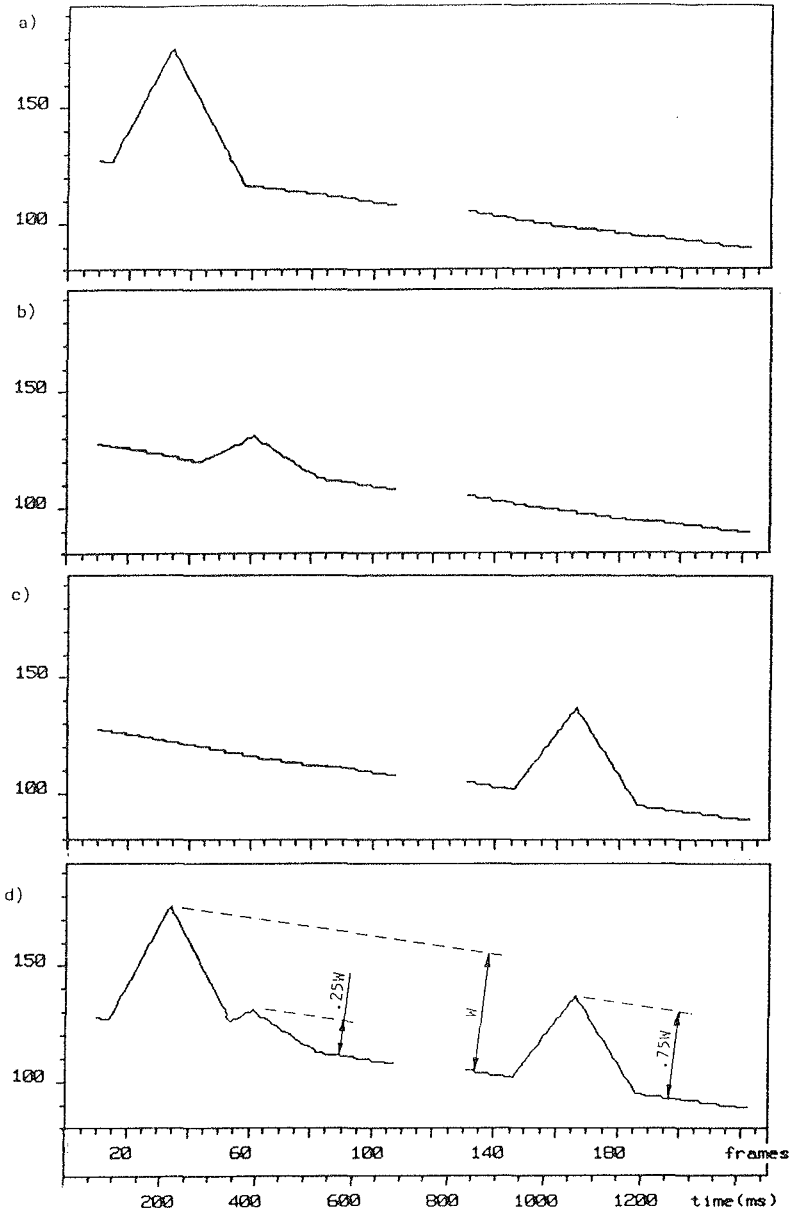


FIGURE 19b. POTENTIAL STAGES IN THE SYNTHESIS OF THE F_0 CURVE WHERE THE FIRST TWO PITCH OBTRUSIONS OVERLAP. THE FINAL OUTPUT IN (d) IS OBTAINED BY CONNECTING THE HIGHEST POINTS IN THE INTERMEDIARY CURVES (a-c).

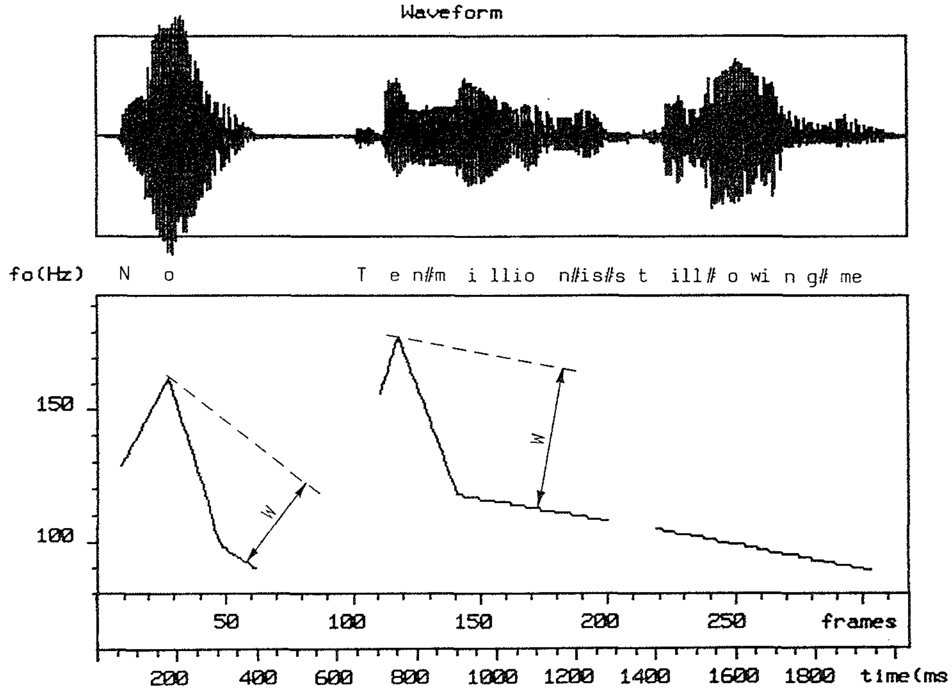


FIGURE 20. SYNTHESIZED F_0 CURVE FOR SENTENCE (5g) WITH FOCUS ON SUBJECT MODIFIER 'TEN'. PROMINENCE ON 'NO' NOT ACCOUNTED FOR BY THE MODEL PRESENTED HERE BUT RATHER ASSUMED TO BE ASSIGNED BY OTHER RULES.

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.

AUTOMATIC PROSODIC ANALYSIS FOR SWEDISH SPEECH RECOGNITION[1]

David House*, Gösta Bruce*, Francisco Lacerda+ and Björn Lindblom+

ABSTRACT

A mingogram reading experiment was carried out in which an expert in Swedish prosody was presented with computer simulated mingograms of unknown Swedish sentences and asked to identify the following categories: stressed and unstressed syllables, grave and acute word accents, focal accent, and terminal juncture. Out of a total of 178 occurrences of the different categories, 151 were correctly identified (85%). The categories were identified by using the Fo contour, energy envelope and a duplex oscillogram. On the basis of this experiment, a set of preliminary, hierarchically ordered automatic analysis rules have been formulated using Fo movement patterns synchronized with energy envelope peaks to define the prosodic categories. These rules have been tested by using two non-expert mingogram readers and are being implemented on an automatic prosodic analysis system.

INTRODUCTION AND BACKGROUND

Prosodic information contained in the speech signal can be used by a speech recognition system to limit lexical access and provide information concerning phrase boundaries and syntactic structure. In Swedish the prosodic categories of stress, word accent, focal accent, and initial and terminal juncture are ready candidates for automatic recognition rule formulation.

[1]A shorter version of this paper was presented at the European Conference on Speech Technology, Edinburgh, 2-4 September, 1987

*Dept. of Linguistics and Phonetics, Lund University, S-223 62 Lund, Sweden.
+Institute of Linguistics, Stockholm University, S-106 91 Stockholm, Sweden.

This paper represents a status report from an ongoing joint research project shared by the Phonetics Departments at the Universities of Lund and Stockholm. The project, "Prosodic Parsing for Swedish Speech Recognition", is sponsored by the National Swedish Board for Technical Development and is part of the National Swedish Speech Recognition Effort in Speech Technology.

Research in prosodic recognition is a natural continuation of a long tradition of research in prosody in Lund and Stockholm. In Lund, this research has been aimed at investigating prosodic phenomena such as rhythm, accentuation and intonation and their relationship both to linguistic structure (phonology, morphology, syntax, semantics, pragmatics and text linguistics) and to the physical realization in the form of tonal and temporal patterns (Bruce, 1977; Bruce and Gårding, 1978; Gårding and Bruce, 1981; Gårding, 1982). Prosody research in Stockholm has primarily dealt with the temporal organization of speech (Lindblom, et al., 1981; Lyberg, 1981) and the relationships between prosody, grammar and speech perception (Svensson, 1974).

Prosodic recognition can constitute one component of a larger, phonetically structured recognition system. The intention is to build on the existing knowledge of Swedish prosody thereby facilitating and enhancing Swedish Speech Recognition. This combination of prosody and recognition is a relatively new area of research. See Lea (1980) and Vaissière (1983).

The primary goal of the project is to develop a method for extracting relevant prosodic information from a speech signal. Some issues relating to this goal are 1) What criteria can we use to recognize prosodic categories, 2) What kind of acoustic invariance relates to prosodic categories, and 3) What degree of success can we achieve in recognizing prosodic categories. Futhermore, by using a recognition approach to prosody, we hope to reach a better understanding of the mechanisms involved in human perception of prosody.

Among the many functions of prosody in spoken language communication, there are two that can be considered fundamental: the weighting function and the grouping function (Bruce, 1985). Speakers use stress and accentuation, for example, to give weight to syllables, words and phrases (related to semantics). Speakers use juncture to signal phrase boundaries and coherence signals within a phrase (related to syntax). The prosodic categories used in the project are STRESS (stressed and unstressed syllables), WORD ACCENTS (acute and grave), FOCUS (focal and non-focal accents), and JUNCTURE (connective and boundary signals for phrases).

In Swedish, the basic dichotomy between stressed and unstressed syllables exists as in English. This division provides the basic rhythmical structure of spoken Swedish, but gives no information about word boundaries or the number of words in an utterance.

In both Swedish and Norwegian, the primary stressed syllable is characterized by having one of two tonal accents: Acute (Accent I) or Grave (Accent II). Identification of ACUTE accent can restrict and thereby facilitate the lexical search. Identification of GRAVE accent provides us with morphological information which can also facilitate lexical access. For example we know that the syllable following the stressed syllable of a grave accent word belongs to the same word.

The identification of focal accents is important for a recognition system since a focal accent tells us that a word is emphasized and bears important information. Focal accents have a predictive value for both semantics and syntax.

The correct identification of juncture will assist a recognition system in isolating phrases. These phrases, or information chunks, are somewhat related to syntax. An interesting and challenging aspect of juncture is that the Fo representation of initial juncture bears a strong resemblance to that of the acute focal word accent while the representation of final juncture resembles that of a grave word accent. Moreover, juncture representations can interact and interfere with other prosodic categories.

MINGOGRAM READING (EXPERT READER)

Promising results from previous mingogram reading experiments (Welin and Lindblom, 1980) led us to use this method as a basis for choosing acoustic criteria for prosody recognition and as a background for recognition rule formulation. By interviewing an expert reader, we should be able to isolate the most salient cues for use by the recognizer.

Ten test sentences, unknown to the reader, were carefully designed so that both the placement of the prosodic categories and the syntax of the sentences were varied. Each sentence contained 10 to 15 syllables with 2 to 5 stressed syllables in each sentence. (See Appendix 1 for a list of the test sentences.) The speech material was recorded in a Stockholm dialect of Swedish in an anechoic chamber, digitized and stored in disk files (one per sentence). The speech signal was sampled at 20kHz (16-bit/sample). Fundamental frequency contours were obtained by digital processing of the speech signal using a classical instantaneous Fo measuring technique for the Fo extraction. The Fo signal was smoothed and plotted in synchrony with the input speech signal and an intensity curve. The plots were stored, with low sampling frequency, in a 3-channel file for subsequent analysis by the automatic prosodic analysis system. Simulated mingograms of the speech signal displaying a duplex oscillogram, the Fo contour and a bandpass-filtered (1500-3500 Hz) intensity curve were presented to the expert reader (see Figure 1 for a sample mingogram). The reader was given the task of identifying the above mentioned prosodic categories on the basis of the mingogram registration.

Of a total of 178 occurrences of the different categories, 151 were correctly identified (85%). These results break down into categories as follows: GRAVE ACCENTS (both focal and non-focal) 13 of 13 (100%), GRAVE FOCAL ACCENTS 7 of 8 (88%), ACUTE ACCENTS (both focal and non-focal) 20 of 23 (87%), ACUTE FOCAL ACCENTS 12 of 13 (92%), ACUTE FOCAL FINAL ACCENTS 2 of 2 (100%), STRESSED SYLLABLES 37 of 37 (100%), UNSTRESSED SYLLABLES 60 of 82 (73%), and TERMINAL JUNCTURE 2 of 2 (100%). A confusion matrix display of these results is presented in Figure 2. Clearly there is considerable prosodic information available in the acoustic signal alone. It is even possible that these results could be improved by adjusting the reader's tendency to identify unstressed syllables as stressed, that is to say, by moving his stressed-unstressed boundary more toward the stressed syllables.

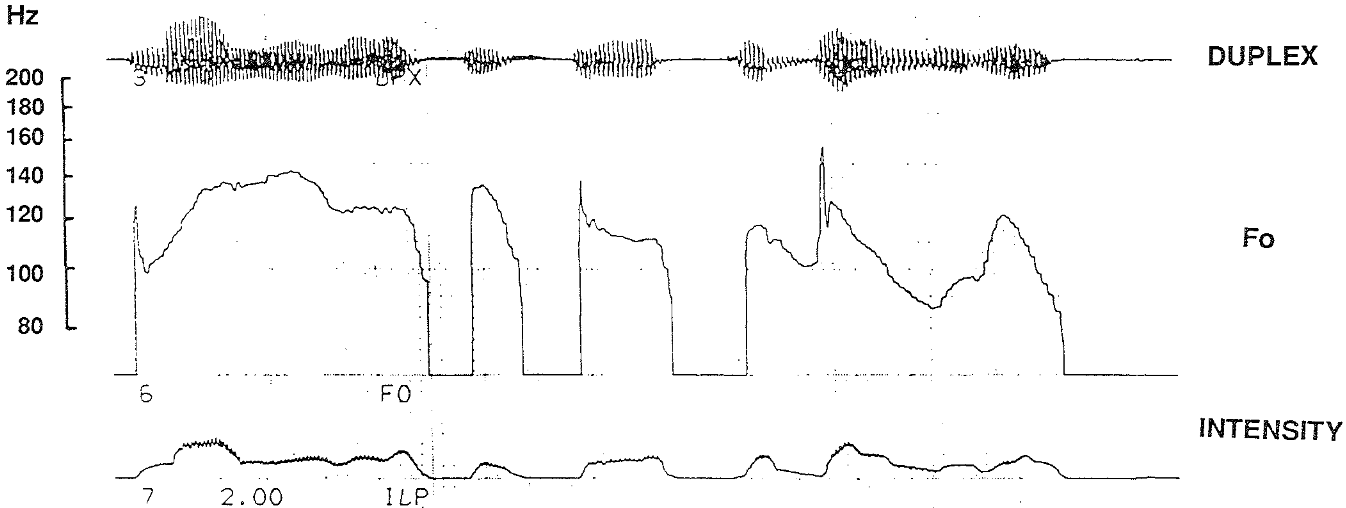


Figure 1. Sample mingogram used for the mingogram reading experiment.

Confusion matrix for expert reader: sentence set 1

IDENTIFIED AS (in percent):

CATEGORY	G	GF	A	AF	AFT	TJ	SS	US	Ø	TOTAL
G	100									13/13 100%
GF	12	88								7/8 88%
A	4		87				9			20/23 87%
AF				92			8			12/13 92%
AFT					100					2/2 100%
TJ										
SS							100			37/37 100%
US							18	73	8	60/82 73%
Ø								(2)		

All categories = 151/178
85%

G = Grave
GF = Grave-focal
A = Acute
AF = Acute-focal
AFT = Acute-focal-terminal
TJ = Terminal juncture
SS = Stressed syllable
US = Unstressed syllable
Ø = Not a syllable or missed syllable
TOTAL = Total correct identification of the category in number and percent

Figure 2. Results of the mingogram reading experiment: Expert reader.

RECOGNITION RULE FORMULATION

On the basis of the mingogram reading test results, descriptive rules were formulated and ordered so that the categories that were easiest to identify, i.e. those showing the greatest degree of signal invariance, should be identified first leaving the categories with more variable patterns to be identified last. Fo movement patterns were deemed to be the most salient information, and the rules, therefore, are based on these movements. However, a falling Fo contour, for example, can signal quite a number of prosodic categories. A vowel containing an Fo fall can be a stressed vowel with a grave accent, a grave focal accent, an acute focal final accent, or it can even be an unstressed vowel in final position. Invariance, therefore, lies not in the Fo movements alone, but in the synchronization of these movements with vowel onset along with the range and steepness of the Fo movement. A complete English version of the rules is presented in Appendix 2.

The first rule category is a falling Fo of a certain steepness and range where the beginning of the fall is synchronized with the vowel onset. This signals a stressed vowel with a grave accent. The reader first attempts to find all the grave accents in the sentences. Then the reader is instructed to look at the syllable following the identified grave accent. If there is a high or rising Fo in the following syllable, then the grave accent is also a focal accent. A rising Fo synchronized with a vowel onset signals an acute focal accent and a down-stepping of Fo from one vowel to the next signals an acute non-focal accent in the vowel receiving the down-stepping. Finally, a steep falling Fo signals terminal juncture. Each rule is elaborated with secondary rules concerning relative Fo highs and lows, range and steepness of movement, and restrictions such as the fact that a grave-accent fall cannot be directly followed by another grave-accent fall. Identification of word accents gives identification of primary stress since a stressed vowel will generally have one of the two word accents. Thus, the identification of stressed and unstressed vowels is mainly arrived at indirectly, although the reader must also make use of duration and amplitude cues as formulated in rules 6 and 7.

RULE TESTING (NON-EXPERT READERS)

Two non-expert mingogram readers were given the same task as the expert reader. These readers spent nearly an hour working on each sentence. Of the 178 occurrences of the different prosodic categories, the first reader identified 138 (78%). A confusion matrix is presented in Figure 3. The second reader, given a new set of ten prosodically comparable sentences, identified 139 of 202 category occurrences (69%). See Figure 4 for the complete results.

The major difficulties for both non-expert readers were found, surprisingly enough, in the identification of grave and grave focal accents. The reason for this could lie in the ordering of the rules. The Fo fall for grave accent was deemed to be the easiest to identify and to be the most invariant category marker. Therefore this rule was placed first. The readers, however, showed a tendency to first reject Fo patterns that did not exactly fit the grave rule description. Once the grave category was rejected, the readers continued on to the following rules and may have felt forced to assign an acute or acute-focal-final category to the Fo fall even though the grave category would have been a better fit, simply because they had already rejected the grave category.

RULE TESTING (AUTOMATIC ANALYSIS)

The rules are currently being implemented on an automatic prosodic analysis system using a curve fitting program which is presented with the same data as the mingogram readers with the addition of segmentation information. At this stage of the project, the system uses a strategy that can be described in the following main steps:

1) Parametric description of the Fo contours of manually marked vowel segments
2) Selection of the "best" global parametric description for each vowel contour
3) Classification of the parameterized contours on the basis of a set of rules for identification of stress categories.

Confusion matrix for non-expert reader 1: sentence set 1

IDENTIFIED AS (in percent):

CATEGORY	G	GF	A	AF	AFT	TJ	SS	US	Ø	TOTAL
G	54		8		8			31		7/13 54%
GF		38	12		12			38		3/8 38%
A			78					22		18/23 78%
AF				85				15		11/13 85%
AFT					100					2/2 100%
TJ										
SS							73	27		27/37 73%
US							11	85	4	70/82 85%
Ø				(1)				(3)		

All categories = 138/178
78%

G = Grave
GF = Grave-focal
A = Acute
AF = Acute-focal
AFT = Acute-focal-terminal
TJ = Terminal juncture
SS = Stressed syllable
US = Unstressed syllable
Ø = Not a syllable or missed syllable
TOTAL = Total correct identification of the category in number and percent

Figure 3. Results of the mingogram reading experiment: Non-expert reader 1.

Confusion matrix for non-expert reader 2: sentence set 2

IDENTIFIED AS (in percent):

CATEGORY	G	GF	A	AF	AFT	TJ	SS	US	Ø	TOTAL
G	56		6	6	25		6			9/16 56%
GF	18	36			36		9			4/11 36%
A			71				5	24		15/21 71%
AF			14	71				14		10/14 71%
AFT					100					4/4 100%
TJ	18	9				54		18		6/11 54%
SS							90	10		35/39 90%
US							31	65	3	56/86 65%
Ø										

All categories = 139/202
69%

G = Grave
GF = Grave-focal
A = Acute
AF = Acute-focal
AFT = Acute-focal-terminal
TJ = Terminal juncture
SS = Stressed syllable
US = Unstressed syllable
Ø = Not a syllable or missed syllable
TOTAL = Total correct identification of the category in number and percent

Figure 4. Results of the mingogram reading experiment: Non-expert reader 2.

Confusion matrix for program testing 1.0: automatic parsing sentence set 2

IDENTIFIED AS (in percent):

CATEGORY	G	GF	A	AF	AFT	TJ	SS	US	Ø	TOTAL
G	81					6		12		13/16 81%
GF	27	54				9		9		6/11 54%
A	5	10	24					57	5	5/21 24%
AF	7	7		28				50	7	4/14 28%
AFT	25	25						50		0/4 0%
TJ	18	18	9			45		9		5/11 45%
SS							59	38	2	23/39 59%
US							10	88	1	76/86 88%
Ø										

All categories = 132/202
65%

G = Grave
GF = Grave-focal
A = Acute
AF = Acute-focal
AFT = Acute-focal-terminal
TJ = Terminal juncture
SS = Stressed syllable
US = Unstressed syllable
Ø = Not a syllable or missed syllable
TOTAL = Total correct identification of the category in number and percent

Figure 5. Results of rule testing: Automatic prosodic analysis system.

This simple curve-fitting procedure is being used as a rough test of the main components of the rules. A preliminary running of the program on the second set of sentences produced a promising 81% correct for grave accents. Out of the 202 category occurrences, the program identified 132 (65%). See Figure 5 for the complete results. The major difficulty was in identifying acute and acute focal accents, and final juncture. These preliminary results differ considerably from the results of the mingogram reading experiment. Performance may be improved if the parametrization of the contours, at least in its present form, is abandonned. Some of the errors result not from the rules themselves, but rather from the discrepancies between calculated and actual Fo values.

REFERENCES

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

Bruce, G. 1985. Structure and functions of prosody. In Guerin and Carré (eds.) Proceedings of the French Swedish Seminar on Speech, Grenoble.

Bruce, G. and E. Gårding. 1978. A prosodic typology for Swedish dialects. In Gårding et al. (eds.) Nordic Prosody. Department of Linguistics, Lund University, 219-228.

Gårding E. 1982. Swedish Prosody. Phonetica 39, 288-301.

Gårding, E. and G. Bruce. 1981. A presentation of the Lund model for Swedish intonation. In Fretheim (ed.) Nordic Prosody II. Tapir, Trondheim, 33-39.

Lea, W. 1980. Prosodic aids to speech recognition. In Lea (ed.) Trends in Speech Recognition, Prentice-Hall, N.J. 166-205.

Lindblom, B., B. Lyberg and K. Holmgren. 1981. Durational patterns of Swedish phonology. Do they reflect short-term memory processes? Indiana University Linguistics Club, Bloomington.

Lyberg, B. 1981. Temporal properties of spoken Swedish. Monographs from the Institute of Linguistics, University of Stockholm no. 6.

Svensson, S-G. 1974. Prosody and grammar in speech perception. Monographs from the Institute of Linguistics, University of Stockholm no. 2.

Vaissière, J. 1983. A suprasegmental component in a French speech recognition system: reducing the number of lexical hypotheses and detecting the main boundary. Recherches acoustiques CNET Lannion, 7 82/83.

Welin, C.W. and B. Lindblom. 1980. The identification of prosodic information from acoustic records by phoneticians. Journal of the Acoustical Society of America, 99th ASA Meeting, S 65.

APPENDIX 1: TEST SENTENCES

SENTENCE SET 1

1. Hon skriver ständigt om sina resor i sin dagbok.
2. Hon bekymrar sig över sitt skrikande barn.
3. Betoningen kommer sist i ordet.
4. Den intressantaste tavlan var en målning av molnen.
5. Mannen reser snart till Bollerup.
6. Vi flyger när vädret är perfekt.
7. Mannen målar, det vill säga han är yrkesmålare.
8. I gryningen ska vi vandra till flodens mynning.
9. Dom läser poesi hela natten.
10. Demonstranterna betraktades av polisen.

SENTENCE SET 2

1. Han skriver en klagodikt över sitt öde.
2. Hon lämnar mig många moderna målningar.
3. Flickan, hon som var på posten, skickade ett paket.
4. Det var den skämtande mannen som var starkast.
5. Hon talar försiktigt när hon vet att mikrofonen är på.
6. Vi får ofta arbeta på natten.
7. Dom har haft kontakt med honom via ett telefonsamtal.
8. Böcker och tidskrifter finns i huset bredvid.
9. Den vackra himlen inspirerade honom.
10. Mannen som gick, han med en blå jacka, betalade kontant.

APPENDIX 2: RULES USED BY MINGOGRAM READERS TO EXTRACT PROSODIC CATEGORIES

1. **Fo-FALL synchronized with a longer vowel having greater amplitude gives GRAVE** (risk of confusion with final juncture) (h)
    a) The beginning of the Fo fall is synchronized with the beginning (often abrupt onset) of a stressed vowel = longer duration + greater amplitude (6).
    b) Fo at the beginning of the fall is higher than or as high as the preceding Fo top.
    c) Fo at the end of the fall is markedly lower then the preceding Fo low.
    d) Range of the fall is greater than most other Fo movements.
    e) Steepness of the fall is relatively constant (45%).
    f) Duration of the fall runs throughout the entire vowel segment.
    g) **WARNING:** An Fo fall for grave accent cannot be directly followed by another grave-accent fall. A syllable containing a grave-accent fall is always followed by another syllable which belongs to the same word. This second syllable is either unstressed or carries secondary stress.
    h) **WARNING:** Risk of confusion with no. 5 below (acute+focus+final juncture)

2. **HIGH or RISING Fo in the vowel following an identified grave accent gives GRAVE+FOCUS**
    a) An Fo rise or a high Fo in the vowel following a grave-accent fall signals grave+focus.
    b) The Fo top in the vowel is almost as high, as high or higher than the previous Fo top.
    c) **NB!** If Fo is low during the vowel, the syllable is part of a non-focal grave accent.

3. **Fo-RISE synchronized with a longer vowel having greater amplitude gives ACUTE+FOCUS** (risk of confusion with initial juncture)
    a) The beginning of the Fo rise is synchronized with the beginning (often abrupt onset) of a stressed vowel = longer duration + greater amplitude (6).
    b) Fo at the beginning of the rise can be lower or the same as Fo in the preceding syllable.
    c) Fo at the end of the rise is higher than the preceding Fo top.
    d) Range of the rise is somewhat less than that of a grave accent fall and usually less than that of the rise for grave+focus (g).
    e) Steepness of the rise is relatively constant (30° - 60°) but varies more than the fall for grave accent.

f) The end of the rise (Fo top) comes near the boundary to the post-tonic syllable or in the post-tonic syllable.
g) **WARNING:** If the syllable comes at the beginning of an utterance, the range of the rise must be greater (= grave accent fall), otherwise the probable category is unstressed+initial juncture.
h) **WARNING:** If an identified grave accent fall is followed by one or more unstressed vowels separating the fall from an Fo rise in a stressed vowel, the sequence can be interpreted as either a compound word (grave+secondary stress) or as grave (non-focal) + acute focal. An Fo hump between the fall and rise signals grave + acute focal. If there is no hump the probable category is grave + secondary stress (compound word).

4. **DOWN-STEPPING of Fo from the previous vowel to a longer vowel having greater amplitude gives <u>ACUTE</u> (non-focal)**
   a) Fo is lower in a stressed vowel (longer duration + greater amplitude) than in the preceding vowel.
   b) Fo can even be falling in the stressed vowel. The important point is that Fo is higher in the pretonic syllable.
   c) The range of the downstepping or fall is less than that of the grave-accent fall.
   d) The range is less before focus, it can be greater after focus.

5. **Steep Fo-FALL preceded by a slight Fo-rise in a longer vowel having greater amplitude gives <u>ACUTE+FOCUS+FINAL JUNCTURE</u>**
   a) An Fo fall in the beginning of a stressed vowel. This fall can be preceded by a slight Fo rise in the same vowel.
   b) The end of the fall is often very low and is normally followed by a pause.
   c) The range of the fall is large and is very similar to the grave-accent fall.
   d) Steepness is constant and although similar to that of the grave-accent fall this fall is often steeper (70°).

6. **LONGER VOWEL + GREATER AMPLITUDE gives <u>STRESSED VOWEL</u>**

7. **SHORTER VOWEL + LESS AMPLITUDE gives <u>UNSTRESSED VOWEL</u>**

EFFECT OF TEMPO AND TONAL CONTEXT ON FUNDAMENTAL FREQUENCY CONTOURS IN JAPANESE

Yasuko Nagano-Madsen

Abstract

Effect of tempo and tonal context on fundamental frequency contours was studied by using Japanese reiterant speech "nanana". Eight logically possible tonal patterns spoken with statement intonation were tested at three tempos. The Fo values of a number of reference points in the middle of an utterance varied systematically as a function of tempo and tonal context while those of other points close to the beginning and end of an utterance were not much affected by the two variables. When the notion of "grid" is applied to represent the overall shape made by the eight Fo contours, the relationship between all the reference points appeared basically constant across tempo variation while the slope of grid varied systematically as a function of tempo. Increase of tempo gradually raised the Fo values of Lows in the middle of utterance so that the Fo contour takes a short-cut between the beginning, peak, and end reference points. The phrase final L was found to be least variable across tempo and contextual variations. The relevance of these phonetic findings to the phonological analysis of tone is discussed briefly.

Introduction

Many studies have been devoted to "reduction" and "coarticulation" phenomena primarily in relation to the invariance issue. These studies have typically dealt with segmentals and less with suprasegmentals. Very little is known, for example, as to how Fo patterns and the relation between Fo and time dimension are affected by stylistic and tempo variations. Likewise, not much is known as to how the phonetic mechanisms of tonal coarticulation are affected by tempo change. One study of tempo effects on tonal and temporal patterns in Swedish dialects was made by Gårding (1975).

As for the tonal(1) patterns of Japanese, it has been pointed out that "casual speech" and "fast speech" should not be equated (Nagano-Madsen, 1987). In casual speech, the tonal contrast is often reduced in pitch range and sometimes even neutralized phonetically. On the other hand, in formal speech such as laboratory recording, a speaker tends to speak clearly. He becomes more aware of the tonal patterns and tries to keep them as much as possible even when directed to speak very fast. Such fast speech sounds like a "shout" presumably because of increased vocal effort. Thus, reduction and neutralization phenomena often observable in casual speech tend not to appear as a mere function of fast tempo in the laboratory.

In the present paper, the results of an exploratory study of tempo and contextual effect on the acoustic manifestations of various tonal patterns in the Kochi dialect of Japanese are reported. It controlled two variables (tempo and context) while other aspects of speech such as formality of style, type of utterance (statement), vocal effort, and segmental composition were kept constant.

The main questions are: what are the acoustic manifestations of the tone represented as H and L in this dialect of Japanese; which features vary and which features remain constant across tempo variation; is the tempo variation systematically represented in the Fo contour; does the degree of tonal coarticulation increase as tempo increases; if so, in what direction are the phonetic mechanisms of coarticulation directed?

Experiment

The test material consisted of trisyllabic reiterant Japanese speech /nanana/ with eight logically possible tonal patterns using binary (H and L) analysis. A real word or a phrase(2) corresponding to a particular tonal pattern was written in brackets as a reference. The test material is shown below:

Tonal pattern	reference phrase	lexicon
1. HHH	/hasiga/	"the edge is"
2. HHL	/hasida/	"it is an edge"
3. HLL	/hasida/	"it is a bridge"
4. LHL	/hasida/	"it is a chopstick"
5. LHH	/hasiga/	"a chopstick is"
6. LLH	/hadasi/	"bare feet"
7. LLL	/(hi)basiga/	"a fire-chopstick is"
8. HLH	/hasida(me)/	"a bridge is no good"

Note that no. 7 and 8 do not occur in the same environment as the other test items. All these phrases were meant as possible answers to the question "what did you say?"

A female speaker (present author) of the Kochi dialect, subdivision of the Kyoto(3) dialect, read the list very clearly with statement intonation at four different rates of utterance. The four speeds were: very fast, normal, slow (approximately adagio) and very slow (approximately largo). For the two degrees of slow, a metronome was placed outside the studio window to prevent great deviation in tempo. A minimum of five tokens were obtained for each tonal pattern at each rate. An informal listening test of the recorded material confirmed that all the words were perceived as having the tonal pattern intended.

A second female informant of the Kyoto dialect found the reiterant version difficult. Consequently, only (near) minimal sets of HLL and LLH patterns were recorded at three different rates for comparison.

Mingograms showing intensity curves, fundamental frequency curves, and duplex oscillograms were made for the entire material with a paper speed of 50 mm/s. Syllable duration and word duration were measured to the nearest 5 ms.

The beginning (BEG) and the syllable boundary (i.e. at the end of the vowel [a]) were chosen as reference points. A syllable boundary was chosen as a reference because when there is a clear turning point(4), it typically occurs there. The syllable boundaries were designated L1, H2 etc. depending on the position in the phrase and on whether they correspond to a High tone or a Low tone. H3 and L3 generally correspond to the end of the utterance, except for some cases at slow tempo where there is a Fo peak in the last syllable followed by a slight fall. For example, the HLL pattern has the following reference points: BEG, H1, L2, L3 (=END).

It was found that the slowest version (largo) had frequent short silent intervals within the utterance. Consequently, only one slow tempo (adagio) was measured and used for the discussion.

Results and Discussion

1. Acoustic manifestation of H and L

What are the acoustic manifestations of the H and L tones in

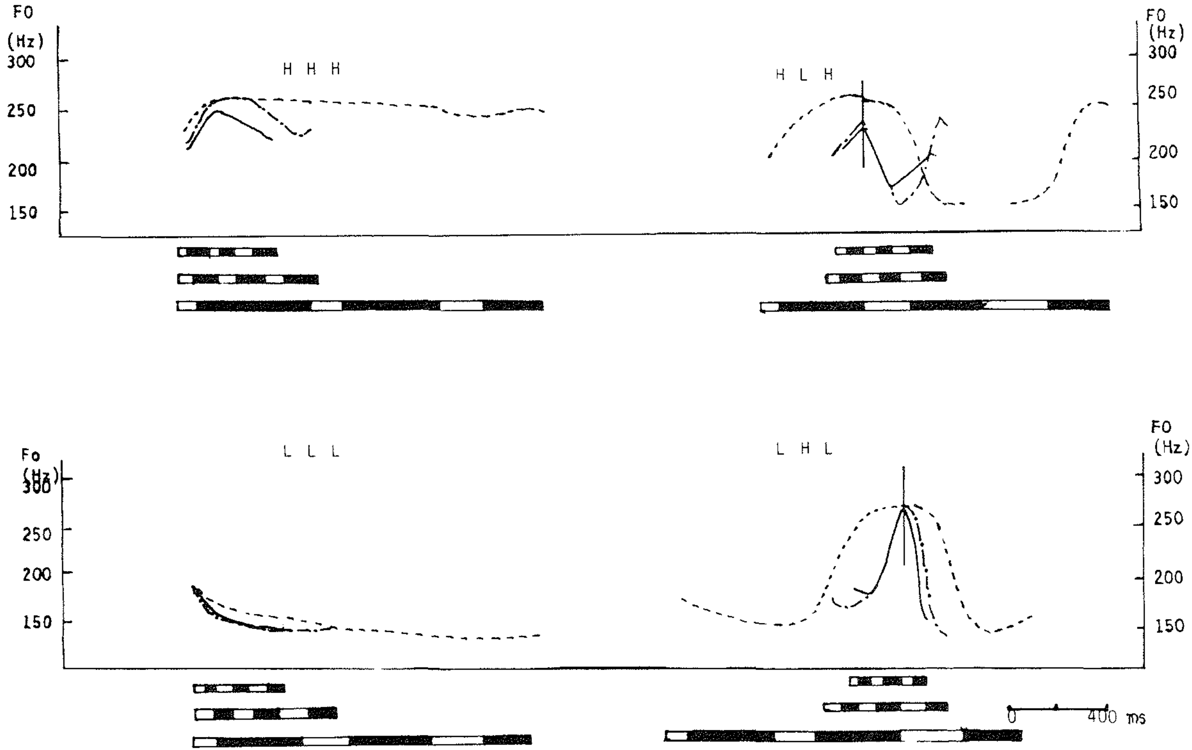


Fig. 1a. Superimposed, typical Fo-contours of the test phrases at different tempos and their timing patterns. The line-up point is at the beginning for HHH and LLL and at the VC-boundary of HL for HLH and LHL.

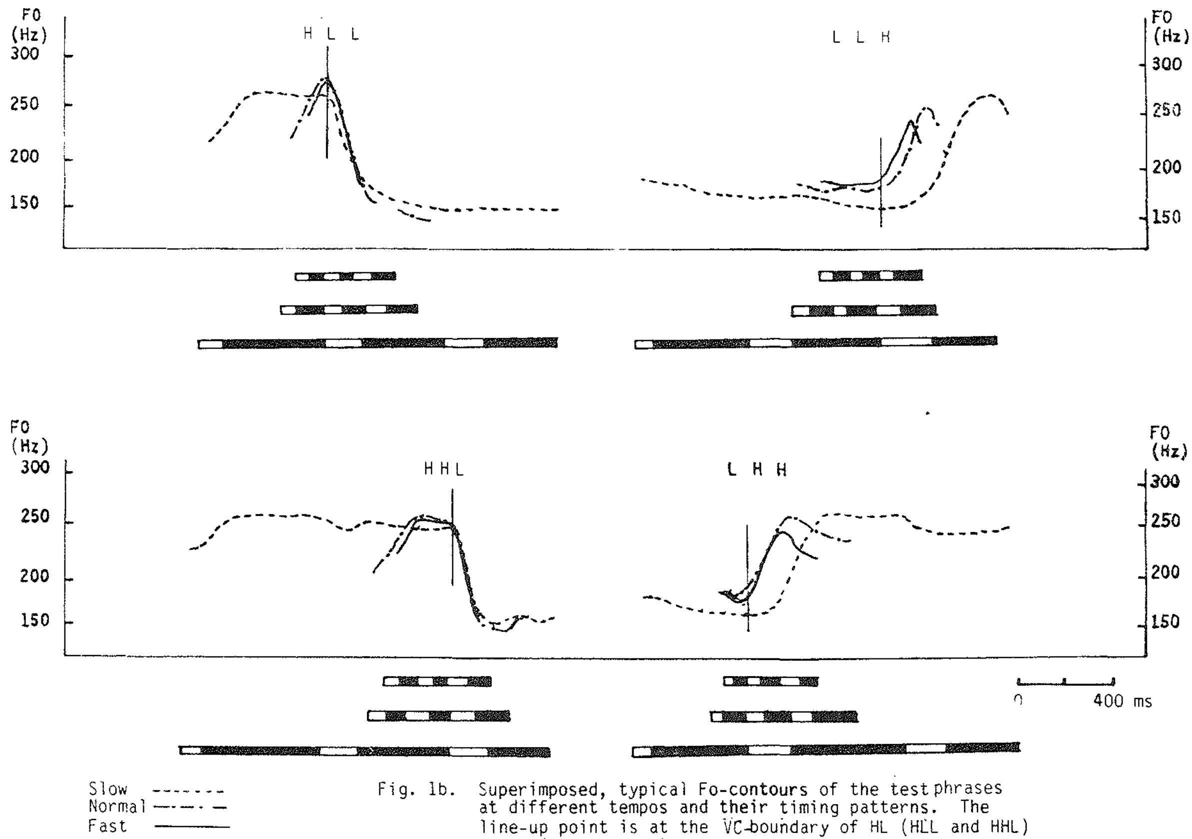


Fig. 1b. Superimposed, typical Fo-contours of the test phrases at different tempos and their timing patterns. The line-up point is at the VC-boundary of HL (HLL and HHL) or LH (LLH and LHH).

statement intonation in this dialect of Japanese? Figures 1 (a) and (b) show mingogram tracings of Fo and timing for a typical sample of each tonal pattern at different tempos. The line-up point is BEG for the level patterns (HHH and LLL) and the VC boundary where the first pitch change occurs for other patterns. It is only at slow tempo that Hs and Ls can be identified as a continuation of Fo at a certain pitch level, i.e. about 140 to 165 Hz for a L tone and about 240 to 270 Hz for a H tone. At normal and fast speeds, the Fo contour is continuously moving up and down unless there is a succession of the same tone elements.

The timing of Fo in relation to segments is a keynote in accent analyses in Japanese as well as in Swedish. The present data from the Kochi dialect of Japanese also show that the most constant acoustic feature across tempo is the relationship between rise and fall in connection with the segmental timing. Since the test material used only /nanana/ in order to avoid Fo variations caused by segments, these relationships appear regularly. Unless preceded by the same tone element, H is typically associated with Fo rise while L is associated with Fo fall, pitch change taking place in the vicinity of the syllable boundary. However, the association with Fo fall for a L is only clearly manifested when it is preceded by a H. In initial position, the manifestation of a L is basically level Fo but may decline or slightly move upward depending on tempo and tonal context.

In the present data, the maxima before the fall were the most easily identifiable turning points and were more rigorously fixed to the syllable boundary than the other kind of turning point, i.e. minima before rise (cf. HLL and LLH in fig. 1b). Since the same situation has been observed in a number of studies of intonation from different languages, this may be an automatic consequence of the production mechanisms involved. In relation to the asymmetry between a fall and a rise, Hirose (1981) notes that "the activation of muscle is achieved by asynchronous excitation of many different motor units, whereas at the time of relaxation all the units can stop their activity almost synchronously".

2. Effect of tempo on Fo

How is the tempo variation manifested in the acoustic signals? Is it only the time dimension that is affected while the frequency dimension remains basically the same? Figures 1 (a) and (b) show that all the Fo curves do undergo change in pitch range as well as change in temporal pattern.

An increase of tempo most notably affects the level portion of the Fo contour making the curves look steeper while the rising and falling parts are left less affected.

In the frequency dimension, the degree and direction of change differs in a more complicated way. The Fo values for Ls and Hs as well as BEG are presented in Tables 1 and 2. An overall tendency is that of centralization, i.e. Fo values of Hs become lower and those of Ls become higher as tempo increases. This tendency appears quite regularly for all Hs and Ls except for L3 whereas BEG and L3 were less affected by tempo variation.

The foregoing observations make more sense when the same Fo contours in Figures (1) and (2) are rearranged according to each tempo (cf. Figure 2). It appears that the eight Fo contours together make an overall shape that becomes gradually steeper as tempo increases. Within this overall shape, the relationship between Hs and Ls is basically constant except for the HLH pattern for which L and Hs are much centralized.

In a model of intonation developed at Lund, such a global shape is interpreted as a manifestation of declarative sentence intonation and expressed by two (upper and lower) reference lines whose slope is supposed to be dependent on the length of the phrase and the initial and final frequencies (Bruce and Gårding, 1978). Later the term "grid" is given to those reference lines (Gårding, 1983). The tendency for the slope of the grid or top line to vary according to tempo is reported for Chinese (Gårding and Zhang, 1987), English (Cooper and Sorensen, 1981), and Swedish (Gårding, 1975). The role of grid (phrase component) for intonational phonology has been discussed at some length by Ladd (1984, 1985).

3. Contextual effect

How are the Fo values of Ls and Hs affected by their tonal context? Does L1 become gradually higher as it changes its environment from _LL, _LH, _HL, and _HH? Is H2 higher in the environment of H_H than in the environment of L_L? Does the influence of context increase as tempo increases? If so, in what way?

Contextual effects were examined from the figures presented in Table 1. It was seen that even at slow tempo, the effect of context was present. For example, the Fo value of L2 is lowest in LLL and highest in LLH at all tempos. The difference in Fo values between

Table 1. Mean Fo values in Hz ($\bar{x}$), standard deviations (s), and the number of tokens (n) for each Low and High in different tonal context.

Tone	Context	Slow			Normal			Fast		
		$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n
L1	L H H	165	5.5	7	191	8.2	5	190	5.7	5
	L H L	165	5.5	7	191	8.2	5	186	5.1	5
	L L H	172	5.5	6	183	5.1	5	181	5.3	5
	L L L	150	0	6	157	3.7	5	161	2.0	5
L2	L L H	161	5.3	6	184	3.7	5	189	4.1	5
	H L L	150	0.7	7	173	3.4	5	184	10.1	5
	H L H	153	3.8	6	163	6.0	5	181	1.9	6
	L L L	142	1.8	6	150	7.1	5	153	3.1	5
L3	H H L	160	0.8	5	156	4.9	5	laryngealized		
	H L L	148	2.0	7	142	1.9	5			
	L H L	143	2.1	5	148	3.2	5			
	L L L	141	1.0	5	141	0.8	5			
H1	H L L	262	2.1	7	273	2.9	5	260	6.9	5
	H H L	260	2.7	7	262	6.2	5	255	1.5	6
	H H H	263	8.0	7	252	6.0	5	243	2.0	6
	H L H	264	3.4	6	232	8.0	5	241	5.2	6
H2	L H L	271	1.2	6	268	4.1	5	262	3.0	5
	L H H	260	3.5	7	257	4.3	5	255	4.8	5
	H H L	252	5.2	8	257	2.9	5	247	3.6	6
	H H H	257	7.5	7	251	6.6	5	244	5.6	6
H3	L L H	261	5.8	6	252	7.9	5	242	1.8	5
	L H H	242	6.4	7	230	6.9	5	234	4.0	5
	H H H	250	5.0	7	217	5.1	5	226	4.5	6
	H L H	240	8.7	6	237	3.3	5	209	6.7	6

Table 2. Mean Fo values in Hz ($\bar{x}$), standard deviations (s), and the number of tokens (n) for BEG in different tonal context and tempo.

Context	Slow			Normal			Fast		
	$\bar{x}$	s	n	$\bar{x}$	s	n	$\bar{x}$	s	n
L L H	183	3.4	6	183	6.0	5	178	5.7	5
L H H	185	4.4	7	189	9.8	5	174	4.9	5
L H L	176	3.4	6	172	4.3	5	173	3.4	5
L L L	170	3.9	6	172	6.4	5	168	4.1	5
H H L	238	9.6	7	224	5.5	5	221	2.7	6
H H H	226	3.3	6	218	5.1	5	222	3.7	6
H L L	223	5.2	7	222	2.7	5	226	8.6	5
H L H	214	4.5	6	209	4.9	5	223	4.7	6

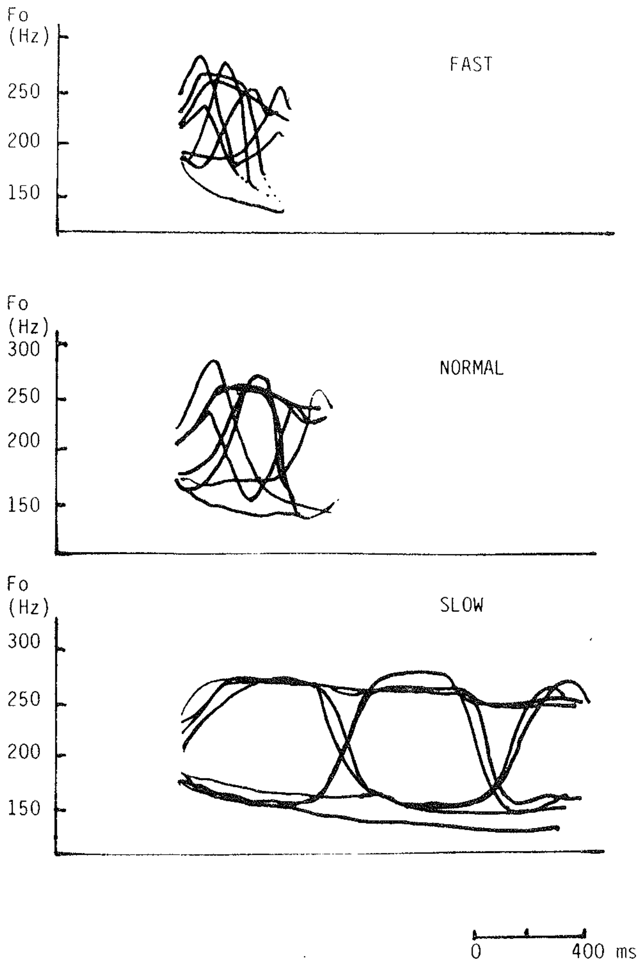


Fig. 2. Superimposed, typical Fo-contours of the test phrases grouped according to the tempo.

the two groups increases with tempo. At fast tempo, this difference comes close to 40 Hz. L3 appeared to be the least variable tone across tempo and contextual variations.

Contextual effects examined within the same tempo showed that all Low tones were affected in their Fo values not only by an adjacent but also by a non-adjacent High tone in the same word. On the other hand, High tones were not affected by any Low tones except for H1 at normal and fast speed and H3 at fast speed in the HLH pattern. Note that the Fo values for L1 and L2 increase systematically with tempo. Thus, the Fo contour corresponding to the first two syllables in the LLH pattern moves downwards before a rise at slow tempo but becomes level at normal tempo and then becomes a gradual rise at fast tempo. Since the Fo values of BEG and END points were not much affected by tempo, it looks like that the Fo contour takes a gradual short-cut from BEG to H and to the END.

Contrary to our preliminary assumption, the Fo value of H2 was highest in the environment of L_L while it was lowest in the environment of H_H at all tempos. Clearly, the tonal coarticulation mechanism can not be captured, at least for this language, by a simple HL analysis.

Why is the assimilation rule asymetrical, i.e. why can HLH become HHH but LHL not become LLL? Why does word final L often resist assimilation rules? These questions can be answered, at least in part, when phonetic mechanisms involved in the production of tone and intonation are understood. In the foregoing section, it has been shown that the recognition of the grid makes the change in Fo values for Hs and Ls consistent, which conforms well with the perceptual impression. Without this notion, the Fo changes are difficult to explain. The L3 tone was found to be least affected by tempo and context effect.

4 Comparison with the second informant

The main difference between the two informants was that of pitch range. For the first informant, the total pitch range varied from 140 Hz to 270 Hz at slow tempo. At slow tempo this informant used basically the same pitch range for Hs and Ls regardless of the tonal pattern in which they occurred. The second informant used a range extending from 160 Hz to 240 Hz at slow tempo. She has two sets of pitch ranges for Hs and Ls depending on whether they are part of a rising movement (e.g. LLH) or part of a falling movement (eg. HLL). The pitch range used for the rising movement was 20-30 Hz more reduced (centralized) than that of falling one. This difference may be due to stylistic variation. The first informant read the list very clearly (hyper speech) while the second informant read it more naturally. It should be also noted that the two dialects (Kochi and Kyoto) have considerable differences impressionistically.

Conclusion

The purpose of this paper has been to show how Fo changes with tempo and to show how tempo affects tonal coarticulation. The major points made were: (1) Fo change as a function of tempo manifests itself as a slope of the grid lines. Seen within the grid, the change in Fo due to tempo did not alter the relationship between Ls and Hs. (2) The mechanisms of tonal coarticulation can not be captured in terms of such linguistic categories as H and L alone. Instead the steady reference points, grid, and the production mechanisms involved for rise and fall should be considered.

The present study has dealt with only limited material. Some of the points made in this paper, therefore, may not be characteristic of more natural and longer utterances from real life. Since word citation forms are known to be inadequate for many other features in prosody (Bruce, 1977), another study is in preparation in which each tonal pattern appears in a different position and context in relation to a sentence.

Acknowledgments

I would like to thank Gösta Bruce and Eva Gårding for their very helpful comments and advice on earlier versions of this paper.

Footnotes

(1) Tokyo Japanese is considered to be a typical example of a pitch-accent system while some dialects of Japanese spoken in the Kansai area have more complicated systems. In this paper, the term "tonal" was used throughout without making distinction between "tonal" and "accentual".

(2) In the original list, this was written in Chinese characters.

(3) It has also been called "Osaka" or "Kansai" dialect.

(4) The turning points (local Fo maxima and minima) were found to be useful for the analysis of intonation in many languages (Gårding, 1983).

Bibliography

Bruce, Gösta. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Gleerups, Lund.

Bruce, Gösta and Gårding, Eva. 1978. A prosodic typology for Swedish dialects. In Nordic Prosody, Department of Linguistics, Lund University:219-228.

Cooper, William and Sorensen, John. 1981. Fundamental frequency in sentence production. Heidelberg, Springer.

Gårding, Eva. 1975. The influence of tempo on rhythmic and tonal patterns in three Swedish dialects. Working Papers 12, Department of Linguistics, Lund University:71-84.

Gårding, Eva. 1983. A generative model of intonation. In Cutler, A. & Ladd, D.R. (eds.) Prosody:Models and Measurements. Heidelberg, Springer.

Gårding, Eva and Zhang, Jialu. 1987. Tempo and prosodic patterns in Chinese and Swedish. Working papers 31, Department of Linguistics, Lund University.

Hirose, Hajime. 1981. Comments in the discussion after Fujisaki et al. 's paper. In Stevens, K. & Hirano, M. (eds.) Vocal Fold Physiology. University of Tokyo Press:361.

Ladd, D. Robert. 1984. Declination: a review and some hypotheses. Phonology yearbook, 1:53-74.

Ladd, D. Robert. 1985. Intonational phonology is not a single autosegmental tier. Work in progress 11, Department of Linguistics, Edinburgh University:47-59.

Nagano-Madsen, Yasuko. 1987. Stylistic variations in the production of Japanese accent. Paper presented at the Mini Conference on Stylistics, Stockholm, May 4-5.

MULTILANGUAGE TRANSLATION OF NUMERALS BY PROLOG

Bengt Sigurd

INTRODUCTION

The present paper describes a program (Nummer.pro), which has been developed for PC (MSDOS) in PDC prolog (Prolog Development Center, DK-2605 Broendby). It translates numerals (English, Burmese, Georgian, French and Japanese) and is based on a general theory of numerals, which presumably can be applied to all numeral systems in the world, although with varying success - related to number of exceptions. Languages will show different types of deviations from the general model (cf below).

The general model accounting for the structure of numerals is formalized by the following rule:

Numeral --> ...(U G3)(U G2)(U G1)(U G0)

In this generative rule U is a unit numeral and G0,G1,G2,etc. are group numerals. The English numeral "two hundred" consists of the unit (U) numeral "two" and the group (G) word "hundred", which is G2 in the formula. There are no other constituents in this case. An expression within parentheses is called a group constituent. The meaning (value) of such a group constituent is the meaning of the unit word multiplied by the meaning of the group word, in this case 2*100=200, as expressed in the ordinary (decimal) mathematical system. A numeral may consist of one or several such group constituents. The numeral "three hundred and thirty three" consists of three group constituents (disregarding "and"):

(three hundred)(thir ty)(three)

Some of the irregularities hinted at above show up in this example. The unit numeral for 3 has different forms (the allomorphs three/thir) in its different occurrences and it seems to be more natural to treat thirty, and the numeral thirteen, as complexes not to be further analyzed synchronically (only etymologically). The English example also shows that there is no group word in the last constituent ("three"). This is the norm in languages, but there are languages such as Burmese, which have unit group words.

The group words in English are neat powers of ten (10), (10*10=100), (10*10*10=1000). We note that group words for 10000, 100000 are lacking in English - but not in the Burmese language, as will be shown. Group words with other values, such as 20 in French and many other languages imply multiplication by 20, etc. We note that the group words are ordered in increasing order right to left. There are obvious exceptions from this ordering rule in teens, e.g. "sixteen" if we assume that "teen" is the representative of the group word with the value 10. The teens are special. In German 32 would be "zwei-und-dreissig" which shows that the units may be placed before the tens even in higher numbers.

The "packing strategy" defined in Hurford(1975) means that the objects to be counted are packed in groups (containers) that are as big as possible. If there are 5200 objects, one does not say "fifty-two hundred", but "five thousand two hundred". If there are 62 objects, one does not say "sixty-twelve" in a decimal culture (but in a vigesimal such as the French). There are exceptions to this strategy, however. If years are counted, one says "fifteen hundred (and) twenty" instead of "one thousand five (hundred and) twenty". It is a moot question whether this violation of the packing strategy will be carried over into the twenty-first century. Will people say e.g. "twenty-one hundred (and) sixty-four" or "two thousand one hundred (and) sixty-four"? The Swedish Academy recommends the first solution for Swedish, but Swedes seem to hesitate. Teens are special and may be allowed to violate the packing strategy.

The meaning (value) of a string of constituents is also clear from our examples: the meaning of the whole expression is the sum of the values of the constituents. The derivation of the value of "three hundred and thirty-three" can then be illustrated by (3*100)+(3*10)+(3)=333. There is thus multiplication within the constituents, but addition between the constituents. Addition and multiplication are the standard operations in numeral systems, but subtraction has also been found, e.g. in Yoruba, where 105 may be expressed either as (20*5) + 5 or (5 from (10 from (20*6))). For detailed information about numerals see Hurford(1975) and Menninger (1958).

OVERVIEW

The program nummer.pro consists of modules for each language, and new modules can easily be added. Each language module understands the numerals of that language, i.e. gives the mathematical value if the operator writes check(lista) with a prefix, e.g. "jcheck" for Japanese, "gcheck" for Georgian, "bcheck" for Burmese. Each language module can also express a number in the language. This is done by writing say(number) with a prefix, e.g. fsay(89) which will return "quatre-vingt-dix-neuf".

Translation is done by deriving the mathematical value of the numeral of the source language by "check" and giving that number to "say" of the target language. The mathematical representation thus serves as an intermediate representation. A number of predicates are defined such as egtra(lista) which calls the English echeck and then Georgian say (gsay). These translation predicates are included in the language modules. The system is thus a multitranslation system for a special field. As the mathematical representations serve as a universal/international language in human communication such a multitranslation system for numerals may not seem useful, but there are certainly situations when it can be used. We note that the difficult task of representing the meaning of natural language is solved for numerals by using the mathematical representation and that the meaning (value) of the whole expression can nicely be computed from the meanings of the constituents.

The modules for English and French are the most comprehensive programs.

UNDERSTANDING ENGLISH NUMERALS

Following is a Prolog program for English numerals up to 999999 (for certain languages less). It understands an English numeral, i.e. gives its decimal values if the numeral is written as an argument (a list within square brackets) of echeck, e.g. echeck([five,hundred,an,fifty,five]). (The word "and" is a built-in predicate, that is why "an" is used.) The program will return the value 555. Note that commas are needed within the words. The program can also go the other way and give numerals for decimal numbers. If we write esay(67), the program will return "sixty, seven". The PDC-Prolog gives an output with quotation marks around all list constituents, which is annoying but this is a matter of cosmetics. An improved print-out without quotation marks and commas is produced by prefixing "cosm" and writing e.g. cosmesay(67).

The program is based on the model presented above, but there are certain irregularities to observe. The teens have been taken as a special group. They are special as they denote 10 by "teen" and have the unit word before, not after as the word for 10 in the "ty"- numerals. The numeral "ten" is considered as a "ten" numeral in the program. The words "eleven","twelve", illustrate words which are best treated as unanalyzable although etymologically they include words equivalent to "one","two" and "leave" ("what is left over ten").

The consituent "num" handles hundreds with or without tens and units. This constituent is then used recursively before thousands in order to handle numerals as high as 999999 (although this particular Prolog takes time to handle such high numbers). We do not cover millions and billions in this exploratory program. The comments are meant to facilitate the understanding of the program.

A PROLOG PROGRAM UNDERSTANDING ENGLISH NUMERALS */

```
code=4500 /* allows memory organization */
domains
tal=real
lista=string*
predicates
num(tal,lista,lista)
enum(tal,lista,lista)
a(lista,lista)
uni(tal,lista,lista)
teen(tal,lista,lista)
deci(tal,lista,lista)
hund(tal,lista,lista)
tus(tal,lista,lista)
cent(tal,lista,lista)
mil(tal,lista,lista)
hy(lista,lista)
esay(tal)
cosmesay(tal) /* with cosmetic concatenation of morphs */
echeck(lista)
cosme(lista)
clauses
num(A,B,B1) :- uni(A,B,B1). /* e.g. two */
num(A,B,B1) :- teen(A,B,B1). /* e.g. twelve */
num(A,B,B1) :- deci(A,B,B1). /* e.g. twenty */
num(D,B,B3) :- deci(A,B,B1),hy(B1,B2),uni(C,B2,B3),D=A+C./* fifty-one */
num(A,B,B1) :- cent(A,B,B1). /* e.g. six hundred */
num(D,B,B3) :- cent(A,B,B1),a(B1,B2),uni(C,B2,B3),D=A+C. /* adds */
num(D,B,B3) :- cent(A,B,B1),a(B1,B2),teen(C,B2,B3),D=A+C.
num(D,B,B3) :- cent(A,B,B1),a(B1,B2),deci(C,B2,B3),D=A+C.
num(E,B,B5) :- cent(A,B,B1),a(B1,B2),deci(C,B2,B3),hy(B3,B4),uni(D,B4,B5
   ,E=A+C+D.
enum(A,B,B1) :- num(A,B,B1).
enum(A,B,B1) :- mil(A,B,B1).
enum(D,B,B3) :- mil(A,B,B1),a(B1,B2),uni(C,B2,B3),D=A+C.
enum(D,B,B3) :- mil(A,B,B1),a(B1,B2),teen(C,B2,B3),D=A+C.
enum(D,B,B3) :- mil(A,B,B1),a(B1,B2),deci(C,B2,B3),D=A+C.
enum(D,B,B2) :- mil(A,B,B1),num(C,B1,B2),D=A+C.
cent(D,B,B2) :- uni(A,B,B1),hund(C,B1,B2),D=A*C. /* multiplies */
mil(D,B,B2) :- num(A,B,B1),tus(C,B1,B2),D=A*C.
a([an|X],X). /* and occupied */
uni(1,[one|X],X).
uni(2,[two|X],X).
uni(3,[three|X],X).
uni(4,[four|X],X).
uni(5,[five|X],X).
uni(6,[six|X],X).
uni(7,[seven|X],X).
uni(8,[eight|X],X).
uni(9,[nine|X],X).
deci(10,[ten|X],X).
deci(20,[twenty|X],X).
deci(30,[thirty|X],X).
deci(40,[fourty|X],X).
deci(50,[fifty|X],X).
deci(60,[sixty|X],X).
deci(70,[seventy|X],X).
deci(80,[eighty|X],X).
deci(90,[ninety|X],X).
```

```
teen(11,[eleven|X],X).
teen(12,[twelve|X],X).
teen(13,[thirteen|X],X).
teen(14,[fourteen|X],X).
teen(15,[fifteen|X],X).
teen(16,[sixteen|X],X).
teen(17,[seventeen|X],X).
teen(18,[eighteen|X],X).
teen(19,[nineteen|X],X).
hund(100,[hundred|X],X).
tus(1000,[thousand|X],X).
echeck(X) :- enum(D,X,[]),write(D),nl. /* understands numeral X */
esay(X) :- enum(X,D,[]),write(D),nl. /* gives numeral for X */
cosmesay(X) :- enum(X,D,[]),cosme(D).
cosme(X) :- X=[H|T],T=[],write(H),nl.
cosme(X) :- X=[H|T],T=[H1|T1],concat(H,H1,H2),X1=[H2|T1],cosme(X1).
```

A PROLOG PROGRAM UNDERSTANDING BURMESE NUMERALS

The following is a representative set of Burmese numerals - somewhat simplified and with commas between morphemes.

1 ta/ta,ku	11 hse,ta	10 hse/ta,hse	21 hna,hse,ta
2 hna	12 hse,hna	20 hna,hse	22 hna,hse,hna
3 thoun	13 hse,thoun	30 thoun,hse	33 thoun,hse,thoun
4 lei	14 hse,lei	40 lei,hse	44 lei,hse,lei
5 nga	15 hse,nga	50 nga,hse	55 nga,hse,nga
6 hcau	16 hse,hcau	60 hcau,hse	66 hcau,hse,hcau
7 hkun	17 hse,hkun	70 hkun,hse	77 hkun,hse,hkun
8 huy	18 hse,hyi	80 hyi,hse	88 hyi,hse,hyi
9 kou	19 hse,kou	90 kou,hse	99 kou,hse,ta
100 ta,ya		101 ta,ya,ta	
1000 ta,htaun		654 hcau,ya,nga,hse,nga	
10000 ta,thaun		1965 ta,htaun,kou,ya,hcau,hse,nga	
100000 ta,thein		6789 hcau,htaun,hkun,ya,hyi,hse,kou	

The structure of the Burmese numerals is given by the rule:

bnumeral --> ..(U thein)(U thaun)(U htaun)(U ya)(U hse)(U ku)

The group words are powers of 10 in increasing order. The numeral for 1 can be with or without "ku", a group word for 1, roughly with the meaning 'pieces'. The unit word 1 may be optionally deleted before "hse" (10), in the numeral for 10; in the teens obligatorily. This cosmetic rule has not been implemented in this explorative program */

```
predicates
bnum(tal,lista,lista)
buni(tal,lista,lista)
bteen(tal,lista,lista)
bdeci(tal,lista,lista)
bhund(tal,lista,lista)
bcent(tal,lista,lista)
btus(tal,lista,lista)
bmil(tal,lista,lista)
btetus(tal,lista,lista)
bdmil(tal,lista,lista)
bsay(tal)
bcheck(lista)
betra(lista)
ebtra(lista)
clauses
bnum(A,B,B1)  :- buni(A,B,B1).
bnum(A,B,B1)  :- bdeci(A,B,B1).
bnum(D,B,B2)  :- bdeci(A,B,B1),buni(C,B1,B2),D=A+C.
bnum(A,B,B1)  :- bcent(A,B,B1).
bnum(D,B,B2)  :- bcent(A,B,B1),buni(C,B1,B2),D=A+C.
bnum(D,B,B2)  :- bcent(A,B,B1),bdeci(C,B1,B2),D=A+C.
bnum(E,B,B3)  :- bcent(A,B,B1),bdeci(C,B1,B2),buni(D,B2,B3),E=A+C+D.
bnum(A,B,B1)  :- bmil(A,B,B1).
```

```
bnum(D,B,B2)  :- bmil(A,B,B1),buni(C,B1,B2),D=A+C.
bnum(D,B,B2)  :- bmil(A,B,B1),bdeci(C,B1,B2),D=A+C.
bnum(E,B,B3)  :- bmil(A,B,B1),bdeci(C,B1,B2),buni(D,B2,B3),E=A+C+D.
bnum(D,B,B2)  :- bmil(A,B,B1),bcent(C,B1,B2),D=A+C.
bnum(E,B,B3)  :- bmil(A,B,B1),bcent(C,B1,B2),bdeci(D,B2,B3),E=A+C+D.
bnum(E,B,B3)  :- bmil(A,B,B1),bcent(C,B1,B2),buni(D,B2,B3),E=A+C+D.
bnum(F,B,B4)  :- bmil(A,B,B1),bcent(C,B1,B2),bdeci(D,B2,B3),buni(E,B3,
   B4),F=A+C+D+E.
bnum(A,B,B1)  :- bdmil(A,B,B1).
bnum(D,B,B2)  :- bdmil(A,B,B1),bmil(C,B1,B2),D=A+C.
bnum(D,B,B2)  :- bdmil(A,B,B1),bcent(C,B1,B2),D=A+C.
bnum(D,B,B2)  :- bdmil(A,B,B1),bdeci(C,B1,B2),D=A+C.
bnum(D,B,B2)  :- bdmil(A,B,B1),buni(C,B1,B2),D=A+C.
bnum(E,B,B3)  :- bdmil(A,B,B1),bmil(C,B1,B2),bcent(D,B2,B3),E=A+C+D.
bnum(E,B,B3)  :- bdmil(A,B,B1),bmil(C,B1,B2),bdeci(D,B2,B3),E=A+C+D.
bnum(E,B,B3)  :- bdmil(A,B,B1),bmil(C,B1,B2),buni(D,B2,B3),E=A+C+D.
bnum(E,B,B3)  :- bdmil(A,B,B1),bcent(C,B1,B2),bdeci(D,B2,B3),E=A+C+D.
bnum(E,B,B3)  :- bdmil(A,B,B1),bcent(C,B1,B2),buni(D,B2,B3),E=A+C+D.
bnum(E,B,B3)  :- bdmil(A,B,B1),bdeci(C,B1,B2),buni(D,B2,B3),E=A+C+D.
bnum(F,B,B4)  :- bdmil(A,B,B1),bmil(C,B1,B2),bcent(D,B2,B3),bdeci(E,B3,
   B4),F=A+C+D+E.
bnum(F,B,B4)  :- bdmil(A,B,B1),bmil(C,B1,B2),bcent(D,B2,B3),buni(E,B3,
   B4),F=A+C+D+E.
bnum(F,B,B4)  :- bdmil(A,B,B1),bcent(C,B1,B2),bdeci(D,B2,B3),buni(E,B3,
   B4),F=A+C+D+E.
bnum(G,B,B5)  :- bdmil(A,B,B1),bmil(C,B1,B2),bcent(D,B2,B3),bdeci(E,B3,
   B4),buni(F,B4,B5),G=A+C+D+E+F.
bdeci(D,B,B2)  :- buni(A,B,B1),bteen(C,B1,B2),D=A*C.
bcent(D,B,B2)  :- buni(A,B,B1),bhund(C,B1,B2),D=A*C.
bmil(D,B,B2)  :- buni(A,B,B1),btus(C,B1,B2),D=A*C.
bdmil(D,B,B2)  :- buni(A,B,B1),btetus(C,B1,B2),D=A*C.
buni(1,[ta|X],X).
buni(2,[hna|X],X).
buni(3,[thoun|X],X).
buni(4,[lei|X],X).
buni(5,[nga|X],X).
buni(6,[hcau|X],X).
buni(7,[hkun|X],X).
buni(8,[hyi|X],X).
buni(9,[kou|X],X).
bteen(10,[hse|X],X).
bhund(100,[ya|X],X).
btus(1000,[htaun|X],X).
btetus(10000,[thaun|X],X).
bcheck(X)  :- bnum(A,X,[]),write(A),nl.
bsay(X)  :- bnum(X,A,[]),write(A),nl.
betra(X)  :- bnum(A,X,[]),esay(A). /* transl Burm X into Eng */
ebtra(X)  :- enum(A,X,[]),bsay(A). /* transl Eng X into Burm */
```

TRANSLATING BETWEEN ENGLISH AND BURMESE

The English program presented above can give the mathematical representation when given an English numeral, or the English numeral when given the mathematical representation. Similarly, the Burmese program gives the mathematical representation, e.g. bcheck([lei,ya,nga]), which results in 405, or the Burmese numeral when given a number, e.g. bsay(4321), which gives "lei,htaun,thoun,ya,hna,hse,ta". The predicate betra(X), has been defined to handle translation between Burmese and English. It means: translate the Burmese numeral X into an English numeral. The predicate ebtra(X) translates from English into Burmese in a similar way.

GEORGIAN NUMERALS

Georgian numerals are illustrated by the following set.

1 erti	11 tertmeTi	10 ati	21 ocdaetri
2 ori	12 tormeTi	20 oci	22 ocdaori
3 sami	13 cameTi	30 ocvdaati	33 ocdacameTi
4 otxi	14 totxmeTi	40 ormoci	44 ormocidaotxi
5 xuti	15 txutmeTi	50 ormodaati	55 ormodatxutmeTi
6 ekvsi	16 tekvsmeTi	60 samoci	66 samocidaekvsi
7 Svidi	17 CvidmeTi	70 samocdaati	77 samocdaCvidmeTi
8 rva	18 tvrameTi	80 otxmoci	88 otxmocdarva
9 cxra	19 cxrameTi	90 otxmocdaati	99 otxmocdacxrameTi
100 asi	101 aserti	110 asati	121 asocdaerti
200 orasi	202 orasori	220 orasoci	240 orasormoci
300 samasi	355 samasormodatxutmeTi		
1000 atasi		2001 ori atas erti	
10000 ati atasi		1969 atascxraassamocdacxra	

The main structure of the Georgian numerals can be given by the following rule.

gnum --> ..(U X)(U asi)(U oci)(U ati)(U)

The Georgian system is a vigesimal system, a system where 20 assumes the value of the group word between 10 and 100.
The formula above is in fact too compact and does not indicate where the additive marker (da) occurs.

The structure is shown by the following formula:

gnum --> .. (gnum as)(U oc) (da gtuni) -i

A Georgian assembles the items into twenties (oc) and the teens and units can be grouped together(gtuni). All numerals end in an -i, which disappears after another vowel (a, in rva,cxra). The teens are generally built by adding the unit word plus meTi "more" to "(a)t" (10). The word for fourty means 2 times 20, the word for fifty-five means 2 times 20 plus 15. Most of these facts are included in the following preliminary Prolog program. */

```
predicates
gnum(tal,lista,lista)
gtuni(tal,lista,lista)
gvingt(tal,lista,lista)
gcent(tal,lista,lista)
gmil(tal,lista,lista)
gsay(tal)
gcheck(lista)
getra(lista)
egtra(lista)
gbtra(lista)
bgtra(lista)
add(lista,lista)
end(lista,lista)
ghu(lista,lista)
gtu(lista,lista)
```

```
clauses
gnum(A,B,B2)  :- gtuni(A,B,B1),end(B1,B2).
gnum(A,B,B2)  :- gvingt(A,B,B1),end(B1,B2).
gnum(D,B,B4)  :- gvingt(A,B,B1),add(B1,B2),gtuni(C,B2,B3),end(B3,B4),
   D=A+C.
gnum(A,B,B2)  :- gcent(A,B,B1),end(B1,B2).
gnum(D,B,B3)  :- gcent(A,B,B1),gvingt(C,B1,B2),end(B2,B3),D=A+C.
gnum(D,B,B3)  :- gcent(A,B,B1),gtuni(C,B1,B2),end(B2,B3),D=A+C.
gnum(D,B,B5)  :- gcent(A,B,B1),gvingt(C,B1,B2),add(B2,B3),gtuni(E,B3,
   B4),end(B4,B5),D=A+C+E.
gnum(A,B,B2)  :- gmil(A,B,B1),end(B1,B2).
gnum(D,B,B3)  :- gmil(A,B,B1),gcent(C,B1,B2),end(B2,B3),D=A+C.
gnum(D,B,B3)  :- gmil(A,B,B1),gvingt(C,B1,B2),end(B2,B3),D=A+C.
gnum(D,B,B3)  :- gmil(A,B,B1),gtuni(C,B1,B2),end(B2,B3),D=A+C.
gnum(E,B,B4)  :- gmil(A,B,B1),gcent(C,B1,B2),gvingt(D,B2,B3),end(B3,B4)
   E=A+C+D.
gnum(E,B,B4)  :- gmil(A,B,B1),gcent(C,B1,B2),gtuni(D,B2,B3),end(B3,B4),
   E=A+C+D.
gnum(E,B,B4)  :- gmil(A,B,B1),gvingt(C,B1,B2),gtuni(D,B2,B3),end(B3,B4).
   E=A+C+D.
add([da|X],X).
end([i|X],X).
gvingt(20,[oc|X],X).
gvingt(40,[ormoc|X],X).
gvingt(60,[samoc|X],X).
gvingt(80,[otxmoc|X],X).
gtuni(1,[ert|X],X).
gtuni(2,[ori|X],X).
gtuni(3,[sam|X],X).
gtuni(4,[otx|X],X).
gtuni(5,[xut|X],X).
gtuni(6,[ekvs|X],X).
gtuni(7,[shvid|X],X).
gtuni(8,[rva|X],X).
gtuni(9,[cxra|X],X).
gtuni(10,[at|X],X).
gtuni(11,[tertmeT|X],X).
gtuni(12,[tormeT|X],X).
gtuni(13,[cameT|X],X).
gtuni(14,[totxmeT|X],X).
gtuni(15,[txutmeT|X],X).
gtuni(16,[tekvsmeT|X],X).
gtuni(17,[cCvidmeT|X],X).
gtuni(18,[tvrameT|X],X).
gtuni(19,[cxramet|X],X).
gcent(100,[at|X],X).
gcent(D,B,B2)  :- gtuni(A,B,B1),ghu(B1,B2),D=A*100.
gmil(D,B,B2)  :- gtuni(A,B,B1),gtu(B1,B2),D=A*1000.
ghu([as|X],X).
gtu([atas|X],X).
gcheck(X)  :- gnum(A,X,[]),write(A),nl.
gsay(X)  :- gnum(X,A,[]),write(A),nl.
getra(X)  :- gnum(A,X,[]),esay(A). /* transl Geo X into Eng */
egtra(X)  :- enum(A,X,[]),gsay(A). /* transl Eng X into Geo */
gbtra(X)  :- gnum(A,X,[]),bsay(A).
bgtra(X)  :- bnum(A,X,[]),gsay(A).
```

FRENCH NUMERALS

A representative set of French numerals is given below.

1 un/une	11 onze	10 dix
2 deux	12 douze	21 vingt et un
3 trois	13 treize	33 trente-trois
4 quatre	14 quatorze	44 quarante-quatre
5 cinque	15 quinze	55 cinquante-cinque
6 six	16 seize	66 soixante-six
7 sept	17 dix-sept	77 soixante-dix-sept
8 huit	18 dix-huit	88 quatre-vingt-huit
9 neuf	19 dix-neuf	99 quatre-vingt-dix-neuf
100 cent		1910 dix-neuf cent dix
1000 mille		2000 deux mille

French has some irregularities,e.g. inserting "et"(and) before un, but only in 21,31,41,51,61,71. A hyphen occurs within all numerals below one hundred. The system shows signs of the (Celtic) vigesimal counting. The following rule shows the main structure of French numerals:

```
fnum --> ..(U mille)(U cent)(ante)-(fteen)
```

It is clear that the "ante" words are of two kinds, those based on vingt (multiplied with 4 in the word for 80) and the others. Note that although the word "soixante" is not based on 20, it takes teens to denote the numerals for 70-79. Gender inflection is disregarded (e.g. "un:une"). The word for 100 ("cent") is plural ("cents") after plural numerals, but not when lower numerals follow. This has been implemented in the program where "cent" is treated separately also because "un" is deleted before "cent" (and "mille"). Most of the basic facts of French numerals are covered by the following Prolog program, where some of the irregularites stand out. */

```
predicates
fnum(tal,lista,lista)
cfnum(tal,lista,lista)
fun(tal,lista,lista)
funi(tal,lista,lista)
fteen(tal,lista,lista)
fvingt(tal,lista,lista)
etante(tal,lista,lista)
ante(tal,lista,lista)
fcent(tal,lista,lista)
fcents(tal,lista,lista)
funmil(tal,lista,lista)
fmil(tal,lista,lista)
ftus(lista,lista)
et(lista,lista)
fsay(tal)
cfsay(tal)
fcheck(lista)
fetra(lista)
eftra(lista)
fbtra(lista)
bftra(lista)
fgtra(lista)
gftra(lista)
clauses
fnum(A,B,B1)  :- funi(A,B,B1).
fnum(A,B,B1)  :- fteen(A,B,B1).
fnum(A,B,B1)  :- ante(A,B,B1).
fnum(A,B,B1)  :- fvingt(A,B,B1).
fnum(D,B,B3)  :- etante(A,B,B1),et(B1,B2),fun(C,B2,B3),D=A+C.
fnum(71,B,B3)  :- fvingt(60,B,B1),et(B1,B2),fteen(11,B2,B3).
fnum(D,B,B3)  :- ante(A,B,B1),hy(B1,B2),funi(C,B2,B3),D=A+C.
fnum(D,B,B3)  :- fvingt(A,B,B1),hy(B1,B2),funi(C,B2,B3),D=A+C.
fnum(D,B,B3)  :- fvingt(A,B,B1),hy(B1,B2),fteen(C,B2,B3),D=A+C.
fnum(A,B,B1)  :- fcent(A,B,B1).
fnum(C,B,B2)  :- fcent(A,B,B1),funi(D,B1,B2),C=A+D.
fnum(C,B,B2)  :- fcent(A,B,B1),fteen(D,B1,B2),C=A+D.
fnum(C,B,B2)  :- fcent(A,B,B1),fvingt(D,B1,B2),C=A+D.
fnum(C,B,B2)  :- fcent(A,B,B1),ante(D,B1,B2),C=A+D.
fnum(C,B,B4)  :- fcent(A,B,B1),etante(D,B1,B2),et(B2,B3),fun(E,B3,B4),
   C=A+D+E.
fnum(C,B,B4)  :- fcent(A,B,B1),ante(D,B1,B2),hy(B2,B3),funi(E,B3,B4),
   C=A+D+E.
fnum(C,B,B4)  :- fcent(A,B,B1),fvingt(D,B1,B2),hy(B2,B3),funi(E,B3,B4),
   C=A+D+E.
fnum(C,B,B4)  :- fcent(A,B,B1),fvingt(D,B1,B2),hy(B2,B3),fteen(E,B3,B4)
   C=A+D+E.
fnum(D,B,B2)  :- funi(A,B,B1),fcents(C,B1,B2),D=A*C.
fnum(D,B,B3)  :- funi(A,B,B1),fcent(C,B1,B2),funi(E,B2,B3),D=A*C+E.
fnum(D,B,B3)  :- funi(A,B,B1),fcent(C,B1,B2),fteen(E,B2,B3),D=A*C+E.
fnum(D,B,B3)  :- funi(A,B,B1),fcent(C,B1,B2),ante(E,B2,B3),D=A*C+E.
fnum(D,B,B3)  :- funi(A,B,B1),fcent(C,B1,B2),fvingt(E,B2,B3),D=A*C+E.
fnum(D,B,B5)  :- funi(A,B,B1),fcent(C,B1,B2),ante(E,B2,B3),hy(B3,B4),
   funi(F,B4,B5),D=A*C+E+F.
fnum(D,B,B5)  :- funi(A,B,B1),fcent(C,B1,B2),etante(E,B2,B3),et(B3,B4),
   fun(F,B4,B5),D=A*C+E+F.
fnum(D,B,B5)  :- funi(A,B,B1),fcent(C,B1,B2),fvingt(E,B2,B3),hy(B3,B4),
   funi(F,B4,B5),D=A*C+E+F.
fnum(D,B,B5)  :- funi(A,B,B1),fcent(C,B1,B2),fvingt(E,B2,B3),hy(B3,B4),
   fteen(F,B4,B5),D=A*C+E+F.
```

```
cfnum(A,B,B1)  :-  fnum(A,B,B1).
cfnum(A,B,B1)  :-  funmil(A,B,B1).
cfnum(D,B,B2)  :-  funmil(A,B,B1),fnum(C,B1,B2),D=A+C.
cfnum(A,B,B1)  :-  fmil(A,B,B1).
cfnum(D,B,B2)  :-  fmil(A,B,B1),fnum(C,B1,B2),D=A+C.
hy([_|X],X).
et([et|X],X).
fun(1,[un|X],X).
funi(1,[un|X],X).
funi(2,[deux|X],X).
funi(3,[trois|X],X).
funi(4,[quatre|X],X).
funi(5,[cinque|X],X).
funi(6,[six|X],X).
funi(7,[sept|X],X).
funi(8,[huity|X],X).
funi(9,[neuf|X],X).
fteen(10,[dix|X],X).
fteen(11,[onze|X],X).
fteen(12,[douce|X],X).
fteen(13,[treize|X],X).
fteen(14,[quatorze|X],X).
fteen(15,[quinze|X],X).
fteen(16,[seize|X],X).
fteen(17,[dix_sept|X],X).
fteen(18,[dix_huit|X],X).
fteen(19,[dix_neuf|X],X).
ante(20,[vingt|X],X).
ante(30,[trente|X],X).
ante(40,[quarante|X],X).
ante(50,[cinquante|X],X).
etante(20,[vingt|X],X).
etante(30,[trente|X],X).
etante(40,[quarante|X],X).
etante(50,[cinquante|X],X).
etante(60,[soixante|X],X).
fvingt(60,[soixante|X],X).
fvingt(80,[quatre_vingt|X],X).
fcent(100,[cent|X],X).
fcents(100,[cents|X],X).
ftus([mille|X],X).
fmil(C,B,B2) :- fnum(A,B,B1),ftus(B1,B2),C=A*1000.
funmil(1000,B,B1) :- ftus(B,B1).
fsay(X) :- cfnum(X,Y,[]),write(Y),nl.
cfsay(X) :- cfnum(X,Y,[]),cosme(Y).
fcheck(Y) :- cfnum(X,Y,[]),write(X),nl.
fetra(X)  :- cfnum(A,X,[]),esay(A),nl.
eftra(X)  :- enum(A,X,[]),fsay(A),nl.
fbtra(X)  :- cfnum(A,X,[]),bsay(A),nl.
bftra(X)  :- bnum(A,X,[]),fsay(A),nl.
fgtra(X)  :- cfnum(A,X,[]),gsay(A),nl.
gftra(X)  :- gnum(A,X,[]),fsay(A),nl.
```

JAPANESE NUMERALS

The following is a representative set of Japanese numerals

1 ichi	11 ju_ichi	10 ju
2 ni	12 ju_ni	22 ni_ju_ni
3 san	13 ju_san	33 san_ju_san
4 yon (shi)	14 ju_yon	44 yon_ju_yon
5 go	15 ju_go	55 go_ju_go
6 roku	16 ju_roku	66 roku_ju_roku
7 nana (shichi)	17 ju_nana	77 nana_ju_nana
8 hachi	18 ju_hachi	88 hachi_ju_hachi
9 ku (kyu)	19 ju_ku	99 ku_ju_ku
100 hyaku	150 hyaku_go_ju	
200 ni_hyaku		
300 san_hyaku		
1000 sen		
2000 ni_sen		
10 000 ichi_man		

The system is fairly regular if we disregard phonological variation and some other complications, which imply e.g. that san_hyaku is rendered by san_byaku, the word for 600 is roppyaku, that for 800, happyaku, 900, kyuhyaku, 3000, sanzen, 8000, hassen, 9000, kyusen. We also disregard the problem of classifiers in Japanese (and Chinese). We suggest the following preliminary program */

```
predicates
jnum(tal,lista,lista)
juni(tal,lista,lista)
jten(lista,lista)
jhu(lista,lista)
jcent(tal,lista,lista)
jdec(tal,lista,lista)
jteen(tal,lista,lista)
jcheck(lista)
jsay(tal)
jbtra(lista)
bjtra(lista)
ejtra(lista)
jetra(lista)
fjtra(lista)
jftra(lista)
jgtra(lista)
gjtra(lista)
clauses
jnum(A,B,B1) :- juni(A,B,B1).
jnum(A,B,B1) :- jteen(A,B,B1).
jnum(A,B,B1) :- jdec(A,B,B1).
jnum(D,B,B2) :- jdec(A,B,B1),juni(C,B1,B2),D=A+C.
jnum(A,B,B1) :- jcent(A,B,B1).
jnum(D,B,B2) :- jcent(A,B,B1),juni(C,B1,B2),D=A+C.
jnum(D,B,B2) :- jcent(A,B,B1),jdec(C,B1,B2),D=A+C.
jnum(D,B,B3) :- jcent(A,B,B1),jdec(C,B1,B2),juni(E,B2,B3),D=A+C+E.
jteen(D,B,B2) :- jten(B,B1),juni(C,B1,B2),D=10+C.
jdec(D,B,B2) :- juni(A,B,B1),jten(B1,B2),D=A*10.
jcent(D,B,B2) :- juni(A,B,B1),jhu(B1,B2),D=A*100.
juni(1,[ta|X],X).
juni(2,[ku|X],X).
juni(3,[san|X],X).
juni(4,[yon|X],X).
juni(5,[go|X],X).
juni(6,[roku|X],X).
juni(7,[nana|X],X).
juni(8,[hachi|X],X).
juni(9,[ku|X],X).
jten([ju|X],X).
jhu([hyaku|X],X).
```

```
jcheck(X) :- jnum(A,X,[]),write(A),nl.
jsay(X) :- jnum(X,Y,[]),write(Y),nl.
jetra(X) :- jnum(A,X,[]),esay(A). /* transl Jap X into Eng */
ejtra(X) :- enum(A,X,[]),jsay(A).
jbtra(X) :- jnum(A,X,[]),bsay(A).
bjtra(X) :- bnum(A,X,[]),jsay(A).
jgtra(X) :- jnum(A,X,[]),gsay(A).
gjtra(X) :- gnum(A,X,[]),jsay(A).
fjtra(X) :- cfnum(A,X,[]),jsay(A).
jftra(X) :- jnum(A,X,[]),fsay(A).
```

CONCLUSION: TYPOLOGICAL FEATURES OF NUMERALS

Studies of numerals and computer programs which understand or translate numerals show that numeral systems are variants of certain themes (parameters). The lexical material (the morphemes) varies, of course, but in addition we can see variation in the following features.

1. Deletion of 1 (uni-deletion) before group words, e.g. before hundred, optionally; mostly obligatorily before low group words.
2. Insertion of conjuction (and-,et-,&-insertion), between high and low group words.
3. Order between low group constituents, e.g. units before tens.
4. Group word values, typically 10, 20 (as in French, Danish, Welsh, Yoruba, Mixtec), but also 5 (as in Ainu, Bantu, Khmer).
5. Group word value relations, e.g. powers of 10, multiples of 5 or 20. The series of group words is often irregular and broken, and e.g. a word for 10 000 and higher may be lacking.
6. Combination of morphemes into words or word groups and hyphenation. In languages with old orthography such as French the rules for the occurrence of hyphens are very rigid.

These features show what is generally to be taken into account when designing a computer program understanding or translating between languages. The programs above may be used as modules in a more comprehensive translating system. Other languages can easily be added. Prolog programs can reflect the linguistic intutions about numeral systems well, but they are certainly not the fastest programs which can be implemented for handling numerals. The programs presented need cosmetic improvement, but they should be good illustrations of one Prolog approach to numerals.

REFERENCES

Corstius, H.B. (ed), Grammars for number names. Foundations of Language series Vol 7, Dordrecht (1968:Reidel)

Hurford, J.R. The linguistic theory of numerals, Cambridge (1975:Cambridge University Press)

Menninger, K. Zahlwort und ziffer. Ein Kulturgeschichte der Zahl, Goettingen (1958:Vandenhoeck & Ruprecht)

Sigurd, B. From mathematical into linguistic representations - and vice versa. Papers from the institute of linguistics, Stockholm (PILUS),15 (1972)

" " Round numbers. Language in Society,1988 (to appear)

REFERENT GRAMMAR (RG) IN COMPUTER COMPREHENSION, GENERATION AND TRANSLATION OF TEXT (SWETRA)

Bengt Sigurd

ABSTRACT AND INTRODUCTION

Theories of text and discourse have long assumed the existence of semantic objects called discourse referents, "things being talked about and referred to by pronouns and definite noun phrases"(cf. e.g. Karttunen (1976), Sigurd (1982), Sidner (1983)). Such referents are also often alluded to in sentence grammars, above all in theories of pronominalization and control - often in terms of antecedents - but the place of the referents in grammar is not clear, and referents are not to be found in textbook generative rules. The type of grammar to be described in this paper, named Referent Grammar (RG), presumes that referents are of crucial importance and places them directly into the syntactico-semantic representation of a sentence. The referents are introduced as variables in the noun phrase rules, carried into the sentence representation and identified on the basis of the syntactic patterns and markers of the language. The referents of sentence grammar can also be related to the referents of discourse, if rules for the text grammaticality of definite noun phrases and pronouns are included. We call that version of referent grammar Text Referent Grammar (TRG). The referents are assumed to have psychological reality.

Referent grammars have so far been written and implemented in Prolog for fragments of Swedish, English, French and Georgian (in decreasing order of detail; Georgian has been included in order to test referent grammar on a typologically very different language). The grammars have been used in analysing and generating sentences and text and generating answers to questions.

The grammars have also been used in a small-scale experiment translating between Swedish, English, French and Georgian (SWETRA).

The programs have been developed and run on a VAX 11/730 using a Prolog version, which the Department of Linguistics has obtained by courtesy of the Department of Cognitive Science, Sussex University (Sussex POPLOG Prolog, version 10,1985).

Referent grammar has been inspired by the potentials of the Definite Clause Grammar (DCG) formalism supported by many Prolog programs. It has also been influenced by ideas put forward in papers by Robin Cooper, e.g. Cooper(1985) and by comments from colleagues.

SOME REFERENT GRAMMAR RULES OF ENGLISH

A simple English declarative intransitive sentence is described by the following rule written in DCG.

```
sent(d,_,s(subj(X),pred(Y))) --> nps(X),vip(Y),[.],{agr(X,Y)}.
```

The predicate "sent" starts the built-in parser, "nps" is short for subject noun phrase, "vip" is short for intransitive verb phrase (not specified here) and the condition "agr" is short for agreement (not specified here). The first argument of the basic predicate "sent" specifies the mode of the sentence, typically ranging over d(eclarative), q(uestion), i(mperative). The second argument (_ in this case) will be a focused constituent in sentences where there is one - as in questions and topicalized sentences. Further "sent" arguments characterizing the sentence type can be included. It seems natural e.g. to have a further argument to distinguish aorist sentences in Georgian, which have a very different setup of cases. The representation s(subj(X),pred(Y)) is the expression which can be stored as a fact in the data base or used in the translation system. It can be shown as a tree in programs supporting this facility.

Subject and object noun phrases must be distinguished even in English (textbooks give an oversimplified view of the np problem), as pronouns have different forms in the two cases. Reflexive noun phrases must be distinguished in referent grammar, as they have a a very special referent meaning, namely that the referent of the subject is referred to in this constituent too. A referent is introduced into an indefinite noun phrase by a rule of the following type:

```
np(np(R1,nom(A,indef))) --> indart(X),n(Y),{fm(Y,A),gensym(rf,R1)}.
```

A phrase, such as "a girl" will be parsed, since "a" is an indefinite article (indart), specified later in the program (we disregard the difference between "a" and "an" here) and because "girl" is a noun (n). The predicate fm(Y,A) gives A as the translation (form-meaning relation) of Y. Translations of all words are to be given later in the program. The meaning of the noun is thus inserted directly into the representation, using a kind of "Semantic English" as a common semantic language for all languages - English is no doubt the most wide-spread language.The variable R1 will be given a successive number by the function "gensym", which will deliver rf1,rf2,rf3, etc. allowing us to count the number of referents introduced into the text. If we do not want to keep track of the number of referents, we may leave out "gensym". Prolog will then give successive numbers for the variables R1 in different noun phrases for each sentence. It is a characteristic feature of referent grammar that a referent is assumed to be one of the constituents of the np, namely the sister of the name of it, constituted by the predicate nom(inal) with its arguments, i.e. the meaning of the words in the noun phrase (derived by fm).

A sentence representation of "A boy ran" could then be: s(subj(np(rf17,nom(boy,sg,indef))),pred(v(run,past))), where the figure 17 indicates that this is the 17th referent introduced. In a transitive sentence, e.g. "A boy hit a girl", the rules would give a sentence representation where the referents of the boy and the girl would be represented by different numbers. A reflexive sentence, such as "The boy hit himself" would, however, be treated differently, and the subject and the object referents would get the same number.

The relative clause is typically marked by a relative pronoun or another relative marker in natural languages. The basic meaning of the relative marker is that a referent in the relative clause is identical to a referent in the matrix clause (the antecedent). Referent grammar can adopt this traditional idea, by inserting a copy of the "antecedent referent" into the relative clause by the np rules. One such rule is the following:

```
np(np(R1,nom(A,indef,rs(subj(np(R1,nom(A,indef)),pred(W)))))) -->
         indart(X),n(Y),rel(Z),vip(W)..,{fm(Y,A))}...
```

It states that the subject inside the relative clause (rs) is the same as the head of the np ("the antecedent"). This rule would handle cases such as "A boy,who ran,..". Similarly referents may be inserted into minor clauses, such as infinitival complements of the verb "promise". The identity of the referents in reflexive sentences, in complex noun phrases with relative clauses and verb phrases with complements can be secured directly by the phrase structure rules in referent grammar. It is also possible to scan the representations later in the derivation and identify referents according to the configurational patterns (indirect referent identification). Dahl(1986) includes some ideas of indirect referent identification.

REFERENT GRAMMAR IN TEXT COMPREHENSION

Definite and indefinite noun phrases are often treated by the same rule in generative grammars. There is, however, a basic difference between them, which appears clearly when one tries to relate the referents of sentence grammar to the referents of text. In a text grammar a definite NP should not be accepted (text grammatical) unless the referent can be identified. This can be done in the Prolog implementation by searching the data base for a previous mention of the referent. When trying to parse "the boy" the grammar looks for a fact in its data base (stored previously) where a "boy" is mentioned. If such a referent is found, the same referent number is assigned to the referent of "the boy". A referent grammar can thus assign the same referent number to the referent of "a boy", the referent of "the boy", the subject of the relative clause and the object represented by the reflexive pronoun "himself" in the following text:

"A boy(rf1) ran. The boy(rf1), who(rf1) ran, hit himself(rf1)."

Pronouns also occur with referents, which should be identified by looking in the text stored as semantic representations in the data base. Clearly a pronoun has to look for a referent identified by a noun with the proper gender features, as is well known.

A text referent grammar should be able to identify referents as indicated by numbers in the following short text:

"A boy(rf1) had a dog(rf2). He(rf1) liked a girl(rf3). He(rf1) gave it(rf2) her(rf3)."

The identification of referents with referents in the previous text or in the verbal and non-verbal context is a most complicated matter, however, which might, in fact, be out of reach for computers, as will be illustrated by the following little case study.

A REFERENT GRAMMATICAL ANALYSIS OF THE OPENING PARAGRAPH OF GUNNLAUG ORMSTUNGAS SAGA

An English translation of the first sentences of the famous Icelandic saga about Gunnlaug Ormstunga is given below. The numbers after noun phrases are the numbers given to the referents in the order they are introduced.

> Once upon a time there was a man(rf1), who(rf1) was called Torsten. His(rf1) father(rf2) was Egil(rf3), who(rf3) was son of Skalla-grim(rf4). His(rf4) father(rf5) was Night-Ulf(rf6), who(rf6) emigrated from Norway(rf7). His(rf1) mother(rf8) was called Asgerd, and was daughter of Bjorn(rf9). Torsten(rf1) lived at the farm Borg(rf10) in Borgarford(rf11). He(rf1) was a rich man and a great chief.

The phrase "once upon a time there was" is a well-known saga opener and the following phrase "a man, who was called X" is also common. The phrase "once upon a time there was" is a stylistic marker with equivalents in other languages, e.g. French "il-y-avait" Swedish "det var en gång", but it is empty from a semantic point of view. The sentence could have started by "there was a man", but the standard phrase has to be included in a grammar which is to understand sagas. The meaning of the first sentence is that a man called Torsten existed once and a referent grammar can give the following expression.

```
s(subj(r1,nom(man,indef,rs(subj(r1,nom(man,indef)),pred(v(name,past)),
       pf(Torsten)))),pred(v(exist,past)),tadv(once))
```

Note that the subject in the relative clause has the same referent as its head(rf1), given automatically by the grammar rules as discussed above. The analysis presumes that there are grammatical rules which identify "Once upon a time there was" as one chunk semantically represented as pred(v(exist,past)),tadv(once). The noun phrase rules will identify "the man" and gensym will assign the number rf1 to this first referent. We note that "was" has the meaning "existed" here; there are other meanings of "be" as will be seen below and as is well-known. The phrase "was called" was rendered by "name,past" and we note that this predicate appears in different syntactic shapes in different languages, reflexive in French "qui s'appellait", active verb in Swedish "som hette", or passive construction "som kallades". The current referent grammar takes the name Torsten to be a predicative (predicate filler, abbreviated pf).

The second sentence talks about Torstens father, Egil and his father Skalla-grim. Egil Skalla-grimsson is a well-known Icelandic hero. The sentence is "His(rf1) father(rf2) was Egil(rf3), who(rf3) was son of Skalla-grim(rf4). The rules should give the following representation.

```
s(subj(np(rf2,nom(father,def,gen(rf1,nom(pro,ma))))),pred(v(ide,past)),
     pf(np(rf3,nom(egil,rs(subj(np(rf3,nom(egil))),pred(v(ide,past)),
     pf(n(son),prep(of)),np(rf4,nom(skalla-grim)))))))
```

The representation of the expression "son of" is preliminary. The choice must be made considering how this meaning is expressed in the languages of the world. The important thing is the chain of referents in the representation. The representation above presumes certain rules of genitive noun phrases, where several referents are involved. It is assumed that the head of a genitive noun phrase is definite, as is suggested by the inacceptability of its occurence in a presentation construction such as "Once upon a time there was his father." The genitive noun phrase seems to introduce the head noun by indicating its relation to the determining (inflected) noun. The rule in the current grammar gives the head a new referent number. All the complexities of genitive noun phrases will not be outlined here. The cases in the text include a pronoun ("his"). We assume that Torsten's father and Egil are taken to be different referents in this sentence; the purpose of the sentence is the identification of Torsten's father with Egil, who is a well-known person. While "his" in this sentence refers to Torsten, the next "his" refers to Skalla-grim. In pronunciation the last "his father" should rather be pronounced with stress on "his", but this cannot be rendered in writing. The referent of "his" can only be found by consulting the data base. We refrain from a detailed analysis of the next sentences which use very much the same vocabulary and syntax. The identification of referents poses many problems (cf. also Johnson & Klein,1986).

The referents discussed above are nominal, but there are others at least the size of a sentence. A typical case is given by such evaluative sentences as "That was fine", occurring after and clearly referring to a previous sentence, e.g. in the following sequence: "Sweden beat Britain. That was surprising." One may express this as follows in order to make the reference clear: "(The fact) that Sweden beat Britain was surprising" or "It was surprising that Sweden beat Britain." If each sentence is given a sentence number s1,s2, etc. we may keep track of these referents separately.

Referent number										Text
1	2	3	4	5	6	7	8	9	10	
+										Once upon a time there was a man(rf1),
+										who(rf1) was called Torsten.
+	+	+								His(rf1) father(rf2) was Egil(rf3),
		+	+							who(rf3) was the son of Skalla-grim(rf4).
			+	+	+					His(rf4) father(rf5) was Night-Ulf(rf6),
					+	+				who(rf6) emigrated from from Norway(rf7).
+							+			His(rf1) mother(rf8) was called Asgerd,
								+		and was the daughter of Bjorn(rf9).
+									+	Torsten(rf1) lived at the farm Borg(rf10).

Fig.1. Referent dynamic diagram of the first sentences of Gunnlaug Ormstungas saga. Each + marks a reference to that referent.

One way of presenting the referent dynamics in the opening paragraph is given in fig.1. It shows which referents are focused in each sentence and how the author returns to the hero Torsten after the presentation of his relatives. The text presents some problems of referent identification, which are, in fact, hard for both humans and computers. How do we know that "his" in "his mother" refers to Torsten and not to Night-Ulf, Skalla-grim or Egil? The reason seems to be the previous phrase "his father", which creates expectations of a parallel phrase "his mother". Note that there is another phrase "his father", but it seems probable that the first phrase "his father" gets its parallel phrase first. (A kind of push-down store strategy). The diagram also lists the typical predicates used in this opening.

COMMENTS ON THE SWEDISH MODULE OF THE PROGRAM

The figure (below) shows a fragment of a Prolog grammar which can analyse and generate some Swedish core sentences. Such a grammar has also been used in the translation between Swedish and English, French and Georgian. Similar grammars have been written for these languages. The first rules show the syntactic structures hinted at by the English translations with the rules. Note that Swedish has inverted word order whenever the subject does not occur first in the sentence. The grammar ("sent") has a special slot for mode, where the value d occurs when the sentence declarative, and q, when the sentence is a question. Declarative sentences and yes/no-questions are assumed to have identical representations, the only difference being in the mode variable. The next argument of "sent" is a slot for a focused constituent. An initial adverb is assumed to be a focused constituent in Swedish and so is the finite verb in yes/no questions and the question word in wh-questions. Swedish may also focus the verb in declarative sentences using an auxiliary verb corresponding to English "do". This construction is illustrated by the syntactic rule including [gjorde].

Some later rules are also shown indicating how definite and indefinite nps with relative clauses are handled. Swedish has very complex agreement within the noun phrase. All nouns are divided into neuter and non-neuter nouns (which is a well-known stumbling-block for foreigners). Furthermore, the number and the definiteness of the noun phrase has to be distinguished. The agreement rules suggested look at one pair of words at a time, e.g. first the article and the noun, then the adjective and the noun in order to make sure that e.g. "det lilla barnet" (the little child) and "en liten flicka" (a little girl) come out right.

```
——┼ 71├ 80  (EDITING: subtrans.pl) ————————————————————————————————
SWEDISH */
sent(d,_,s(subj(X),pred(Y))) --> nps(X),vip(Y),[.].   /* x ran */
sent(d,Z,s(subj(X),pred(Y),advl(Z))) --> adv(Z),vip(Y),nps(X),[.].  /* fast ra|
sent(d,Z,s(subj(X),pred(Y),advl(Z),advl(W))) --> adv(Z),vip(Y),nps(X),adv(W),[|
sent(d,_,s(subj(X),pred(Y),advl(Z))) --> nps(X),vip(Y),adv(Z),[.].  /* ran fas|
sent(d,_,s(subj(X),pred(Y),advl(Z),advl(W))) --> nps(X),vip(Y),adv(Z),adv(W),[|
sent(q,Y,s(subj(X),pred(Y))) --> vip(Y),nps(X),[?]. /* ran x? */
sent(q,Y,s(subj(X),pred(Y),advl(Z))) --> vip(Y),nps(X),adv(Z),[?].
sent(q,X,s(subj(X),pred(Y))) --> npq(X),vip(Y),[?]. /* who ran */
sent(q,X,s(subj(X),pred(Y),advl(Z))) --> npq(X),vip(Y),adv(Z),[?].
sent(q,X,s(subj(X),pred(Y),advl(Z),advl(W))) --> npq(X),vip(Y),adv(Z),adv(W),[|
sent(d,Y,s(subj(X),pred(Y))) --> vip(Y),[gjorde],nps(X),[.]. /* ran did x */
sent(d,_,s(subj(X),pred(Y),aobj(Z))) --> nps(X),vtp(Y),npo(Z),[.]. /* x hit y |
sent(d,_,s(subj(X),pred(Y),aobj(Z),advl(W))) --> nps(X),vtp(Y),npo(Z),adv(W),[|
sent(d,_,s(subj(X),pred(Y),aobj(X))) --> nps(X),vtp(Y),npr(Z),[.]. /* x hit hi|
sent(q,Y,s(subj(X),pred(Y),aobj(Z))) --> vtp(Y),nps(X),npo(Z),[?]. /* hit x y |
sent(q,Y,s(subj(X),pred(Y),aobj(Z),advl(W))) --> vtp(Y),nps(X),npo(Z),adv(W),[|
sent(q,X,s(subj(X),pred(Y),aobj(Z))) --> npq(X),vtp(Y),npo(Z),[?]. /* who hit |
sent(d,Z,s(subj(X),pred(Y),aobj(Z))) --> npq(Z),vtp(Y),nps(X),[?]. /* whom hit|
sent(d,_,s(subj(X),pred(Y),dobj(Z),aobj(W))) --> nps(X),vttp(Y),npo(Z),npo(W),|
sent(q,_,s(subj(X),pred(Y),dobj(X),aobj(W))) --> vttp(Y),nps(X),npo(Z),npo(W),|
sent(d,_,s(subj(X),pred(Z,Y))) --> nps(X),cop(Y),pf(Z),[.],{pagr(X,Z)}. /* x i|
sent(q,Z,s(subj(X),pred(Z,Y))) --> cop(Y),nps(X),pf(Z),[?],{pagr(X,Z)}. /* is |
```

Some syntactic rules used in an implementation of referent grammar (SWETRA)

REFERENT GRAMMAR IN AUTOMATIC TRANSLATION (SWETRA)

Referent grammar is used in the Swedish translation system SWETRA. Representations with "Semantic English" are used as an intermediate level. They have to be modified in some ways when going into another language and some interlanguage rules have been worked out. A number of basic sentences can be translated. Case marking and agreement are seen as "local" problems in each language. Such markings disappear when sentences are translated into the normalized Semantic English representations making translation possible. The following are some examples of sentences which can be translated. The predicate setrans means translate from Swedish into English, sftrans means translate from Swedish into French, etc. There is also a predicate which translates from Swedish into English, into French and back to Swedish. The vocabulary is very restricted as yet and the system is only a prototype intended to indicate the potentials of referent grammar in automatic translation. (The figure shows further examples)

setrans(en liten flicka,	a little girl,
som beundrade en pojke,	who admired a boy,
som en hund bet, sprang.)	whom a dog bit, ran.
sftrans(den lilla flickan sprang.)	la petite fille courut.
sgtrans(en hund bet flickan.)	jagli gogos kbenda.
sgtrans(den flickan gav ett barn en hund.)	gogo svilfs jagli darboda.

Even sentences with pronouns in different order in the languages can be handled correctly, as witnessed by the following case.

English: He gave it her. (Grammatical in some variants of English)

French: Il le lui donnait.

Swedish: Han gav henne den.

```
?- sgtrans([en,flicka,sprang,.],X).
X = [gogo, darboda, .] ?
yes

?- sgtrans([en,hund,bet,flickan,.],X).
X = [jagli, gogos, kbenda, .] ?
yes

?- setrans([den,gick,snabbt,.],X).

X = [it, went, fast, .] ?
?- setrans([ubaaten,styrde,mot,karlskrona,.],X).

X = [the, submarine, headed, towards, the, carlscrona, .] ?
yes

?-
?- setrans([ubaaten,foeljde,en,fiskebaat,.],X).
** (1) Call : intse(s(subj(np(_1, nom(m(submarine, sg), m(def)))), pred(v(m(fo
      llow, past))), aobj(np(_2, nom(m(fishingboat, sg), m(indef))))),
      _3)?
** (1) Exit : intse(s(subj(np(_1, nom(m(submarine, sg), m(def)))), pred(v(m(fo
      llow, past))), aobj(np(_2, nom(m(fishingboat, sg), m(indef))))),
      es(subj(np(_1, nom(m(submarine, sg), m(def)))), pred(v(m(follow,
      past))), aobj(np(_2, nom(m(fishingboat, sg), m(indef))))))?

X = [the, submarine, followed, a, fishing_boat, .] ?
      _3)?

?- setrans([en,jagare,gick,fraan,karlskrona,.],X).
** (1) Call : intse(s(subj(np(_1, nom(m(destroyer, sg), m(indef)))), pred(v(m(
      go, past))), advl(adv(_2, nom(m(from), np(_3, nom(m(carlscrona),
      m(def))))))), _4)?
** (1) Exit : intse(s(subj(np(_1, nom(m(destroyer, sg), m(indef)))), pred(v(m(
      go, past))), advl(adv(_2, nom(m(from), np(_3, nom(m(carlscrona),
      m(def))))))), es(subj(np(_1, nom(m(destroyer, sg), m(indef)))),
      pred(v(m(go, past))), advl(adv(_2, nom(m(from), np(_3, nom(m(carlscro
      na), m(def))))))))?

X = [a, destroyer, went, from, the, carlscrona, .] ?
yes
?-
?- setrans([ubaaten,doek,naer,jagaren,anlaende,.],X).
** (1) Call : intse(s(subj(np(_1, nom(m(submarine, sg), m(def)))), pred(v(m(di
      ve, past))), advl(adv(m(when), _2, rs(subj(np(_3, nom(m(destroyer,
      sg), m(def)))), pred(v(m(arrive, past))), advl(adv(_2)))))), _4)?
** (1) Exit : intse(s(subj(np(_1, nom(m(submarine, sg), m(def)))), pred(v(m(di
      ve, past))), advl(adv(m(when), _2, rs(subj(np(_3, nom(m(destroyer,
      sg), m(def)))), pred(v(m(arrive, past))), advl(adv(_2)))))), es(subj(
      np(_1, nom(m(submarine, sg), m(def)))), pred(v(m(dive, past))),
      advl(adv(m(when), _2, rs(subj(np(_3, nom(m(destroyer, sg), m(def)))),
      pred(v(m(arrive, past))), advl(adv(_2)))))))?

X = [the, submarine, dived, when, the, destroyer, arrived, .] ?
yes
```

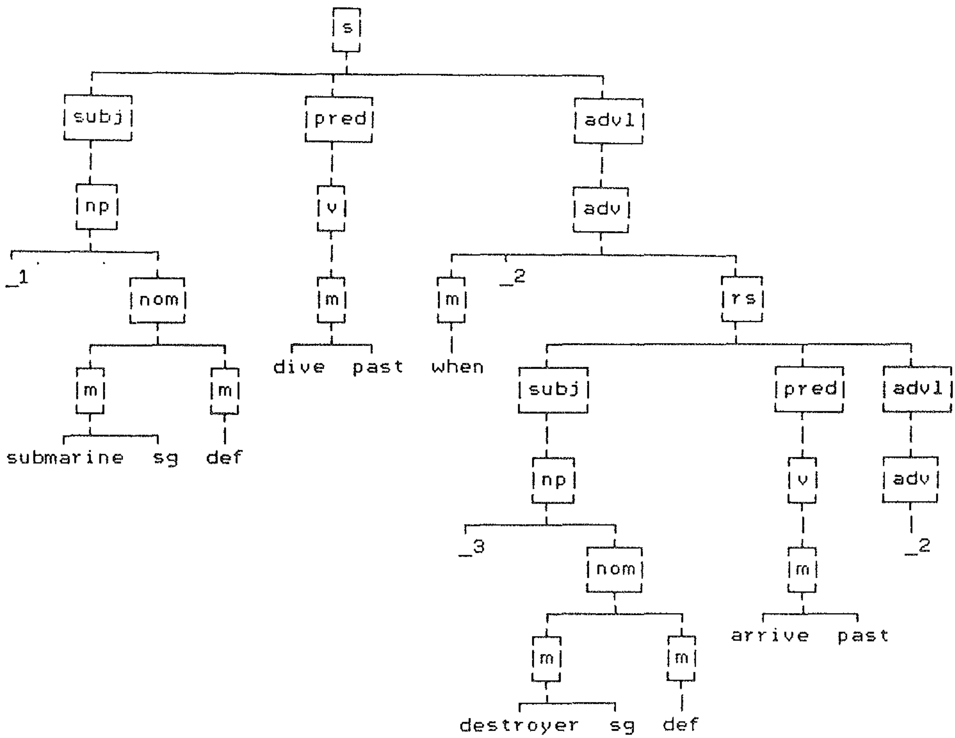
Sample translation by SWETRA. Note the intermediate representations with the referent variables.

The choice of the pronouns, which depends on the features (number, gender, as e.g. French le/la,les) poses a knotty problem, however. A dog (chien) would require a following "le", while a flower (fleur) would require "la", both corresponding only to English "it". The dog (hund) requires "den" in Swedish, a child (barn) would require "det" as a pronoun. One way to handle this situation is to use text referent grammar, keeping track of the referents in the text and the information about them. If no such solution can be found the goals of automatic translation must be more modest.

Experiments in automatic translation using referent grammar are carried out under the project title SWETRA (Swedish Computer Translation Reseach), supported by The Swedish Research Council for the Humanities and the Social Sciences. The experiments follow different lines. One line will emphasize the multilanguage translation potentials developing modules for a number of languages. This will allow translation of certain types of information e.g. weather forecasts between that set of languages. Such a system translating some 20 000 words between the EG languages is in fact the main goal of EUROTRA.

Another line followed by SWETRA is pairwise translation, e.g. between English and Georgian, between Japanese and Swedish in order to examine the problems of translation between typologically dissimilar languages. Special interest has been taken in automatic translation between Georgian and English as the tense-aspect system of Georgian requires special arrangements. An illustrative extract from a Georgian module is shown.

Translation between English and Swedish in a certain field - marine events - has been studied in some detail and a print-out from a session translating text dealing with a submarine and a destroyer outside Carlscrona is shown.



TREE DIAGRAM OF THE SYNTACTIC-SEMANTIC REFERENT REPRESENTATION OF "UBÅTEN DÖK NÄR JAGAREN ANLÄNDE", WHICH IS ALSO USED AS INTERMEDIATE REPRESENTATION IN TRANSLATION

HIGH RESOLUTION REFERENT GRAMMAR (HRG) IN ANALYSIS AND GENERATION OF TEXT

High resolution referent grammar (HRG) is a kind of grammar where the content of a sentence is dissolved into smaller pieces, each including referents which show the connections of the pieces. It is an extension of Referent Grammar (RG) presented above. Using a simple example, the sentence "Bill hit Tom" is dissolved into: somebody did something, what he did was to hit, he who hit was Bill, he who was hit was Tom. Using a formalism suitable for Prolog implementation the meaning of the sentence is represented as:

```
s(n(P,N1,N2),d(hit,P),d(bill,N1),d(tom,N2)). (Bill hit Tom)
```

Expressed in words the formula says: It is a fact that there is a predicate (P) with two arguments (N1 and N2) combined into a nexus (n);cf. the use of this term by Jespersen. The predicate (P) is denoted (d) by "hit", the first argument (N1) is denoted by Bill and the second argument by Tom. These four facts are gathered inside a sentential fact (s). The nexus fact is the first. It states that there is a finite predicate, which is a 2-place predicate. The next facts specify the predicate and the two nouns. The letters P and N are used as variable names.

It is convenient to talk about these facts (nexus, denotation) as "factoms", making up "factecules", using the parallelism of atoms and molecules. The terms of the factoms can then be termed "refons", as they refer and are the structural equivalents of electrons and protons.

The factoms can be combined in different ways corresponding to different sentence types. The following table shows the set-up of factoms for some basic types of sentences.

No of factoms, set-up		Sample sentences (Eng, Swe)
2	n(P,N),d(A,P)	There was running, Det sprangs
3	n(P,N),d(A,P),d(B,N)	Tom ran,Tom sprang (intrans)
3	n(P,N1,N2),d(A,P),d(B,N2)	Tom was hit,Tom slogs (ag-less)
4	n(P,N1,N2),d(A,P),d(B,N1),d(C,N2)	Bill hit Tom,Bill slog Tom (tr)
4	n(P,N1,N2),d(A,P),d(C,N2),d(B,N1)	Tom was hit by Bill (pass)
4	n(P,N1,N2,N3),d(A,P),d(B,N2),d(C,N3)	Bill was given a book (ag-less)

The basic order in the sentence factecules above is: the nexus(predication, finite) combination first, the character of the predicate next and the characters of the arguments in order next. Such formulas offer the order of the factoms in the factecules as a way of differentiating between active and passive, as is shown above.

Note that what characterizes an impersonal construction is the lack of information about the N, which seems intuitively correct. Similarly the agentless passive lacks information about the character of the agent, i.e. the first N (N1) in the nexus factom.

EXPLAINING COMPREHENSION BY HRG

Comprehension is seen as a parsing process resulting in a factecule which is stored in the memory (data base). The individual factoms are stored in a second step. It is an open question whether further implications from these first facts, e.g. factoms equivalent to relative nps should be stored during the comprehension process or derived when needed.

This means that when the sentence "A boy hit a dog" or Swedish "En pojke slog en hund" is to be understood, a factecule such as s(n(p1,n1,n2),d(hit,p1),d(boy,n1),d(dog,n2)) is stored (asserted) along with the individual factoms: n(p1,n1,n2),d(hit,p1) etc. New constellations, e.g. s(n(p1,n1,n2),d(hit,p1)) equivalent to "There was hitting", and s(n(p1,n1,n2),d(hit,p1),d(dog,n2)) equivalent to "A dog was hit" are also true (and available as potential sentences). The existence of the factom d(boy,n1) in the data base makes it possible to identify a later mention of the boy (the boy,he,the scoundrel) with n1. The factom equivalent to the np "the boy who was hit": d(s(n(p1,n1,n2),d(hit,p1),d(dog,n2)),n2) may also be made available.

The representation uses lower case letters with numbers instead of upper case letters in accordance with Prolog conventions. The values p1,p2 etc and n1,n2 etc are assigned during the parsing procedure and stored with the factoms. This means that it is possible to follow the build-up of the database on the basis of the sequence of sentences. It is also possible to know which is the preceding verb, noun etc., which is of interest in interpreting anaphora and proforms.
(See printout of sample session)

PRO-FORMS IN HRG

Referent grammar is specifically designed to handle referential expressions. There are a number of referential expressions in natural languages. We will only discuss some here. The "It" in "It is true that Bill ran" refers to the following clause "that Bill ran", but in a sentence such as "It/That is true" the "it" refers to a preceding sentence. In the little text "Bill ran. So did Tom" the "so" refers to the action (running) mentioned just before, the denotation of the previous predicate. This information has thus to be extracted in order for the latter sentence to be understood and the factecule and factoms corresponding to "Tom ran" stored. Natural languages utilize reference extensively and in an economic way.

The most frequent cases are definite noun phrases and pronouns, whose interpretation has been discussed extensively for a long period and cause computerized text comprehension systems great problems. (See References)

High resolution grammar stores individual factoms as the comprehension proceeds and gives each referent an individual number. A referent established previously may be referred to later, thus being used as a discourse referent. After the sentence "A boy ran" is processed the factoms n(p1,n1),d(ran,p1),d(boy,n1) are stored in the data base (simulating human memory). If the next sentence is "The boy hit a dog", the referent of "the boy" is identifed with the same referent number (n1) found to be used with a boy in a factom. The dog will be given a new number (n2), however, and stored in a factom as d(dog,n2). A following noun phrase "the dog" can then be taken to refer to this n2. More technically, the factom d(dog,X) is looked for and n2 is found to be a possible value of X.

Similarly, a pronoun is taken to refer to a current referent mentioned in a factom, if the requirements of number, gender etc are met. This is handled by implicational rules in the Prolog implementation. If there is a factom d(boy,n9), then d(he,n9) is available. Pronouns are often ambiguous, and we have not as yet a solution to this problem, which also pesters human communication. It is not reasonable to expect that computers should be able to solve problems of ambiguous reference, which human beings hardly can solve, in spite of all their experience and information processing capacity.

It is also a well-known fact that discourse referents are referred to by hyponymic expressions. A hyponymic expression such as "the animal" can thus be used about a dog ("Bill hit the dog. Tom hit the animal too"). This problem can be solved by including hyponymic implicational rules. If there is a factom d(dog,n7) then d(animal,n9) is available, assuming that animal is given as a hyponym of dog.

Verbal hyponyms are handled in a similar way. If d(ran,p9) is true then d(moved,p9) is available, as moved is a hyponym of ran. It is also known that various pejorative and affectionate expressions may serve as referent expressions ("Bill hit the dog. Tom hit the beast too"). This could also be handled by implicational rules, but it seems dangerous to state generally that if somebody is called a boy he can also be called a scoundrel. As has been observed by Merle Horne(1986) coreferential terms such as "the animal", "the beast" and pro-forms of nouns and verbs do not take focal stress (are destressed).

One of the rationales for High Resolution Referent Grammar is to offer units which are suitable for handling reference expressions of natural languages. The factecules corresponding to sentences are clearly needed for the referring "it,this" in cases such as "Philosophy is important. It/This is true." But there seems to be need for a still bigger unit in cases where the whole content of the preceding text is referred to. Such a discourse referent may be called "circumscriptive" (a term used by Bonnie Webber in a lecture) or "summative" as it sums up what has been said. A typical case is when a person says e.g. "This is all I have to say."

One may carry the resolution suggested above further in order to satisfy further needs of referential processes. One may consider whether tense should not be factored out from the main predicate and "Bill ran" be represented by n(t1,p1,n1),d(run,p1),d(past,t1), d(bill,n1), where we have a tense variable (t) in the first factom. One reason for doing this could be that one may (possibly) say: "Bill ran. So does Tom". In that case only the component "run", not the tense, is to be copied into the factecule of the second sentence.

The sentence type "So did Bill" (Swedish "Det gjorde Bill ocksaa") seems to be a transitive sentence in both languages, where "so/det" is the preposed (topicalized) object. Swedish uses "gjorde" (goera) as an equivalent to English "did" (do) in this case, but not generally. The "so/det" may refer to the denotation of the verb (as shown) but also to the verb including an object (the whole VP) as in: "Tom hit a dog. So did Bill", where the last sentence means that Bill hit a dog, not only that he hit. This suggests that factoms corresponding to the whole predicate phrase should be copied.

GENERIC SENTENCES

Generic sentences such as: "(All) boys are animals","A boy is (always) an animal" (Swedish "(All) pojkar aer djur", "En pojke aer (alltid) ett djur", and even "Pojken aer ett djur" (The boy is an animal) require special interpretation, as is well-known. There is an extensive linguistic and philosophical literature on the problem. The standard philosophical interpretation is: "If something is a boy, then it is an animal", but this seems rather to be a secondary implication. We suggest that generic sentences be rendered by special factoms (g), and that a later implicational rule allow the program to infer facts as suggested by the standard philosophical interpretation. The implication is straightforward given our analysis in factoms (and in Prolog). If there is a factom d(bird,n9), and a generic factom g(animal,bird), then d(animal,n9) is true, or more generally (in Prolog) d(X,N) :- d(M,N),g(X,M). There is, of course, much more to be said about this problem.

COMMENTS ON AN EXPERIMENTAL PROLOG IMPLEMENTATION OF HRG

The following program only includes a few basic types of sentences and a few vocabulary items. It is to be noted that the built-in predicate "gensym" is called to give successive referent numbers to verbs and nouns. This is handled on the sentence level for verbs which get p1,p2,p3 etc successively and in the np submodule for nouns. The numbers generated in the lower np module have to be carried into the nexus factom by the rules of the "sent" predicate.

The basic predicate "sent" processes some Swedish sentences, but equivalent English sample sentences are given as comments. This program understands sentences such as: [en_pojke,sprang] (A boy ran), [en_pojke,slog,en_hund] (A boy hit a dog) and can store the corresponding factecules and factoms if asked to by using "lagra (X)" (store). The first line in the program reads:" if something (X) is a noun phrase (np) and the following item (Y) a verb, then build the structure (tree) indicated to the left (as an argument with s as the root)". In that structure the values of the variables are given by gensym, except for referring expressions, where the value is taken from a factom stored previously.

The meaning is dissolved into factoms as mentioned above. The word recognition procedures offer the semantic representations of the words which are given in "Semantic English", a representation which is also used in the intermediate representations of the machine translation system (SWETRA) mentioned above. Only simple noun and verb phrases are covered and there are only a few lexical items in this demo program.

Some basic sentence types are included in the program. The predicate with the prefix g (gsent) is used in generation, when gensym is not to be used. There are interactive commands in the program, which generate all possible sentences within the grammar (say) and all possible sentences which are true (truesay).

The program includes intransitive, transitive (active and passive, with and without agent), impersonal, generic and coordinated sentences. Some sentences with verb pro (so,det) are also covered. The factecule which is the first argument of sent shows how the different sentences are analysed.

If we write sent(T,[en,pojke,sprang],[]), we get T= s(n(p1,n1), d(ran,p1),d(boy,n1)), which may be spelled out as "there is an event involving P and N, where P is "ran" and N is "boy". The factoms may be combined in several ways to make up different factecules, as mentioned. If asked to generate true sentences by the command truesay the program will search the data base for primary and implicated factecules and combine primary and implicated factoms into all possible factecules. It will e.g. generate "Det sprangs" (there was running) as the factecule s(n(p1,n1),d(ran,p1)) can be constructed.

The factoms from a sentence such as ([en_pojke,slog,en_hund] ("A boy hit a dog") may be combined in various ways as indicated by the syntactic rules. Beside the original sentence it may evoke:[en_hund, slog,s,av,en_pojke] ("a dog was hit by a boy"), [en_hund,slog,s] ("a dog was hit"), [det,slog,s] ("there was hitting"). Coordinated sentences, e.g. [en_pojke,sprang,och,hoppade] ("A boy ran and jumped") are also available, if the same boy is known to have jumped. If asked to generate all possible sentences on the basis of the factoms stored one may get unexpected coordinated sentences, such as [en_pojke,sprang,och, sprang] ("A boy ran and ran"), on the basis of the previous sentence [en_pojke,sprang]. The reduplicated verb, however, carries a special meaning! Definite nouns, pronouns, and some hyponyms and pejoratives can be used if the conditions are met (as discussed above).

There are various interactive predicates. Storing is handled by the procedure "lagra(X)". The whole factecule is later split into factoms, which are also stored (asserted). The predicate "truesay"gives all sentences which can be rendered by combinations of the factoms learnt or derived by the system. The predicate "answer(X)" gives answers (or rather confirmations) to declarative sentences, which have been given or which can be derived as true by the implicational rules. Note that the predicate "gensym(a,X)" has to be loaded by "library(gensym)". The predicate "library(useful)" should also be loaded in order to make "append" and some other commands available. Debugging and extension of the program is still going on, but the program can analyze some short sample texts nicely (as illustrated).

```
yes

?- library(gensym).
yes

?- load(nex).
yes

?- lagra([en_pojke,slog,en_hund]).
n(p4, n7, n8)d(hit, p4)d(m(boy, sg), n7)d(m(dog, sg), n8)yes

?- lagra([den,sprang]).
n(p7, n8)d(ran, p7)d(rpro, n8)yes

?- lagra([det,gjorde,pojken,ocksaa]).
n(p10, n7)d(ran, p10)d(m(boy, sg), n7)yes

?- lagra([djuret,lekte]).
n(p13, n8)d(played, p13)d(m(animal, sg), n8)yes

?- lagra([det,gjorde,han,ocksaa]).
n(p16, n7)d(played, p16)d(mpro, n7)yes

?-
```

Analysis and comprehension of a short Swedish text corresponding to English: A boy hit a dog. It ran. So did the boy. The animal played. So did he.

Note that the referents n7 and n8 introduced in the first sentence are identified thoughout the text.

PROGRAM
Initialize by lagra([en_pojke,sprang]). This version generates np at module level and inserts the gensym generated noun variable (N) at the sentence level. The verb variable (P) is generated at sentence level. Relativization is handled by deriving d-facts from s-facts.
Various interactive commands available:lagra,answer,say,truesay,facts.

```
sent(s(n(P1,N1),d(Y,P1),d(Z,N1))) --> np(X),v(Y),{gensym(p,P1),
   asserta(lastv(P1)),d(Z,N1)=X}.
/* current(N1)? */
gsent(s(n(P1,N1),d(Y,P1),d(X,N1))) --> gnp(X),v(Y). /* for generation */

/* following takes verb value from previous sentence */
sent(s(n(Z,N1),d(Y,Z),d(W,N1))) --> [det],[gjorde],np(X),[ocksaa],
   {d(W,N1)=X,lastv(Z),d(Y,Z)}.
gsent(s(n(P1,N1),d(Y,P1),d(W,N1))) --> [det],[gjorde],gnp(X),[ocksaa],
   {lastv(Z),d(Y,Z),d(W,N1)=X}.

/* sentence with unstressed hyponymic coreferent with  verb, np additive
    ocksaa  */
sent(s(n(P1,N1),d(Y,P1),d(X,N1))) --> np(X),v(Y),[ocksaa],{d(X,N1),
    hyp(Z,Y),d(Z,P2),gensym(p,P1)}.
/* sentence with stressed (nonhypo) verb, verb additive ocksaa */
sent(s(n(P1,N1),d(Y,P1),d(X,N1))) --> np(X),v(Y),[ocksaa],{gensym(p,P1),
    asserta(lastv(P1))}.
/* following coordinates predicates */
sent(s(n(P1,N1),d(Y,P1),n(P2,N1),d(Z,P2),d(W,N1))) --> np(X),v(Y),
   [och],v(Z),{d(W,N1)=X,gensym(p,P1),gensym(p,P2)}.
gsent(s(n(P1,N1),d(Y,P1),n(P2,N1),d(Z,P2),d(W,N1))) --> gnp(X),v(Y),
   [och],v(Z),{d(W,N1)=X}.

/* following establishes 2 factoms  */
sent(s(n(P1,N1),d(Y,P1))) --> [det],v(Y),[s],{gensym(p,P1),asserta
   (lastv(P1))}. /* there is dancing */
gsent(s(n(P1,N1),d(Y,P1))) --> [det],v(Y),[s].
/* following is a predicative non-generic (plural) sentence */
sent(s(n(P1,N1),d(Y,P1),d(W,N1))) --> ni(X),[aer],pf(Y),{gensym(p,P1),
   d(W,N1)=X}.
gsent(s(n(P1,N1),d(Y,P1),d(W,N1))) --> gni(X),[aer],pf(Y),{asserta
   (lastv(P1))}.
/* following are some tentative generic sentences */
sent(s(g(Y,X))) --> [alla],ni(X),v(Y).
gsent(s(g(Y,X))) --> [alla],ni(X),v(Y).
sent(s(g(Y,X))) --> [alla],ni(X),[aer],pf(Y). /* all boys are nice */
gsent(s(g(Y,X))) --> [alla],ni(X),[aer],pf(Y).
sent(s(ng(Y,X))) --> [inga],ni(X),[aer],pf(Y). /* no boys are nice */
/* following establishes 4 factoms */
sent(s(n(P1,N1,N2),d(Y,P1),d(W,N1),d(U,N2))) --> np(X),vt(Y),np(Z),
   {gensym(p,P1),asserta(lastv(P1)),d(W,N1)=X,d(U,N2)=Z}.
gsent(s(n(P1,N1,N2),d(Y,P1),d(W,N1),d(U,N2))) --> gnp(X),vt(Y),gnp(Z),
   {d(W,N1)=X,d(U,N2)=Z}.
sent(s(n(P1,N1,N2),d(Y,P1),d(W,N1),d(U,N2))) --> np(Z),vt(Y),[s_av],
   np(X),{gensym(p,P1),asserta(lastv(P1)),d(W,N1)=X,d(U,N2)=Z}.
```

```
gsent(s(n(P1,N1,N2),d(Y,P1),d(W,N1),d(U,N2))) -->gnp(Z),vt(Y),[s_av],
   gnp(X),{d(W,N1)=X,d(U,N2)=Z}.
sent(s(n(P1,N1,N2),d(Y,P1),d(U,N2))) --> np(Z),vt(Y),[s],{gensym(p,P1),
   d(U,N2)=Z}. /* pass without ag */
gsent(s(n(P1,N1,N2),d(Y,P1),d(U,N2))) --> gnp(Z),vt(Y),[s],{d(U,N2)=Z}.
sent(s(n(P1,N1,N2),d(Y,P1))) --> [det],vt(Y),[s],{gensym(p,P1)}. /*
   impersonal trans */
gsent(s(n(P1,N1,N2),d(Y,P1))) --> [det],vt(Y),[s].
np(d(X,N1)) --> [A],{isn(A),bet(A,X),gensym(n,N1)}. /* indefinite
   (new N1) */
gnp(d(X,N1)) --> [A],{isn(A),bet(A,X)}.
np(d(X,N1)) --> [A],{isnd(A),bet(A,X),d(X,N1)}. /* definite np (old N1) *
gnp(d(X,N1)) --> [A],{isnd(A),bet(A,X),d(X,N1)}.
/* relative nps */
np(d(L,N1)) --> np(X),[som],v(Y),{d(W,N1)=X,L=s(n(P1,N1),d(Y,P1),
   d(W,N1)),L}.
np(d(L,N1)) --> np(X),[som],vt(Y),np(Z),{d(W,N1)=X,L=s(n(P1,N1,N2),
   d(Y,P1),d(W,N1),Z),L}.
np(d(L,N2)) --> np(Z),[som],np(X),vt(Y),{d(W,N2)=Z,L=s(n(P1,N1,N2),
   d(Y,P1),X,d(W,N2)),L}.
/*Lexicon */
isn(ett_barn).
bet(ett_barn,m(child,sg)).
isn(ett_djur).
bet(ett_djur,m(animal,sg)).
ne(m(animal,sg)).
ne(m(child,sg)).
isn(en_hund).
bet(en_hund,m(dog,sg)).
re(m(dog,sg)).
isn(en_idiot).
bet(en_idiot,m(idiot,sg)).
ma(m(idiot,sg)).
fe(m(idiot,sg)).
isn(en_pojke).
bet(en_pojke,m(boy,sg)).
ma(m(boy,sg)).
isn(en_flicka).
bet(en_flicka,m(girl,sg)).
fe(m(girl,sg)).
/* definite. Gender based on bedeutung (in indefinite) */
isnd(pojken).
bet(pojken,m(boy,sg)).
isnd(flickan).
bet(flickan,m(girl,sg)).
isnd(barnet).
bet(barnet,m(child,sg)).
isnd(idioten).
bet(idioten,m(idiot,sg)).
isnd(den_idioten). /* used in pejorative sense */
bet(den_idioten,m(idiot,sg)).
isnd(den_boven).
bet(den_boven,m(scoundrel,sg)).
isnd(djuret).
bet(djuret,m(animal,sg)).
isnd(hunden).
bet(hunden,m(dog,sg)).
isnd(den).
bet(den,rpro).
isnd(han).
bet(han,mpro).
```

```
isnd(hon).
bet(hon,fpro).
isnd(det).
bet(det,npro).
/* pronoun identification.If eg n9 is denoted by boy it can be denoted
   by han  */
d(rpro,Y) :- re(X),d(X,Y). /* rpro is substitutable if noun is re (den) *,
d(mpro,Y) :- ma(X),d(X,Y). /* mpro is substitutable if noun is ma (han) *,
d(fpro,Y) :- fe(X),d(X,Y). /* fpro is substitutable if noun is fe (hon) *,
d(npro,Y) :- ne(X),d(X,Y). /* npro is substitutable in noun is ne (det) *,

/* pejorative and hypochoristic (affectionate) substitution rules  */
d(Y,Z) :- pej(X,Y),d(X,Z). /* if n9 is a boy, n9 is also a scoundral */
pej(m(boy,sg),m(idiot,sg)).
pej(m(boy,sg),m(scoundrel,sg)).
pej(m(dog,sg),m(scoundrel,sg)).
/* hyponym substitution rule  */
d(Y,Z) :- hyp(X,Y),d(X,Z).
hyp(m(dog,sg),m(animal,sg)).
hyp(jumped,played). /* played is assumed to be hypon ym of jump here */
ni(X) --> [A],{isni(A),bet(A,X)}.
isni(pojkar).
bet(pojkar,m(boy,pl)).
isni(hundar).
bet(hundar,m(dog,pl)).
v(X) --> [A],{isv(A),bet(A,X)}.
isv(sprang).
bet(sprang,ran).
isv(gick).
bet(gick,went).
isv(hoppade).
bet(hoppade,jumped).
isv(lekte).
bet(lekte,played).
vt(X) --> [A],{isvt(A),bet(A,X)}.
isvt(slog).
bet(slog,hit).
isvt(tog).
bet(tog,took).
isvt(gillade).
bet(gillade,liked).
pf(X) --> [A],{ispf(A),bet(A,X)}.
ispf(pigga).
bet(pigga,alert).
ispf(trevliga).
bet(trevliga,nice).
ispf(dumma).
bet(dumma,silly).
/*
```

```
STORING (Lagra[]).
stores and prints factecule and (by extore) each factom */
lagra(X) :- sent(L,X,[]),print(L),asserta(L),extore(L).
extore(L) :- s(X,Y)=L,asserta(X),asserta(Y).
extore(L) :- s(X,Y,Z)=L,asserta(X),asserta(Y),asserta(Z).
extore(L) :- s(X,Y,Z,W)=L,asserta(X),asserta(Y),asserta(Z),asserta(W).

/* derives new facts from generic information;
the quantifier alla */
d(A,X) :- g(A,m(B,pl)),d(m(B,_),X),print(alla). /*  if generally (g) B
   is A and X is B then X is A */
/*  construction of potential factecules from factoms */
s(n(A,B),d(C,A)) :- n(A,B),d(C,A),print(impersonalitr).
s(n(A,B),d(X,A),d(Y,B)) :- n(A,B),d(X,A),d(Y,B),print(itr).
s(n(A,B),d(X,A),d(C,B),d(Y,C),d(Z,B)) :- n(A,B),d(X,A),d(C,B),d(Y,C),
   d(Z,B),print(coo).
s(n(A,B,C),d(X,A)) :- n(A,B,C),d(X,A),print(imperstr).
s(n(A,B,C),d(X,A),d(Z,C)) :- n(A,B,C),d(X,A),d(Z,C),print(passtr).
s(n(A,B,C),d(X,A),d(Y,B),d(Z,C)) :- n(A,B,C),d(X,A),d(Y,B),d(Z,C),
   print(tr).
/* derives relative np facts (d(X,Y) from factecules (s(X,Y,Z))) */
d(L,N1) :- L=s(n(P1,N1),X,Y),L. /* if N1 ran, N1  ran is true */
d(L,N1) :- L=s(n(P1,N1,N2),X,Y,Z),L. /* if N1 hit N2, N1 who hit N2 is
    true */
/* Answers declarative statement by ja(yes) if true */
answer(X) :- gsent(L,X,[]),L,print(ja),nl,print(L).
/* prints sentences which are true given primary and implicated facts */
truesay :- gsent(L,X,[]),L,print(X),;.
/* prints all sentences */
say :- sent(L,X,[]),print(X),;.
/* prints possible sentences and their factecules */
fsay :- sent(L,X,[]),print(X),print(L),;.
/* prints all factoms */
facts :- F=n(P,N),F,print(F),;.
facts :- F=n(P,N,M),F,print(F),;.
facts :- F=j(X,A),F,print(F),;.
facts :- F=d(X,Y),F,print(F),;.
```

REFERENCES

Cooper,R. Meaning representation in Montague Grammar and Situation Semantics. Workshop on Theoretical Approaches to Natural Language Understanding. May,1985. Halifax, Nova Scotia

Dahl,O. The interpretation of bound pronouns. In F.Karlsson (ed), Papers from the fifth Scandinavian conference of computational linguistics, Helsinki(1986:Dept of linguist.)

Horne,M. Focal prominence and the 'phonol ogical phrase' within some recent theories. Studia Linguistica 1986:2, 101-121

Johnson,M. & I.Klein. Discourse anaphora and parsing. Proc of COLING 1986, Bonn. (1986)

Kamp,H. A theory of truth and semantic representation. In Groenendijk,J.A.G., Janssen,T.M.V. & Stokhof,M.B.J. (eds), Formal methods in the study of language, Vol 136. Amsterdam: Mathematical Centre Tracts.

Karttunen,L. Discourse Referents, in J.McCawley (ed), Syntax and Semantics, vol 7. New York (1976:Academic Press)

Sidner,C. Focusing in the comprehension of definite anaphora. In Brady,A. & R.C.Berwick (eds), Computational Models of discourse. Cambridge, Mass.(1983:MIT press)

Sigurd,B. Text representation in a text production model. In Allen,S. (ed), Text processing, Proc. of Nobel Symposium, Stockholm. (1982: Almqvist & Wiksell International)

A NOTE ON THE DURATION OF CHINESE QUESTIONS

Jan-Olof Svantesson

One of the most common ways of forming yes/no questions from statements in spoken Standard Chinese is by merely changing the sentence intonation, leaving all else intact. Questions formed in this way have the same meaning as questions formed with the final sentence particle *ma* or by repeating the verb with the negation *bu* between the repetitions. For instance, corresponding to the statement *Wāng Yī chōu xiāngyān* 'Wang Yi smokes cigarettes' there are three questions, all meaning 'Does Wang Yi smoke cigarettes?' (see e.g. Li and Thompson 1981:520 ff. for some differences in the use these types of questions):

(1) Wāng Yī chōu xiāngyān?

(2) Wāng Yī chōu xiāngyān ma?

(3) Wāng Yī chōu bu chōu xiāngyān?

Questions of type (1) differ from the corresponding statements mainly by having higher pitch, particularly over the final phrase. In languages which can form yes/no questions with intonation only, it is often the case that in addition to the pitch differences, questions have shorter duration than the corresponding statement. This is true for instance in Polish (see data given in Petecka 1985) and Cairo Arabic (unpublished measurements by Kjell Norlin). In this note I present data that show this to be the case in Standard Chinese as well.

For this purpose, a material consisting of four sentences with different tonal patterns uttered in different contexts, including neutral statement and yes/no question was used. The four sentences were:

1.	Wāng Yī chōu xiāngyān	'Wang Yi smokes cigarettes.'
2.	Sòng Yán mài niúròu	'Song Yan sells beef.'
3.	Shěn Lǐ mǎi yǔsǎn	'Shen Li buys an umbrella.'
4.	Wāng Lǐ chuān yǔyī	'Wang Li wears a raincoat.'

This material was recorded in Lund and Stockholm with four speakers of Standard Chinese. The first three speakers had grown up in Peking, and the

Table 1. Duration of statements (S) and questions (Q) in ms.

Sentence:	1		2		3		4	
	S	Q	S	Q	S	Q	S	Q
Speaker 1	890	910	830	720	910	850	970	890
	1190	950	1090	920	1190	950	1010	830
	990	780	910	830	970	900	1060	850
mean:	1023	880	943	823	1023	900	1013	857
Speaker 2	1645	1080	1110	970	1210	1070	1540	1080
	1310	1380	1330	1140	1650	1050	1400	1140
	1190	1050	1030	990	1110	1050	1300	1050
mean:	1383	1170	1157	1033	1323	1057	1413	1090
Speaker 3	1090	910	1180	760	1150	890	990	790
	1130	870	1050	760	2000	830	1110	760
	1070	880	890	770	1320	800	970	810
mean:	1097	887	1040	763	1490	840	1023	787
Speaker 4	1170	850	1190	890	1210	880	1320	930
	1230	930	1180	800	1060	840	1090	870
	1180	890	1250	930	1150	970	1300	990
mean:	1193	890	1207	873	1140	897	1237	930

fourth, who was originally from Liaoning province in the Northeast, had come to Peking in his mid-teens.

Each speaker read the entire material three times. The duration of each sentence was measured from spectrograms, and the results are given in Table 1.

As seen in Table 1, for each combination of speaker and sentence, the average duration was longer for the statement than for the question, the average duration of a question being about 78% of that of the corresponding statement. As might be expected, a three-way analysis of variance (with the factors speaker, sentence and statement/question) showed that the duration difference between statements and questions was highly significant ($p<0.0005$; $F(1,88)=73.72$).

References

Li, Charles and Sandra Thompson. 1981. *Mandarin Chinese. A functional reference grammar.* Berkeley: University of California Press.

Petecka, Janina. 1985. 'A study of question intonation in Polish'. *Working Papers* (Dept. of Linguistics, Lund University) 28, 151-73.

SOME DATA ON THE DURATION OF CHINESE STOPS AND AFFRICATES

Jan-Olof Svantesson

The purpose of this note is to report on some measurements of the duration and voice onset time (i.e. the duration of the aspiration phase) for Standard Chinese stops and affricates. Such data seem not to be readily available except in Chinese sources: Feng 1985 gives data on the total duration and the duration of the closure phase for Chinese stops and affricates, and Qi and Zhang 1982 measured the total duration of Chinese consonant sounds (excluding the closure phase for stops and affricates).

As is well known, there are two series of stops and affricates in Standard Chinese, voiceless unaspirated and aspirated (here given with IPA symbols and the standard *pinyin* transcription within <...>):

Unaspirated:	p <b>	t <d>	k <g>	ts <z>	tʂ <zh>	tɕ <j>
Aspirated:	p‘ <p>	t‘ <t>	k‘ <k>	ts‘ <c>	tʂ‘ <ch>	tɕ‘ <q>

Stops and affricates occur only syllable initially in Chinese, and for this investigation monosyllabic words were used. They were read in a carrier sentence Wǒ pǎ ___ tsì ɕiěxǎo <Wǒ bǎ ___ zì xiěhǎo> ‘I finish writing the character ___’. The word containing the investigated consonant was thus in a focussed position.

In the test syllables, the stops and affricates were followed by the vowel /a/, except for the palatals, which occur only before /i/ and /y/. The following syllables were used:

pā	<bā>	p‘āi	<pāi>
tāo	<dāo>	t‘ā	<tā>
kāo	<gāo>	k‘āi	<kāi>
tsāo	<zāo>	ts‘āo	<cāo>
tʂāo	<zhāo>	tʂ‘āo	<chāo>
tɕiā	<jiā>	tɕ‘iāo	<qiāo>

These words embedded in the carrier sentence were read twice by four speakers of Standard Chinese, three of whom had grown up in Peking; the fourth, originally from Liaoning province, had been living in Peking since his mid teens.

For each stop and affricate, the duration of the closure phase, the aspiration phase, and for the affricates, the fricative phase, were measured from oscillograms made with the ILS program package. The results are given in Table 1.

The results show that the aspiration phase of Chinese stops is relatively long (90-110 ms), as might be expected in a language where aspiration is the only phonetic property which differentiates two series of stops and affricates. This can be compared with languages such as Swedish or Arabic, which have one series of voiceless, somewhat aspirated stops, and one series of voiced stops, and where aspiration is usually analyzed as a non-distinctive property of the voiceless stops. The duration of the aspiration phase in these languages is about 30-50 ms (Löfqvist 1976; Gårding et al. 1986), much shorter than in Chinese stops.

As is usually the case, aspiration increases with the backness of the place of articulation, both for stops and affricates. In particular, 'unaspirated' [k] has a VOT of about 30 ms.

Aspiration is shorter in the affricates than in the stops, around 70 ms. Since there is a gradual transition from the fricative to the aspiration phase, it was difficult in some cases to find a clear boundary point between them. For the unaspirated affricates, no aspiration phase could be detected. It can be noted that the alveopalatal affricate [tʂ] has a relatively short fricative phase, which may be related to the fact that there is no stop with this place of articulation.

As is often noted in descriptions of Chinese (e.g. Kratochvíl 1968:26), the unaspirated stops and affricates may become voiced in some speech styles. (Because of the syllable structure of Chinese, syllable initial consonants are always preceded and followed by voiced sounds, unless they are utterance initial.) The exact circumstances under which this takes place are not known, but it seems that only stops and affricates in unstressed syllables may become voiced. Voicing was not observed in my test words, which were in a focussed position in the sentence frame, but especially for speaker 2, /p/ in the object marker /pǎ/, and /ts/ in the word /tsì/ 'character' were often voiced. The oscillogram of *Wǒ pǎ pā tsì ɕiěxǎo* shown in Figure 1 illustrates this. In the sentence frame, *tsì* forms a noun phrase (e.g. *pā tsì* 'the character *pā*') together with the preceding word. The object marker *pǎ* is syntactically in construction with this noun phrase, but is phonetically cliticised to the subject *wǒ* 'I'.

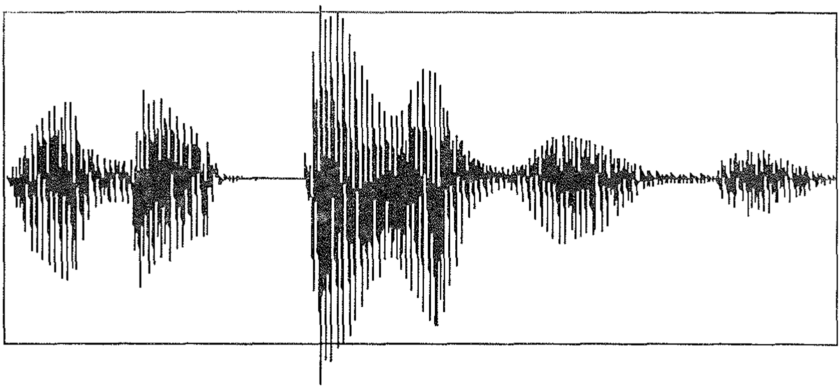


Figure 1. Oscillogram of wǒ pǎ pā tsì ɕiěxǎo (Speaker 2).

References

Féng Lóng. 1985. 'Běijīnghuà yǔliú zhōng shēng yùn diào de shícháng'. In Lín Tāo and Wáng Lǐjiā (eds.) *Běijīng yǔyīn shíyàn lù*, 131-195. Běijīng: Běijīng dàxué chūbǎnshè.

Gårding, Eva, Mona Lindau, Kjell Norlin and Jan-Olof Svantesson. 1986. 'Final report: phonetic analyses of some non-European languages (LUCLA)'. *Working Papers* (Dept. of Linguistics, Lund University) 29, 115-38.

Kratochvíl, Paul. 1968. *The Chinese language today*. London: Hutchinson.

Löfqvist, Anders. 1976. 'Closure duration and aspiration for Swedish stops'. *Working Papers* (Dept. of Linguistics, Lund University) 13, 1-39.

Qí Shìqián and Zhāng Jiālù. 1982. 'Hànyǔ pǔtōnghuà fǔyīn yīncháng fēnxì (A study of duration of Chinese consonants)'. *Shēngxué xuébào* (Acta acustica) 7:1, 8-13.

Table 1. Duration of Standard Chinese stops and affricates (ms).

Speaker:		1		2		3		4		mean
[p]	Cl.	115	127	120	141	96	111	140	125	122
	Asp.	8	6	7	4	13	10	7	8	8
	Tot.	123	133	127	145	109	121	147	133	130
[t]	Cl.	123	127	146	110	126	129	166	127	132
	Asp.	8	7	13	16	10	10	11	12	11
	Tot.	131	134	159	126	136	139	177	139	143
[k]	Cl.	92	126	126	104	86	114	106	94	106
	Asp.	26	21	30	36	29	30	43	36	31
	Tot.	118	147	156	140	115	144	149	130	137
[pʻ]	Cl.	74	94	179	90	104	101	120	95	107
	Asp.	112	112	80	80	61	105	102	101	94
	Tot.	186	206	259	170	165	206	222	196	201
[tʻ]	Cl.	88	105	101	84	89	94	119	101	98
	Asp.	124	123	96	82	93	108	97	93	102
	Tot.	212	228	197	166	182	202	216	194	200
[kʻ]	Cl.	74	74	107	68	80	104	96	96	87
	Asp	129	102	86	77	91	119	129	127	108
	Tot.	203	176	193	145	171	223	225	223	195
[ts]	Cl.	106	87	137	110	86	76	104	110	102
	Fric.	36	59	37	40	47	52	77	58	51
	Asp.	-	-	-	-	-	-	-	-	-
	Tot.	142	146	174	150	133	128	181	168	153
[tʂ]	Cl.	90	101	123	85	123	88	89	81	98
	Fric.	36	43	20	30	25	29	41	46	34
	Asp.	-	-	-	-	-	-	-	-	-
	Tot.	126	144	143	115	148	117	130	127	131
[tɕ]	Cl.	71	72	103	72	53	56	90	86	75
	Fric.	62	64	58	62	67	75	78	78	68
	Asp	-	-	-	-	-	-	-	-	-
	Tot.	133	136	161	134	120	131	168	164	143
[tsʻ]	Cl.	71	76	135	83	85	89	77	89	88
	Fric.	58	65	48	37	36	44	56	39	48
	Asp.	71	49	52	54	75	58	97	69	66
	Tot.	200	190	235	174	196	191	230	197	202
[tʂʻ]	Cl.	77	66	104	103	96	88	69	69	84
	Fric.	49	50	49	44	29	41	51	61	47
	Asp.	86	79	56	59	68	82	88	71	74
	Tot.	212	195	209	206	193	211	208	201	204
[tɕʻ]	Cl.	56	59	96	68	67	72	69	79	71
	Fric.	52	56	55	59	54	46	51	56	54
	Asp.	82	98	58	53	53	57	99	94	74
	Tot.	190	213	209	180	174	175	219	229	199

DISCOURSE INTONATION IN SWEDISH
An Exploratory Study in Communication Games

Madeleine Wulffson

When recently asked if intonation was being included in a current project involving natural speech, a Swedish linguist replied, "Well, no. It is important, but it's too difficult." A phonetician, on the other hand, lamented that traditional phonetics have no established methodology for dealing with pragmatic meaning. So the linguist, who deals daily with meaning in communication, needs also systematic access to intonation. And the phonetician, who deals daily with the intricacies of intonation, needs also systematic access to the meaning it carries. The question is, how to go about it?

Intonation plays a significant pragmatic role in at least three major ways. Firstly in the transmission of information and the negotiation of meaning. Secondly in the creation and maintenance of social and psychological relationships, and thirdly for signalling discourse structure. The question has been to find a new approach and develop a new model for Swedish which encompasses and incorporates these discourse functions.

As it happens, such a model has in fact been developed in England which appears extremely promising in this regard for Swedish as well. The present study is based on the model for Discourse Intonation developed by David Brazil at Birmingham, and its applications to the Swedish language. For details and a comprehensive exposition of the model, the reader is referred in particular to Brazil 1985, "The Communicative Value of Intonation in English".

At the beginning of these pilot studies the question was raised, "Is this model relevant to languages other than English, (Swedish in particular) and if so, in what ways?". Approximately one year's study and research has shown that the basic system can indeed be applied to Swedish, and also that it will provide an illuminating tool for discourse analysts and phoneticians alike. A descriptive overview of the model will be given in the first section of this paper, with particular attention being paid to its Swedish applications.

The second section will demonstrate ways in which the system can be applied to a particular type of discourse. The present material consists of a series of 18 recordings (3-15 min. each) of dialog activities which have been dubbed "Communication Games", or "K-Lekar" in Swedish. These games represent spontaneous natural conversations between two people working cooperatively towards a mutually recognized goal. Occasionally material from other languages have also been taken into consideration when required for elucidation of specific points. The relatively large volume of material is a key aspect of the methodology of this investigation, which aims to find the common discourse denominators of Swedish intonational phenomena, and discover how meaning emerges through them. The model is of a nature such that this type of approach is rendered both manageable and rewarding.

Section 1.

The Discourse Model of Intonation: Some Introductory Notes

The conceptual categories represented in the Birmingham Model of Discourse Intonation are of an abstract extra-linguistic nature, inherent in human communication. Regardless of language, human beings continually assess the state of convergence between themselves and their interlocutors. Based on that ongoing assessment of common and separate worlds, they make decisions as to how to classify those worlds while speaking, on informational as well as social and psychological levels. They distinguish between 'more important' and 'less important' bits of information. They create and occupy various roles in their personal relationships as well as in society. Accordingly, they also exercise various degrees of control over the ongoing discourse, simultaneously giving it structure and cohesion. They classify things as being in line with or contrary to expectations. They also set up expectations about the type of response they consider to be agreeable or appropriate at any given moment. (Whether they receive that response or not is an entirely different matter, up to someone else to decide.) They may project truth as falsehood, and falsehood as truth. In short, they classify and manipulate the world around them in a dynamic and nuanced way. According to Brazil, intonation plays a pivotal role in the realization of such functions, in a framework of meaning increments which are conceptually separable from the type of meaning conveyed through grammar and lexis.

We are dealing here with a moment-by moment, real time model where speakers classify existence along lines which are accessible to themselves and their hearers in the unique here and now setting of the discourse. The number of composite elements is small, and the formula is simple, but the model effectively represents far-reaching and complex phenomena. Abstract formal oppositions acquire concrete local significance in given contexts.

The Discourse Model postulates a finite set of meaningful linguistic oppositions which can be singled out on a perceptual auditory level from the more or less constantly varying stream of speech. The meaning components here described represent the result of a speaker having made an either/or choice. The independent variables are functional in nature. For example, "if there is a 'falling pitch', it is not the fall itself which is of interest but rather the function of the language item that carries it." (Brazil 1985) A formal functional choice may result in a wide range of phonetic realizations.

The basic factors which contribute to the realization of the functional oppositions within each tone unit are: PROMINENCE, TONE, KEY, and TERMINATION. Further within the domain of these systems are ORIENTATION ('direct/oblique'), and DOMINANCE or DISCOURSE CONTROL. Study of the work of Brazil is recommended for the reader who would like to gain a deeper understanding of the model. In the meantime, the following will serve as a general guideline:

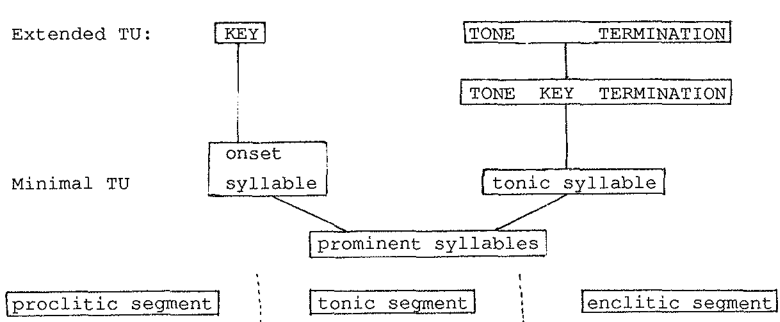
PROMINENCE refers to 'a selection from sets available at successive places along the time dimension.' 'An incidence of prominence fixes the domain of the other variables of tone, key, and termination.' (Brazil 1985) A syllable or stretch of speech may be assigned prominence for the purpose of sense or intonation selection. For example, if one Swede asks another 'Vilket kort spelade du?' (Which card did you play?), and the other replies 'HJÄRTerDAM!' (the queen of hearts), it represents a selection from an existential set of 52 - the deck of cards. If the question had been: 'Vilken DAM spelade du?'- (Which queen did you play?) with the answer, 'HJÄRTerdam', then there are only 4 choices in the existential set of hjärter, spader, ruter, and klöver. On the other hand, if the question had still been 'Vilken dam spelade du?', but the answer had been, for example, 'HJÄRTerDAM' there would seem to be no motivation for 'DAM' to be prominent. But let's say, for example, that the speaker wished to concentrate on the card game instead of answering questions, he might convey this with low termination on the word 'DAM', in order to be left alone! So prominence may be assigned for the purpose of making a choice within any of the other intonational systems of tone, key or termination. Actually, in Central Swedish the question here is complicated by the instance of accent 2 or grave accent, on the syllable 'HJÄRT' , which can affect or modify the distribution of meaningful discourse elements to a considerable degree. The discourse implications of the tone accents in Swedish will be more closely explored and mapped out in a later publication.

TONE refers to basic pitch movement types, each of which carries a distinct abstract meaning increment. The PROCLAIMING TONE, of which there are two versions, the SIMPLE proclaiming and the DOMINANT proclaiming (p ↘ p+ ↗↘) stands for the elements in the discourse which represent a change in the status quo of speaker-hearer understanding. The REFERRING TONE, on the other hand, also with a SIMPLE and a DOMINANT version (r ↘↗ r+ ↗) effectively represents the areas of convergence, or reification of the status quo between speaker and hearer, either on informational or social levels, or both. The dominant version reinforces the basic meaning of a tone and/or affects control of the discourse.

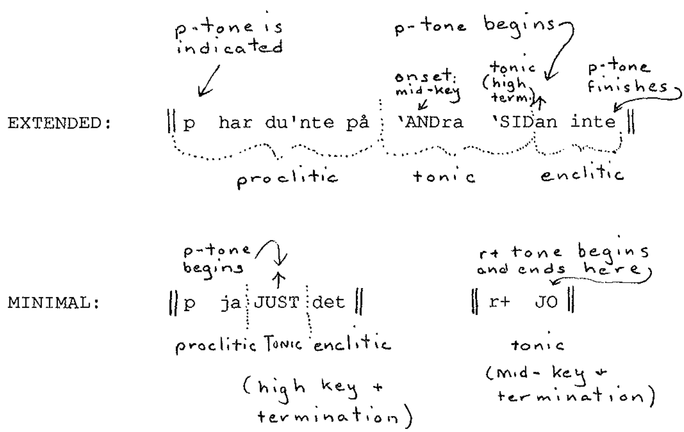
The FIFTH TONE is LEVEL (o →), and remains outside of the interactive proclaiming/referring dichotomy. ORIENTATION refers to the mode of concentration of the speaker. DIRECT orientation refers to the discourse situation in which speaker/hearer interaction is in focus (P/R), whereas OBLIQUE orientation (O/P) functions where the language itself or linguistic organization is in focus.

KEY and TERMINATION deal with the communicative value of relative pitch levels, HIGH, MID or LOW. Within their domain are relationships of CONTRASTIVENESS, ADDITIVENESS, and EQUIVALENCE , as well as the interactive areas of projected and actual responses, ADJUDICATING (high termination and key) and CONCURRING (mid termination and key) or no projected expectations (low termination and key). DISCOURSE STRUCTURING and SEQUENCING are also achieved through key and termination.

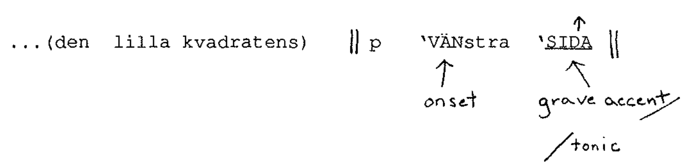
The place of operation for these four sets of speaker options is the TONE UNIT, which can be said to be the building block of verbal communication. According to Brazil, the speaker 'plans' the tone unit, and the hearer 'decodes' it as a whole. A tone unit (TU) in direct orientation consists of ONE (minimal TU) or TWO (extended TU) prominent syllables, one of which is TONIC (= carries a major movement in pitch, or constitutes the beginning of a pitch movement which extends over the syllables that follow.) The diagram below will show the disposition of tone, key, and termination in the TU. Key and termination are determined by the level of pitch in relation to preceding and succeeding prominent syllables. The tonic syllable is the only obligatory portion of a TU. A pause always defines a TU boundary, but a TU is not always defined by a pause. The model differs substantially from other models in this crucial point. It is the instance of a set of meaningful functional choices, and their internal organization, rather than external boundaries which determine the tone unit. But here for the moment are the technical criteria for tone unit organization:



EXAMPLE TONE UNITS



As noted before, grave accent can affect the distribution of prominence and the discourse tones. If an A2 word is 'two-topped', and the second 'top' carries tonic prominence, but there is a word or syllable in the onset position of the tone unit other than the grave accent syllable, then the tone unit will have 3 prominent syllables in the transcription, as opposed to the usual maximum of two in a direct orientation tone unit in English.
For example:



The significance of these systems and subsystems will be made clearer by exemplification in other sections of this article. Only selected elements which are pertinent to subsequent discussions have been mentioned here. A schematic overview of the Discourse Model will be found in the appendix, and will be useful as a point of reference during subsequent reading. In order for the reader to understand the notation system which will be used henceforth, an overview of the transcription conventions is provided here, with special attention paid to the modifications required for Swedish. Please keep in mind that the categories are perceptual and functional in nature and that the representations are stylized and schematic.

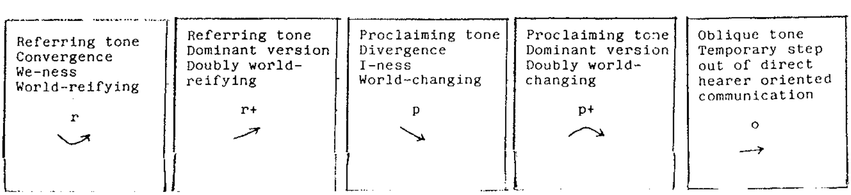
Transcription Conventions for Swedish

1. Tone unit boundaries:
2. Prominent syllables in capital letters with the tonic underlined: ÖVre kaTET
3. Key and termination (relative pitch factors involved at every TU. Key is associated with the onset syllable, and termination with the tonic in an extended in an extended TU. Together on the tonic in a minimal TU.

   a. MID-KEY/TERMINATION are not specially marked.

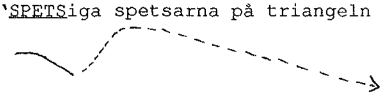
   b. HIGH or LOW key-termination:

   In the case of accent 2 (A2) words in Swedish, where the pitch switch from mid to high key or termination takes place on the syllable following grave accent (A2), this is indicated by an arrow placed above that second non-prominent syllable.
4. Grave accent or accent 2 (A2): (`)
5. P = either /p/ or /p+/ proclaiming. R = either /r/ or /r+/ referring.



Here are some examples of tone contour configurations found in the C-Game data. (Please keep in mind that these lines represent only very general, perceptually significant directions of pitch which have a bearing on the discourse context of meaning. In no way do they intend to represent all the myriad variations found in instrumental representations of Fo phenomena.)

Proclaiming Tone:
(C-Game no. 5): ‖ p 'SPETSiga spetsarna på triangeln ‖

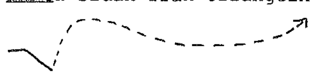


(no.5): ‖ p 'VÄNta nu ‖ ‖ p 'SOCKerLÅdor ‖

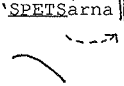
p+ Tone:
(C-Game no.7 and no. 11): ‖ p+ Å (!) ‖ p+ avanCERat (!) ‖ / ‖ P+ TITta! ‖

Referring Tone:
(C-Game no.17): ‖ r BORT från dig ‖ (no.12): ‖ r att komma PÅ nånting ‖

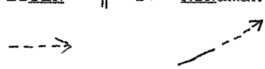
(no.5): ‖ r den 'LÅNGa sidan från triangeln ‖



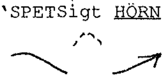
(no.17): ‖ r med dem spetsiga 'SPETSarna ‖



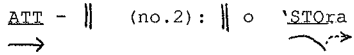
r+ Tone (C-Game no.7): ‖ o en LITEN- ‖ r+ KINaman ‖



(no.17): ‖ r+ det är ju ett 'SPETSigt HÖRN ‖



o Tone
(C-Game no.7): ‖ o för ATT - ‖ (no.2): ‖ o 'STOra - ‖

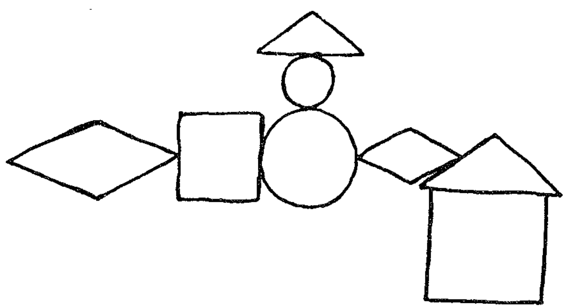


Section 2

The Discourse Model in Action: Communication Games

The Communication Games were chosen from a larger collection of recorded and transcribed material because they afforded the opportunity to make very specific comparisons of highly spontaneous, while at the same time relatively content-controlled material, in a context where the investigator was able to keep track of speaker intent, and also be present during recording sessions.

Pairs of volunteers were each given a set of paper shapes; a large and a small version of a circle, a triangle, a square, and a diamond. As they sat separated by a screen, one of them would lay a pattern on the table which might look like this: (C-Game no.2)



He or she ('A' or the 'instructor') would then give instructions to the other person ('B' or the 'constructor') as to how to achieve the same pattern with his or her shapes. Both participants were invited to discuss the progress of the operation together throughout, and to express comments on the final result. They then exchanged roles and repeated the exercise, which was in turn also recorded.

In the C-Games, as elsewhere, it happens that grammar and lexis represent only part of what is happening to the state of speaker/hearer understanding at any given time. Intonation stands for a very substantial share of the information therein contained.

In the following sections some of the aspects of discourse intonation in the C-Games will be taken up. Only very selective bits have been chosen for consideration.

Discourse Boundary Markers in the C-Games:

We have noted that intonation is important in signalling discourse structure, and it is this function in one of its aspects that will be examined here. Sinclair & Brazil (Teacher Talk 1982) isolate the 'framing' move of the teacher in the classroom. This is characterised as consisting of a small closed class of words or phrases , such as 'Right', 'Now', 'OK', 'Good', etc., that are used by teachers as if to say "At this point I finish one part of what I want to say, and open up another part of the discourse." These boundary markers enclose sections of discourse called 'transactions'. Teachers (Sinclair/Brazil 1982) "use frames after silences of any length, or to mark digressions, supplement any points, at the ends as well as the beginnings of transactions." They are followed by a 'measured pause', and notably carry a high key falling tone. The words do not carry the meanings they are otherwise associated with; for example 'right' is not presented as the opposite of 'left' or 'wrong', but rather as an indication that a structurally significant point has been reached. The high key aspect, signalling a contrastive, 'this, not that' polarity, must therefore attach to the whole discourse unit. The proclaiming tone is used to 'tell' the class of the existence of a boundary. Although there is evidence, already in my data, (which includes, among other things, C-Games in several different languages) that the basic framework of this model remains intact across many languages, still it seems more than reasonable to postulate a sort of intonational convention, which may be culturally, language-wise, or even individually specific. The 'frame' or rather 'frame-type device' in Swedish is probably one of these conventions. In the case of the C-Games, a very distinct framing device is easily recognizable, basically consisting of 2 variations:

|| r+ (↗) ↑<u>MHM</u> || or: || r+ (↗) ↑<u>HAH</u> ||

Since in this case lexis is not involved at all, (as opposed to the 'dummy' lexis of the frame in the English speaking classroom, the meaning component is even more obviously carried by intonational means alone. This was by far the most common distinct structural marker to be found in the C-Games. Concentration here will be on its use at the beginning of the games, although it also occurred not infrequently at significant structural points throughout the games. Particularly between negotiated placements - ie, where both parties had discussed the location of a circle or triangle, and had agreed that it was correct.

In the first phases of the C-Games, in 10 out of 18 cases cases there were markers resembling frames. Of the frames following the 'Swedish' trend (r+ with high key), interestingly enough, only 4 out of 11 came from the instructor, while the rest were marked by the other speaker (the constructor) who at face value might be assumed, at least temporarily, to occupy a non-dominant position. No way. Constructors were quick to establish their 'equal' status in the exercise. Or, as in C-Game no.1, both parties contributed equally to the initial structuring of the discourse: (A = instructor, B = constructor)

A: ‖ r+ j'↑<u>HA</u> ‖ p NU är jag <u>KLAR</u>↓ Jonny ‖ B: r+ ↑<u>MHM</u> ‖

In common with the English classroom frame, this Swedish framing device is almost invariably marked with high key in the C-Games. There is only one frame-like marker which occurs on mid-key in the data, but it comes from, if the reader will pardon the pun, a very 'low key' person.

The major difference is the (r+) rising tone, as opposed to the falling (p) tone. A possible logic behind this could be: I 'refer' (R-tone) to the fact (which you probably have noticed,) that something has been completed and/or something is about to begin, but I am the one taking the initiative, and therefore choose the dominant version of this tone in order to mark this boundary. (Entirely subconsciously, of course.)

In order to set this convention into a broader context, I would like to step briefly outside my Swedish data, and comment on 3 C-Games recorded in American English. A recent perusal of those transcriptions yielded all of 6 examples of a framing device. All featuring high key and falling tone, and all using the lexis of the closed set (Right, Now, etc.,) as described by Brazil & Sinclair (1982) in connection to classroom discourse. A step even further afield is the distinct 'framing'device found in a C-Game in Finnish:

‖ p no ↑<u>NIIN</u> ‖

(high key) following the English style intonational pattern. Actually in one way it can be said that the C-Games bear a strong similarity to the classroom situation, in that there is also here a specific body of concrete information to be transmitted in a more or less structured way within a given space of time.

There is, in fact, one Swedish instance (at the beginning of C-Game no.16,) of a classic English-style framing device:

A: ‖ p (↘) o↑<u>KAY</u> ‖

So this is not in any way to be excluded from the Swedish repertoire.

But from general acute listening (one risks becoming an eavesdropper in this line of work,) plus the preliminary evidence from the data, I believe it can safely be maintained that the non-lexical, high key, rising (r+) tone 'framing' or structural boundary marker device is a significantly prevalent intonational phenomenon in Swedish. A kind of intonational convention which differs partly from the English norm in its realization, but covers the same function and remains fully within the categories of the Discourse model. A 'frame' - in a different 'frame of mind'.

The Role of Key in the C-Games

The term 'key' is used to describe the relative pitch level of the onset syllable in a tone unit. If the tone unit is minimal, then key is associated with the tonic syllable. It is not in any way a question of absolute pitch values . The workings of key can be described in the following manner: if the speaker finds himself at a certain pitch level at the tonic of a tone unit (termination), he has 3 existential choices, in terms of pitch. He can go higher, stay on the same level, or else drop his voice to a lower level. There is no other way he can go. So it is with key. To an onset pitch level which is perceptibly higher (beyond the lesser 'jump' that prominence entails), a contrastive value is attached. For example, in C-Game no.7, 'B' thinks that he is going to be the first constructor:

B: (...jag tror att jag ska göra ungefär som jag gjorde)

‖ r+(↗) `FÖRRa ↑GÅNGen ‖ ... (But 'A' has 'beat him to it'):

A: ‖ p(↘) j'har ↑GJORT en gubbe ‖ p(↘) j'har REDan ↑GJORT en gubbe

'B' says, 'I think I'll do something like I did last time' - but 'A' has a different idea. He says, 'I've already DONE it' (I've already MADE a little man.) 'DONE' is projected as being in compelling contrast with 'will do', through the use of high key. And compelling it is, as A 'wins' the position of instructor.

Mid key in the C-Games is slightly more difficult to exemplify clearly. Not because it does not occur very often - just the opposite. Mid-key tone units are imbued with the quality of additiveness. They are projected as being simply additional bits of information. As in C-Game no.15:

A (och lägger den-) ‖ p till HÖGer om `FYRkanten ‖

B ‖ r MHM ‖ r+ vilken AV dem? ‖

A ‖ r den `STÖRSTA ‖

A instructs B to put one of his shapes to the right of the square. But as we can see, this is not sufficient information for B to carry out the task. So he asks for additional clarification. Which square? (In mid-key.) A in turn 'adds' the required information (= the biggest) also in mid key.

Low key on the other hand imbues a tone unit with the quality of equivalence. In the Saussurian general paradigm, one would never find an instance of 'triangle' being considered synonymous with 'hat'. But in the 'existential paradigm' (Brazil 1985), of C-Game no.12, this is precisely what happens:

A (så den här snögubben han har precis som -eh-) ‖ p+ din

kiNES ‖ o han har tagit PÅ sig 'n HÄR- ‖ r `STORa triANGeln ‖

B ‖ p som HATT ‖ A ‖ r på sitt `LILla HUVud ‖ ...

Player A has already presented a configuration of shapes as representing a snowman, even to the extent of availing himself of the masculine personal pronoun. "This snowman he has, just like your chinaman, put on this big triangle--" Whereupon B chimes in with: "As a hat!" but in low key. The two are presented as existentially equivalent, one and the same. The triangle is a hat. 'A' has no objections, and continues with the same line of imagery: "...on his little head..." in mid-key, of course.

The 'Go-Ahead' Phase in the C-Games: A function of the r+ Tone

The dominant version of the referring tone (r+) has a number of practical functions depending on the context it occurs in. In using the r+ tone a speaker invokes convergence and dominance simultaneously. As Brazil says, we "have to be clear about the distinction between the generalizable meaning of the formal opposition and the precise interpretation that might be placed on it in any one instance"; the local meaning of the '+' option is often affected by very obscure considerations of conversational finesse' (Brazil 1985). While a speaker may, on occasion, select the '+' option, and so project an assumption of his own dominance in a way which is acceptable or offensive to the hearer, we must think of this particular part of the intonation system as serving primarily to facilitate the smooth exchange of control of the discourse.

In the C-Games, instructors regularly break up their information into small 'digestible' chunks. They lead the constructors through each step, always making sure that their partners are 'with' them. They regularly 'call' for response, most often with high termination which sets up the expectation of an active adjudicating response; but also with mid termination in expectation of concurrence: (C-Game no.17:

(A:) ‖o så att spetsen VILar eMOT- ‖p ↑TAKet ‖o på-s ‖

‖o ↑PÅ - ‖ r+ ↑FYRkanten ‖(B:) r+ a↑HA ‖(A.) r+ ↑MHM ‖

‖r och triANGeln ligger alltså '↑OVanför ‖(B:) r+ ↑MHM ‖

‖(A:) r BORT från dig ‖ (B:) r+ ↑MHM ‖.....

At these points in C-Game discourse, the responses are almost invariably non-lexical, or if lexical, are usually limited to 'ja' or 'jaha'. So again it is the intonation alone which carries the meaning factor. Apart from key, which is discussed more thoroughly in other sections of this article, the crucial factor is tone. In all of the 18 C-Games, when the instruction has been correctly understood, the tone of the response is the r+ - ie, the dominant version of the referring tone. In C-Game no.17, there are 36 instances of the r+ tone in the above-described position. In no.18, a shorter recording, there are 21, and so on.

This phenomenon is often called feedback or back-channeling. In the C-Games I would prefer to call it the 'go-ahead' phase or turn, which carries, I hope, a more active connotation than the the other terms. The constructor confirms common ground, as 'requested', (R tone) but at the same time reserves in some sense for himself the prerogative of discernment. He 'goes along' with the instructor (assuming he has understood!), but he also signals that in his judgment (the + factor) the activity may proceed. In this frame of mind, the r+ tone is the appropriate choice.

When it does not occur in the position, and some other tone takes its place, or silence, (with the exception of non-verbal responses such as head-nodding, etc.) it is usually a case of the constructor not understanding the instructions sufficiently to carry out the task, as in

C-Game no. 18:

(A): ‖ o EH- ‖ o ti- PÅ- ‖ p den 'LILla kvadratens- ‖

‖ p 'VÄNstra 'SIDA ‖ (B): r+ MHM ‖ (A): p har jag lagt

den 'LILla ROMBen ‖ ... ‖ (B): r tan GERar de varandra eller... (?) ‖

It is perhaps appropriate to remark at this point that the same phenomenon also prevailed in the C-Games in other languages as well, notably Italian, French, Spanish, German, English, and Finnish. There were in fact only 2 other language C-Games in all, in which the constructor took a different strategy. One was in Spanish and the other Japanese, and in both cases there was consistently a p tone ( ) in the 'go-ahead phase. In the case of Spanish it seemed to be an exception to the general strategy, as the other 3 C-Games in Spanish more or less conformed to the 'general norm' of the r+ tone. A gloss for the line of thinking behind the p tone here might be 'I have hereby registered this information'. The local effect is rather of a 'detached' or 'scientific' attitude. It is also notable that the tone was always carried by the lexical item 'si' and not 'MHM' or 'MHMHM' as is otherwise usually the case.

In Japanese, on the other hand, there may be grounds for postulating a language specific intonational convention. The R tone, as we have said before, invokes common worlds. It is a well-known part of Japanese culture and psychology to 'veil' personal reactions and emotions of any kind. This characteristic is manifestly evident on all manner of levels, not the least intonation. In the Japanese C-Game, the 'go-ahead' phase is realized in the form of a staccato p tone on the word 'hai' (yes). In this way emphasis is diverted from the common ground aspect of the interaction to a simple notification that a message has been heard. It is not polite in Japanese to place oneself in any way on the level of the teacher or instructor. One simply accepts what the teacher has to say. (Please remember the similarites between the C-Games and a teaching situation.) This intonation reflects the socio-cultural ideal of personal reserve so fundamental to the Japanese way of thinking. Already built into the language and intonational conventions lies the cultural psychology of a nation. Which came first?

The p+ Tone in the C-Games

The dominant version of the proclaiming tone, which is characterized by a rise-fall (↗↘) contour, is by far the least common of the tones, and occurs only under rather specific conditions in the C-Games. One of these conditions will be discussed later in section E on the Comparison Phase. This instance can be characterized informally as 'negative discovery/surprise.' However, it must be emphasized that this is not a notional or attitudinal label to be attached to the p+ tone. Notional and attitudinal labelling have diverted attention from the abstract common denominator, which is neither lexical nor emotional nor situational. In the case of the rise-fall tone, such labels have consisted of things like 'surprise', 'sarcasm', 'indignation', and so on. Actually it is the 'doubly world-changing nature' of the p+ tone, plus the social and psychological factor of dominance which, interrelated with other intonational and contextual factors, contribute to the emergence of such locally specific interpretations.

Other than in the comparison phase, there are basically 2 types of instance in which the p+ tone occurs in our material. Take C-Game 7 for example:

(A): (... sitter som hatt, med spetsen uppåt, vad, och sen täcker det, går det ner lite grann över, s'att det liksom, så huvud sticker upp i hatten lite grann)

(B): ‖ p+ Å (!) ‖ p+ avan<u>CER</u>rat (!) ‖ p+ <u>AH</u>(!) ‖

The local gloss here might be 'discovery and appreciation', ie. 'This is both new <u>and</u> clever.' Doubly world-changing. 'I congratulate you (the + factor)'. Several other parallel instances are to be found in the C-Games.

The other main context for p+ tones in the games can be related specifically to the dominance factor inherent in the tone. It occurs where the 'instructor' takes on a 'didactic' sort of role: (C-Game no.11)

(A): ‖ o SÅ- <u>FINNS</u> ‖ p+ den 'LILla kva<u>DRAAT</u>en ‖ p+ <u>LÄNGST</u> ut ‖

‖ p+ till ↑<u>HÖGe</u>r ‖

The instructor speaks extra clearly in what could locally be characterized as a 'teacherly' tone of voice. In an abstract sense it is largely the choice of the dominant version of the tone (used 3 times in this short bit) which creates this effect. A kind of 'now this is what you do and this is how you do it' or 'this is new for you and I am telling you all about it' - approach. Definitely in charge.

Key and Termination: Pitch Concord in the C-Games

The phenomenon of pitch concord is an all pervading feature of the C-Games. The cooperative nature of the activity can account for the high instance of this interactive feature of expected/projected termination/key correspondence between speakers across turns. Briefly, pitch concord refers to the relationship between one speaker's projected expectations of the type of response he will get from the other speaker, either adjudicating or concurring, and also of the key choice of that response. (Please see overview chart.) For example, in C-Game no.2, where one placement of a shape on the table has been completed, and the next begins, we have:

(A): ‖ r+( ) SEN går du ut till klockan ↑TRE ‖ (B): r+ ↑JA ‖

(B): ‖r+( ) på den 'STORa '↑CIRkeln ‖ (B): r+ ↑MHM ... ‖

Player A is checking at each step of the way that B is 'with' him. In both exchanges he is 'asking' for an active, adjudicating response from B, and at the same time projecting an expectation that that response will commence with high key. Now, player B is following A's line of thinking completely, and having accepted the role of constructor, is also fully in agreement with being told what to do. Player A, through high termination, is effectively asking, 'Are you with me or are you not with me here?' B responds both times in the affirmative, and also complies with A's expectations of high key.

In C-Game no.5, we have a rather different configuration: B is re-formulating an instruction to make sure that she has understood correctly (Which she firmly believes that she has.):

(B) (Jaha, naha, s'att det är alltså den rätvinkliga

spetsen så att säga på triangeln) ‖p som ligger mot spetsen på

↑ROMBen ‖ (slight pause) ‖(A): r+ NEJ ‖ (skratt/laugh, laugh)

‖p det är det ↑INte ‖

Now, had her assessment of the situation been correct, the most likely response (as found in almost all similar instances in the other C-Games) would have been an immediate, high key (r+) ‖ JA ‖ or ‖ MHM ‖ which is clearly what she was expecting. A's slight pause, coupled with 'nej' and non-compliance with concord expectations indicates that, indeed, something is amiss. Concord is a sign of 'compliance', or 'togetherness', while non-concord sets the utterance up as being outside the area of speaker-hearer convergence. The total unexpectedness of the answer causes both A and B to burst out laughing.

Towards the beginning of C-Game no.13 we have still another configuration:

(A): ‖p nu BÖRjar jag HANS ‖ r talar OM för dig hur du ska

lägga HÄR ‖ (B): r+ ↑MHM ‖ (A): r MHM ‖ (B): p JA↓ ‖

A's mid termination in the last tone unit of her first utterance sets up the expectation of a simple concurring 'yes' in mid-key. No difficulties have yet arisen, as this is the first placement of the game. The goal has been mutually agreed upon, so there is no reason for her to expect any thing but concurrence. But again, as it is the beginning of the exercise, B is establishing his own role as equal partner in the enterprise. So he opts for a high key adjudicating response instead, perhaps to signify 'Right, that's a good idea in my judgment as well.' His response simultaneously fills a sort of framing or structural boundary marking function as well, and it is he who closes the pitch sequence with a low termination ‖ JA↓ ‖ The + version of the tone, 'concord breaking' and pitch sequence closure are all attributes of dominance. To which B claims his full share. A gloss to the whole sequence might be, 'Right, let's get this show on the road!'

Having given two examples of pitch concord and two examples of concord-breaking, I would like to specify that it was rather difficult to find the instances of non-concord, while concord was to be found everywhere in the C-Games. Indeed this interplay of matching pitch levels between speakers is perhaps the most striking feature of the intonation in the C-Games.

Absence of pitch concord, on the other hand, is usually quite conspicuous. In a recent summary of the Discourse Model, Cooper-Kuhlen (1986) writes that the pitch concord aspect of the model is applicable in relation to the organization of assymetrical discourse, but overly rigid in the case of symmetrical discourse, even to the point of breaking down. This is due to a misunderstanding of the way the system works. Speakers may or may not fulfill pitch concord expectations, and as seen above, this can have a number of causes. The speaker always has the prerogative of choice, whatever is expected of him and whatever the consequences. Although there is perhaps 'no such thing' as totally symmetrical discourse (Per Linell 1987), it can safely be said that the C-Games represent more or less symmetrical discourse, or as Brazil (1985) points out, a situation in which there is an 'ongoing, albeit incipient competition for dominance.' In the C-Games, pitch concord may be viewed as a normative feature which is observably oriented to by the participants. Perhaps it is plausible to postulate a sort of intonational adjacency pair in the tradition and terminology of Sachs, Schegloff and Jefferson. In this way pitch concord and non-pitch concord could be likened to the 'preferred' and 'dispreferred' seconds of conversational analysis.

The Comparison Phase in the C-Games

Out of the 18 C-Games in the study, 14 sets of players managed to negotiate their patterns correctly. The remaining 4 pairs ran into difficulties which resulted in dissimilar patterns. The results, correct or incorrect, were 'unveiled' in the final phase of the activity. Again different strategies were taken , which fell into a small number of basic groups. Here are two examples of the comparison phase in correctly negotiated patterns:

(no.10). (B): ‖ r ↑JO ‖ (A): r+ ↑JA ‖

(no.1): (A): ‖ r+ ↑JA ‖ p HELT rätt ‖ (B): r ↑JA ‖ p `ENaSTÅende ‖

In such cases of correctly negotiated patterns we can look at the initial tone choices of each player, usually on the words 'JA' or 'JO', which can be roughly glossed as follows:

r 'Contrary to expectations,(high key) we are together and successful here (r tone), don't you think?' (high termination) (7 examples)

r 'Yes indeed, (mid key) we are together and successful here. (r tone) Surely you agree(?)' (mid termination) (only 1 example)

r+ 'Contrary to expectations,(high key) I judge (the + factor) that we are together and successful (r tone). Don't you think(?)' (high termination) (3 examples) r+ 'Yes indeed, (mid key) I judge (the + factor) that we are together and have done the exercise correctly (r tone). You will agree. (mid termination) (3 examples)

p 'I tell you (the investigator, not my partner,) that the exercise is correctly done. (p tone) You agree. (mid termination) (1 example)

Worthy of note is: firstly the prevalence (14 to 1) of R tone in this summing up phase. The players have just received confirmation of the success of their joint efforts, and the referring tone reinforces this 'jointness'of a mutual world. In fact the only exception to this is when a player turns to a third party in order to give notification of the correct completion of the activity, with a proclaiming tone. Also worthy of note is the preference for the simple r tone (↘↗) complex with high key/termination. Here the emphasis is on togetherness, ie 'we have done this together'. The simultaneous preference for high key invokes the notion of 'this is not necessarily what one might have expected.

The r+ ( ) tone was also chosen quite often, sometimes before or after a simple r tone. Having no videotapes of these games, I can only speculate that those who chose the dominant version, which in this case seems to imbue the speaker with the role of 'judge', - had their attention fixed on the finished patterns, whereas those who chose the simple r tone (non-dominant) were more concerned with their partners.

Now in the case of the less successful games, it will be even easier to characterize them. The following exemplification from C- Game no.12 is typical:

(B): || r+ ↑AH(!) || p+ (↗↘) jaså du lägger den så ↑DÄR(!) ||

|| p+ JA ↑HA (!) ||

The p+ tone has been said to be the least common tone, both in the C-Games and in general. Here, however, in the context of the unveiling and discovery of a mistake, it is basically the only tone used for commenting on the mistake.

In the 4 'failed' C-Games, there are 5 examples of this tone. In fact, it occurs in every mistaken game. So even though there is no 'rule' that it must occur, this ratio nonetheless gives an indication that it is very likely to occur.

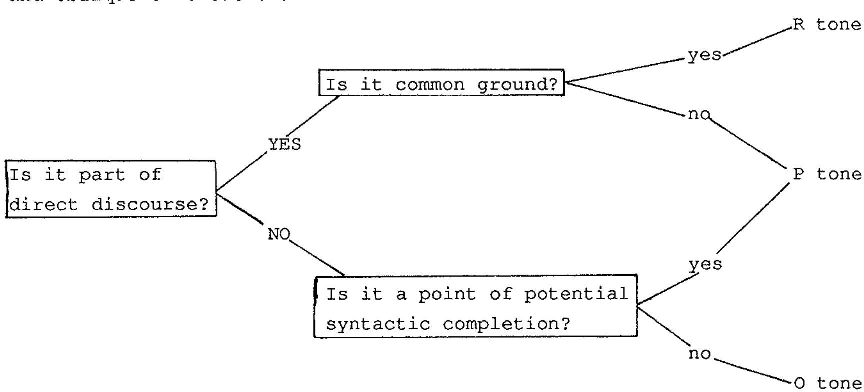
And no wonder. In the C-games, the players have expressly reached an agreement that the exercise has been satisfactorily completed before the unveiling takes place. This is part of the rules of the game.

Remembering that the basic meaning increment of the p+ tone is double world-change, a mistake discovered at the time of the 'unveiling' will by definition be 'new information' to both participants.

Oblique Orientation in the C-Games

In the discourse system a distinction is made between édirect' and 'oblique'orientation (Brazil 1979, 1985). Direct orientation refers to the intonational stance of the speaker when relating to a here-and-now state of understanding. Adjustments are made on a moment-by-moment basis according to the you and me, I and we situation. Oblique orientation on the other hand, can be seen as a temporary side- step from this interactive relationship, for one reason or another, with a consequent concentration on linguistic organization instead. A chunk of language is presented as - just that, a chunk of language.

The following diagram, adapted from Brazil 1985, may rapidly provide a visual clarification of the tone distribution associated with direct and oblique orientation:



The tone unit may also change its character in oblique orientation. In the 'usual'case, the tone unit consists of one or two prominent syllables, one of which is tonic. (In Swedish, exceptions to this rule are described in Section 1. An oblique tone unit has no such limits. Consider the case of rapid recitation of the alphabet in Swedish:

|| ABCDEFGHIJKLMNOPQRSTUVXYZÅÄÖ! ||

A 28 prominence tone unit. Key and termination, on the other hand, retain their respective meaning components in oblique orientation.

Oblique orientation serves a number of functions in discourse, among them stylized announcements (‖ o NORRMALMSTORG ‖) or in instructions (se UPP för DÖRRarNA- o DÖRRarna STÄNGS ‖), prayers, all manner of ritualistic language, some kinds of quotations and recitations, and so on.

In the C-Games, oblique orientation constitutes a relatively high percentage of the tone units involved, and is usually associated with hesitation and verbal planning (remember we said that in oblique orientation the speaker turned attention away from interaction and towards the language in question). Consider the following example from C-Games no.3:

(A): ‖ o ↑SEN - ‖ o ↑I DEN - ‖ o ↑'HÖGRA - ‖ o NEDRE - ‖

‖ p ↑SPETSen ‖ (B): ‖ r+ ↑MHM ‖ p lägger du den

'STORa ↑'CIRKELN ‖

This is the second placement of the game, and A is the first instructor of the set. The entire activity is new to the speaker, and he is fumbling for ways of getting his point across. With o SEN - (=THEN), he marks off the boundary of a new placement with high key (a kind of focussing device). In the tone units that follow ('on the - right - lower - tip...) He is having enough difficulty finding the right words, let alone trying to classify them simultaneously as belonging to common or separate worlds. Oblique orientation is 'all he can manage' at this point.

The discourse takes on a 'chopped-up' quality with these short oblique tone units. When the lexical search for positioning terminology is over, however, the speaker has no difficulty whatever in planning a much longer tone unit in direct orientation, ('You place the big circle'), - which was he was 'waiting to say' all along.

It is interesting that oblique orientation is also widely used in C-Game data in other languages as well, and in similar ways. Please note the following example in French:

(A): ‖ r+ la loSANGE (?) ‖ p il ↑TOUCHE la losange ‖

‖ o il TOUCHE l'ANGLE ‖ o du GRAND - ‖ r du GRAND

losange ‖ (B): r+ OUI ‖ (A): r+ ↑MHM ‖ O LE - ‖

‖ r/o GRAND triANGLE ‖ p/o TOUCHE ‖ p l'a - l'angle ouVERT du GRAND losange || (A): p j'eSPÈRE c'est pas

un test d'intelliGENCE tout ca! ||

Here again we have the 'choppy bits' in O - orientation so typical of this type of discourse. Speaker A (the instructor) clearly is not accustomed to running into such planning difficulties otherwise. This he expresses in no uncertain terms, in an exclamation which he has no difficulty whatever in planning as a unified whole. And very directly.

Further Considerations in the C-Games

There are many interesting areas still to be explored in the communication games, which will no doubt prove illuminating on many levels. Among them would be the following:

1. The different intonational strategies taken for negotiating meaning. In the C-Games there are two basic strategies taken by instructors, with many sub-categories and combinations. The 1st was 'instructions for building the pattern', and the 2nd was 'description of the pattern'. Internally, beyond the basic 'to the left, right, above, below' approach, we have analogies such as the clock and the compass, as well as figurative imagery (a house with sun and moon, a snowman balanced on a pile of sugar-boxes, a Chinaman kicking a ball, a Christmas tree with presents, etc.) Most interesting at the moment from our point of view is the role of intonation, particularly the P/R opposition of common or separate existential worlds used for creating mutual understanding of this type of here-and-now meaning.

2. The way different personalities are expressed through intonation. For example the subject with his 'head in the intellectual clouds' who tends to speak in long semi-oblique tone units, 'getting in' as much as possible into the limited space of a tone unit. Or the subject who has several mental 'channels on' simultaneously, and breaks up or combines his tone units accordingly, in order to 'get in' more than one 'channel' at a time.

3. The inter-relationships of discourse control, and the multi-layered role of intonation in creating and establishing power (dominance), facilitating power exchange, as well as creating a sense of equality between the speakers. What factors are involved in C-game discourse?

4. Male-female differences and relationships intonationally speaking, and what factors in intonation play a role. There are certain more or less sex differentiated conventions in Swedish. What are they and why? How do they show up in the C-games?

5. How different social relationships between the players, plus different attitudes towards the C-Game itself, can affect or alter intonational patterns and strategies. The less formal, the less predictable.

6. Selectivity and the prominence system - how it contributes to the negotiation of meaning in the C-Games.

7. Discourse and relationship structuring through the use of the pitch sequence in the C-Games.

Concluding Remarks

It goes without saying that the C-Games represent only a very limited type of discourse. Nonetheless it is hoped that this brief exposition will have given the reader an indication of some of the ways in which the Discourse Model may be applied to other types of discourse as well. The model, being finite, renders intonational meaning imminently accessible to a wide range of analysts, and not 'too difficult' after all. The functional categories in the model may also provide an impetus towards the search for acoustic correlates and the varying ranges involved in their expression. The Discourse Model's applicability to the Swedish language has been illustrated, and preliminary questions related to the tone accents have been discussed.

There are of course areas of prosody which are not included in the system, such as loudness, voice quality and overall pitch level. This system deals only with linguistic intonation; systematic factors resulting from choices of the all-or-none sort. This view places intonation solidly within language, rather than 'round the edge of language' (Bolinger 1961) - and right up on the level of grammar and lexis in terms of functional utility.

The Discourse Model is definitely not aimed at generating synthetic speech. As intonation is both context-created and context-creating, on both informational and psycho-sociological levels, a computer would need to think and feel like a human being in order to sound like one. On the other hand, it is not at all difficult to imagine that some of the elements of the system could be implemented in a stylized way in synthetic speech programs, which would result in a more humanoid-sounding organization than previously has been achieved.

As to Swedish as a foreign language, and the teaching of Swedish intonation, it is clear that both lesson content and methodology will require some re-thinking in the light of the insights of the Discourse Model. More emphasis will need to be placed in future on the function and meaning of intonation in action rather than on the mastery of static forms. How intonation _works_ and what it _means_. Shall we say, let's not talk about what intonation '_IS_', but rather about what intonation '_DOES_'.

Here, as everywhere else in science, there is still much to explore and much to learn. But a discourse approach to intonation in Swedish surely represents a major step towards the useful and fruitful cultivation of the 'no-man's land' that intonation and intonational meaning have so long inhabited.

Acknowledgements: Many heartfelt thanks to all those who contributed to the realization of these studies. Those who helped with recording, served as informants, gave advice (particularly Gösta Bruce and Robert McAllister), and technical assistance (Mats Dufberg, Lars Eriksson). And everybody at the linguistics and phonetics departments of the universities of Lund and Stockholm.

BIBLIOGRAPHY

Brazil, David. 1985. The Communicative Value of Intonation in English. (Discourse Analysis Monograph no.8) English Language Research & Bleak House Books. University of Birmingham.

Brazil, D. 1985. Phonology: Intonation in Discourse. Extract: Handbook of Discourse Analysis. Academic Press, London.

Brazil, D. 1985. Where is the Edge of Language? Review article. Semiotica 56 - 3/4.

Brazil, D. 1984. Sentences Read Aloud. Offprint: Intonation, Accent, and Rhythm. Studies in Discourse Phonology. Walter de Gruyter, Berlin, New York.

Brazil, D. 1982. Impromptuness & Intonation. Offprint from impromptu speech - a Symposium. Åbo Academy.

Inst. för lingvistik, Lunds Universitet.

Bruce, G. 1977. Swedish Word Accents in Sentence Perspective. Gleerups, Lund.

Brown, G. 1980. Questions of Intonation. Croom-Helm. London.

Bolinger, D. 1986. Intonation and its Parts. Vol.1 E.Arnold.

Cooper-Kuhlen, E. 1986. An Introduction to English Prosody. E. Arnold.

Coulthard & Montgomery (Eds). 1981. Studies in Discourse Analysis. Routledge & Kegan Paul.

Coulthard, M. 1977. An Introduction to Discourse Analysis. Longman.

Crystal, D. 1969. Prosodic Systems and Intonation in English. CUP.

Cruttenden. A. 1985. Intonation. CUP.

Einarsson, Jan. & Hultman, T. 1984. Godmorgon Pojkar och Flickor. Om Språk och Kön i Skolan. Liber. Malmö.

Elert, C.C. 1970. Ljud och Ord i Svenskan. Almqvist & Wiksell.

Garlen, C. 1984. Svenskans Fonologi. Mins 15. Stockholms Univ.

Gårding, E. 1983. Nordic Prosody 3. Chinese and Swedish in a generative model of intonation. Umeå

Garrod, et.al. 1984. Semantic Negotiation and the Dynamics of Conversational Meaning. Glasgow Psychology Report no.1.

Gumperz,J.J. 1982. Discourse Strategies. CUP.

Kjellin, O. 1978. Svensk Prosodi i Praktiken. Studieförlaget Uppsala.

Marshall, R. Rosenkvist, H. 1984. Uttal. Stockholms Universitet.

Linell, P., Gustavsson, L. 1987. Initiativ och Respons. Om Dialogens Dynamik, Dominans, och Koherens. Linköping.

Sinclair & Brazil. 1982. Teacher Talk. Oxford University Press.

Appendix:

DISCOURSE INTONATION OVERVIEW

THE TONE UNIT

Building block of verbal communication | Place of operation of a set of speaker options

PROMINENCE

Selectivity
vs
Non-selectivity

NB:

General vs
Existential
Paradigm

DISCOURSE CONTROL

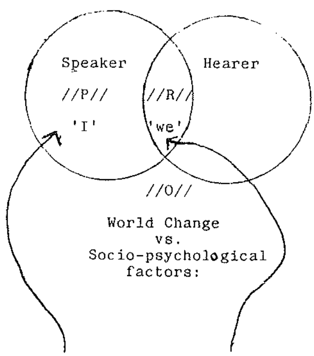
Consider:
Symmetrical
vs.
Assymmetrical
Speaker-Hearer
Relationships:

(+) Tone

Pitch sequence
closure

Non-concord

P vs. R



New	Given
Separateness	Togetherness
Proclaiming	Referring
(+) Dominance	- Dominance
Non-concord	Concord

TONE

p ↘

p+ ↗↘

r ↘↗

r+ ↗

o →

ORIENTATION

Direct - P/R
(Interactive)

Indirect - O/P
(Oblique)

Consider:
Speaker projection
(+ misjudgment!)

Speaker exploitation
(+ manipulation!)

of the VARIOUS CHOICES

KEY

(Inter-TU relationships):

High	Contrastive
Mid	Additive
Low	Equative

(Inter-sequential 'syntagmtic' relationships):

High	Maximal dysjunction (New topic or 'tack')
Mid	Additive (subordinate)
Low	Equative (existentially valid reformulation)

TERMINATION

(Interactive relationships):

	Projected response:	Projected/ invited response key:
High	- adjudicating	High
Mid	- concurring	Mid
Low	- pitch sequence closure	High / Mid / Low

RECENT NUMBERS AND CONTENTS

WP 29	Robert Bannert	From prominent syllables to a skeleton of meaning: a model of prosodically guided speech recognition
	Robert Bannert	Independence and interdependence of prosodic features
	Antonis Botinis	The prosodic structure of Greek: a phonological, acoustic, physiological and perceptual study
	Helen Goodluck	Notes on some particle and prepositional constructions in Swedish and English
	Eva Gårding	A visit to China 29.4.-7.6. 1986
	Eva Gårding, David House	Production and perception of phrases is some Nordic dialects
	Eva Gårding, Mona Lindau, Kjell Norlin, J-O Svantesson	Final report: Phonetic analyses of some non-European languages (LUCLA)
	Lars Gårding	A simple qualitative model for the vibration of the vocal folds
	Magnus Olsson	Backness and roundness harmony in Hungarian
	Shi Bo, Zhang Jialu	Vowel intrinsic pitch in Standard Chinese
	Bengt Sigurd	Understanding measure and comparative sentences through Prolog
	J-O Svantesson	A note on Kammu perception verbs
	Paul Touati	Structures prosodiques du suédois et du français: le cas de la focalisation et du contraste
WP 30	Kjell Norlin PhD dissertation	A phonetic study of emphasis and vowels in Egyptian Arabic

TRAVAUX DE L'INSTITUT DE LINGUISTIQUE DE LUND

Fondés par BERTIL MALMBERG
Publiés par GÖSTA BRUCE et BENGT SIGURD

I *Carl-Gustav Söderberg.* A Typological Study on the Phonetic Structure of English Words with an Instrumental-Phonetic Excursus on English Stress. 1959

II *Peter S Green.* Consonant-Vowel Transitions. A Spectrographic Study. 1959

III *Kerstin Hadding-Koch.* Acoustico-Phonetic Studies in the Intonation of Southern Swedish. 1969

IV *Börje Segerbäck.* La réalisation d'une opposition de tonèmes dans des dissylables chuchotés. Etude de phonétique expérimentale. 1966

V *Velta Ruke-Dravina.* Mehrsprachigkeit im Vorschulalter. 1967

VI *Eva Gårding.* Internal Juncture in Swedish. 1967

VII *Folke Stranger.* Les voyelles nasales françaises. 1969

VIII *Edward Carney.* Hiss Transitions and their Perception. 1970

IX *Faith Ann Johansson.* Immigrant Swedish Phonology. 1973

X *Robert Bannert.* Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage. 1976

XI *Eva Gårding.* The Scandinavian Word Accents. 1977

XII *Gösta Bruce.* Swedish Word Accents in Sentence Perspective. 1977

XIII *Eva Gårding, Gösta Bruce, Robert Bannert* (eds). Nordic Prosody. 1978

XIV *Ewa Söderpalm.* Speech Errors in Normal and Pathological Speech. 1979

XV *Kerstin Nauclér.* Perspectives on Misspellings. 1980

XVI *Per LIndblad.* Svenskans sje- och tje-ljud (Some Swedish Sibilants). 1980

XVII *Eva Magnusson.* The Phonology of Language Disordered Children. 1983

XVIII *Jan-Olof Svantesson.* Kammu Phonology and Morphology. 1983

XIX *Ulrika Nettelbladt.* Developmental Studies of Dysphonology in Children. 1983

Numbers III, VI, X, XIII and XVI are available from the Department of Linguistics and Phonetics, Helgonabacken 12, S-223 62 LUND.

STUDIA LINGUISTICA

Editors BENGT SIGURD and JAN SVARTVIK

One volume of Studia Linguistica is published each year in two parts. The subscription price is 60 SEK (approx. $ 9). Orders may be sent to any international bookseller or to the publishers, CWK Gleerup, Box 1205, S-223 62 LUND, Sweden.

PRAKTISK LINGVISTIK

Nr 1 - 1979 **Out of print**	*Eva Gårding & Robert Bannert.* Optimering av svenskt uttal
Nr 2 - 1979 **Out of print**	*Kenneth Hyltenstam* (Ed). Invandrarundervisning - aspekter på alfabetisering och den första nivån
Nr 3 - 1979 SEK 40,-*	*Robert Bannert.* Ordprosodi i invandrarundervisningen
Nr 4 - 1979 SEK 30,-	*Eva Wigforss.* Svensk pluralbildning hos finska invandrarbarn - en psykolingvistisk utvecklingsstudie
Nr 5 - 1981 **Out of print**	*Robert Bannert.* Svårigheter med svenskt uttal: inventering och prioritering
Nr 6 - 1981 SEK 30,-	*Gösta Bruce* (Ed). Tema prosodi - seminarieuppsatser i fonetik
Nr 7 - 1982 SEK 30,-	*Don Kulick, Karina Vamling & Lars-Åke Henningsson.* Om tolkning
Nr 8 - 1983 SEK 30,-	*Jan Fornell & Bengt Sigurd.* Commentator - ett mikrodatorbaserat forskningsprogram för lingvister
Nr 9 - 1984 SEK 40,-	*Thore Pettersson* (Ed). Färgterminologi - seminarieuppsatser i allmän språkvetenskap
Nr 10 - 1985 SEK 60,-*	*Kerstin Nauclér & Eva Magnusson.* Språket i skrift - om normal och försenad utveckling av tal och skrift
Nr 11 - 1985 SEK 45,-	*Bengt Sigurd, Jean-Jacques Berthout & Robin Cooper.* Språk på dator - hur datorn kan konversera, förstå, tolka morse och undervisa i franska

* Price incl cassette recording

Available from the Department of Linguistics and Phonetics, Helgonabacken 12, S-223 62 LUND.

LINGUISTICS DEPARTMENT
STAFF

Chairman	Kurt Johansson
Director of studies	Thore Pettersson/Kjell Weimer
Library	Magnus Olsson
Office	Janna Geggus, Eva Körner, Britt Nordbeck

General Linguistics

Cairns, Barbara	Fil Kand; extra teacher
Dooley Collberg, Sheila	BA; extra teacher
Eeg-Olofsson, Mats	Fil Kand; Assistant Lecturer
Gustafsson, Lars	Research Assistant
Henningsson, Lars Åke	Fil Kand; extra teacher
Horne, Merle	MA; extra teacher
Malmberg, Bertil	Fil Dr; Professor Emeritus
Nordbeck, Britt	Secretary
Pettersson, Thore	Fil Dr; Docent; Lecturer
Sigurd, Bengt	Fil Dr; Professor; Head of Division
Vamling, Karina	Fil Kand; Assistant; extra teacher
Research Students:	Jean-Jacques Berthout, Barbara Cairns, Hans Dahlbäck, Sheila Dooley Collberg, Boel de Geer, Janna Geggus, Lars-Åke Henningsson, Merle Horne, Gisela Håkansson, Magnus Olsson, Emilio Rivano-Fischer, Karina Vamling

Child Language Research

Bredvad-Jensen, Anne-Christine	Fil Kand; Research Assistant
de Geer, Boel	Fil Kand; Research Student
Mårtensson, Bodil	Fil Kand; Secretary
Söderbergh, Ragnhild	Fil Dr; Professor; Head of Division

Phonetics

Bannert, Robert	Fil Dr; Docent; Research Assistant
Bruce, Gösta	Fil Dr; Professor; Head of Division
Eriksson, Lars	Civ Ing; Computer Engineer
Geggus, Janna	Fil Lic; Secretary
Gårding, Eva	Fil Dr; Professor Emeritus
Hadding, Kerstin	Fil Dr; Professor Emeritus
House, David	BA; Assistant; Research Assistant
Johansson, Kurt	Fil Lic; Lecturer
Jönsson, Karl-Gustav	Technician
Körner, Eva	Secretary
Magnusson, Eva	Fil Dr; Logoped; Research Assistant
Nauclér, Kerstin	Fil Dr; Logoped; Research Assistant
Nettelbladt, Ulrika	Fil Dr; Logoped; extra teacher
Svantesson, Jan-Olof	Fil Dr; Research Assistant
Weimer, Kjell	Fil Kand; extra teacher
Wood, Sidney	Fil Dr; Logoped; Research Fellow
Research Students:	Antonis Botinis, Anne-Christine Bredvad-Jensen, Christina Dravins, David House, Dieter Huber, Boris Larnert, Lili-Ann Rudberg, Gabriella Stenberg-Koch, Paul Touati, Yasuko Nagano Madsen, Kjell Weimer, Ursula Willstedt

Voice and Speech Training

Norman, Lennart	Fil Kand; Voice and Speech Teacher

Address: Department of Linguistics and Phonetics
Lund University
Helgonabacken 12
S-223 62 LUND