PREDICTION AND PERCEPTION OF FOCAL ACCENTS

Anja Elsner
Institut für Kommunikationsforschung und Phonetik (IKP), University of Bonn, Germany

ABSTRACT

Two experiments were carried out in order to test the hypothesis that words with higher informational load are more prominent than others, i. e. they are realized with a focal accent. In a prediction test, subjects had to mark the words in a written text which they would presumably realize with a focal accent. In a perception test, using the same text material as before, but now with spontaneous speech, the same subjects had to mark the words which they perceived with focal accents. Results show that prediction and actual realization of the focal accents differ in several respects. Additionally, acoustic measurements showed clear correlations between subject judgements and acoustical salience of marked words for the perception test.

1. INTRODUCTION

Speakers use prosodic means for highlighting certain words in their utterances differently. Bolinger [1] pointed out that the intention of speakers is more relevant for accenting than syntactic rules: *"Location of sentence accents is not explainable by syntax or morphology - what item has relevantly stronger stress accent in intonational pattern, is a matter of information, not of structure."* So, he concluded that accents can only be predicted when the hearer is a 'mind-reader'. One aim of this paper is to examine this hypothesis.

To start with, a definition of the term 'focus' as used here is given: Prosodic focus is defined as the word with the most prominent accent in a phrase or a sentence. It often marks particularly important elements in an utterance. Bolinger [1] used the term 'point of information focus' to indicate that the degree of prominence which each word receives depends to some extent on its relative importance within the sentence and also on the context of the sentence itself. Words can be 'focused' or 'highlighted' to signal newness or contrast and they are marked by pitch accents.

2. THEORETICAL BACKGROUND

For Ladd [2], the term focus is a semantic notion. In a written text, focus can only be determined by context. In fixed contexts (term is already mentioned or is fixed by a question) a focus can be determined logically: the focus can be on the whole constituent (broad focus) or on a word or syllable (narrow focus). In the acoustic realization a focal accent is placed. In polysyllabic words, the focal accent lies normally on that syllable which is defined by the lexicon (lexical word accent).

In a definition by Cruttenden [3], focus is marked by nucleus placement in intonation groups. The nucleus is the pitch accent which stands out as the most prominent in an intonation group. In most cases, the most prominent accent is the *last* pitch accent in an intonation group. Another rule is that the nucleus is more likely to be placed on lexical items (content words: nouns, main verbs, adjectives, adverbs) than on grammatical items (function words: articles, prepositions, pronouns, etc.).

Broad focus represents a kind of 'normal' focus. The nucleus falls in general on the last lexical item of an intonation group (besides some 'regular' exceptions). Narrow focus is applied in context-bound sentences. In this case, the nucleus can also fall on grammatical items to mark a special contrast or emphasis.

For German spontaneous speech, several rules have been defined to annotate accent position automatically in a database [4]. The rules make use from part-of-speech of the word and the position of the word in a phrase. To apply the rules, syntactic-prosodic phrase boundaries must already be known.

3. DATA

The speech material consists of German spontaneous dialogues (supplied within Verbmobil), in which two participants negotiate the date, time and place for a meeting. Focus accents were perceptually labelled by the author for 11 dialogues (195 utterances with one or more phrases, 502 focal accents) with 10 different speakers (3 female, 7 male).

For the experiments, 80 dialogue utterances (with one or more phrases) were selected from the already labelled ones, most of them were taken out of the context. This was done to test the prediction without special knowledge, so that a prediction of mostly 'normal' focus (broad focus) was expected. The dialogues were selected by perceptual reasons: mainly the dialogues with weaker focal accents were taken to improve the labelling of the author as an additional aim.

categories	V	IV	III	II	I	0	III - V	I - II	I - V
Prediction	3.8	5.8	7.0	8.0	10.8	64.6	16.6	18.8	35.4
Perception	6.1	6.8	7.2	4.4	9.6	66.4	19.6	14.0	33.6

Table 1: *Percentages of words with different subject scores (counted on total number of words).*

categories	V	IV	III	II	I	0	III - V	I - II	I - V
Number of words marked by subjects	63	66	75	46	100	689	204	146	350
Number of words marked by author	63	62	64	11	14	825	189	25	214
Percentage subjects vs. author	100	93.9	85.3	23.9	14	-	92.7	17.1	61.1

Table 2: *Perception of focused words; comparison between test subjects and labelling of the author.*

4. EXPERIMENTS

5 phonetic experts (members of the institute) did the predictions and the perceptual labelling. Between the prediction test and the perception test was an interval of one month. Consequently, in the perception test, the subjects should not be influenced by their former predictions.

4.1. Prediction Test

In the first experiment, the written transcription of the spontaneous utterances was presented to the subjects. They had to underline all words in the text where they would apply a focal accent when speaking the utterance. The assumption is that the subjects will only underline 'important' words, for example content words and words which could contain 'new' information.

Results are shown in table 1, in the row 'prediction'. From all words (1039), a percentage of 35.4 % was marked by the subjects. The categories in the table correspond to the number of subjects which marked the words. The categories 'III-V' and 'I-II' were counted together - in the evaluation of the results they are defined as 'marked' and 'unmarked', respectively. The percentage for the 'marked' category is lower than for the smaller scores. For a written text, there are more possibilities to put focal accents, thus, agreement of subjects is not as high as expected.

4.2. Perception Test

In the second experiment the same subjects had to listen to the original spontaneous utterances from the dialogues. They could freely decide how often they wanted to listen to the utterance. The perceived focal accents had to be underlined in the written transcription. The number of markable focal accents was not fixed.

Results can be seen in table 1, in the row 'perception'. This time, a smaller percentage of words was marked (33.6 %). The percentage in the category 'marked' (III - V) is higher now than for the category 'unmarked' (I - II). The acoustic data reduces the possibilities for markable focal accents, thus, the variation between the subjects becomes lower. Remarkably, there is a high correlation between subjects for the unmarked words. For both prediction and perception about 2/3 of the words are unmarked for all subjects.

4.3. Improvement of Labelling

Another reason to conduct the perception test was to control the labelling of one person, i. e. the author. The labels for focal accents are used to train a detector for automatic recognition of focal accents [5]. So there is the question if judgements of only one person are a reliable base for the detection task. Table 2 shows the scores of the subjects compared with those of the author.

In general, subjects marked more words with focal accent (350 vs. 214 from the author). But the percentages show that there is a high agreement for the higher scores (III - V), i. e. 92.7 %. For the score V, the agreement is even 100 %, for score IV it is 93.9 % respectively. So it is possible to say that most of the author's labels are supported by expert subjects.

For further work in the automatic recognition task, as a consequence of these results, all labels with low agreement (I - II) were examined again by the author, most of them were removed then.

5. SOME EXAMPLES

In German, the marking of the date is normally expressed by denoting the number of the day (in ordinal numbers) followed by the name of the month. In the prediction test, all subjects predicted 'number of the day' to be associated with focal accent, but in the spontaneous data, 'month' almost always was perceived with a higher prominence.

We will present some examples now to examine further if there are general rules for the marking of date, time and place in our negotiation dialogues. The examples are given in English translation, in some cases German word order was kept.

5.1. Marking of date

First, there are presented examples with no difference between prediction and perception (accented words are in bold letters). The month or the weakday have normally the focal accent, except in the case of some time adverbs, like 'afterwards'.

- 'we'll meet on **Monday**'
- 'we'll take a/the **Monday** '
- '**Monday** afternoon'

2. PERCEPTUAL PROMINENCE IN SPEECH SYNTHESIS

The term *prominence* has been given a quantification by [3], resulting in its definition as a measure of perceptual markedness relative to the surrounding phonetic context.

The appeal of this approach lies mainly in the possibility to have an easy description (here: a scale between 0 and 31) of the prosodic characteristics of an utterance reflecting the perceptual impressions instead of directly describing the acoustic correlates of phonological concepts. The link between perceptual prominences and acoustic realisation has been studied and integrated into the prosodic component of a rule based synthesis system [9]. This approach seems to be especially interesting in the study of prosodic focus, since contextual parameters influencing the perception can be controlled in a comparably straightforward way. Previous studies already indicate a possibility of expressing narrow focus using the prominence approach in speech synthesis [13, 9]. In [13], indications were found that subjects preferred to perceive contrastive focus in sentences synthesized with high prominence values on the syllable carrying the main stress of the focal word. Still, the correlation between (absolute) prominence values and perceived contrastive focus was low. We concluded that not only the focal syllable but also the context needed some attention.

3. A CONTRAST EXPRESSING PROMINENCE PATTERN

3.1 The Experiment

Three different declarative sentences were synthesized with contrastive stress on three different positions in each sentence, using five different prominence patterns for each constellation. The most prominent syllable within the focus exponent was chosen as focal syllable. The original prominence values were taken from the VERBMOBIL generation module which calculates prominence values using syntactic and lexical information, if no further semantic/pragmatic information is given. The result is a default prosodic pattern. Those patterns were manipulated according to the methods illustrated in Table 1. The following stimulus sentences were used (SMALL CAPITALS indicate all the possible locations of intended contrastive focus).

ENde MAI bin ICH noch im Urlaub. End of May am I still on vacation.
Es würde mich freuen, wenn WIR noch EInen TerMIN ausmachen It would me please if we another one appointment made.
ANfang MAI hätte ICH noch Zeit. Beginning of May would have I still time

Each sentence was further supplied with two question contexts. The first context matched the contrastive accentuation pattern in the answer, the second context ought to produce an odd (if not ungrammatical) impression if the accent were interpreted as a correction contrast. The following sentences are examples for a such a contrast in a matching (3) and one in a non-matching context (4).

(3) Q: Anfang Mai sind Sie also noch im Urlaub?
Q: Beginning of May are you so still on vacation

A: ENDE Mai bin ich noch im Urlaub
A: END OF May am I still on vacation

(4) Q: Ende Juni sind Sie also noch im Urlaub?
Q: End of June are you so still on vacation?

A: ??ENDE Mai bin ich noch im Urlaub.
A: END OF May am I still on vacation.

The context questions were read aloud and recorded in an anechoic chamber by a male native speaker of German and phonetic expert, who was familiar with the experiment. He was instructed to read the questions as if no context was given to prevent any bias towards a specific prosodic expectation. The resulting 180 question-answer pairs were presented to phonetic experts (n=11). Both question and answers were played via headphones, the questions were furthermore displayed on a computer screen. Subjects were allowed to listen to each answer several times and had to judge each question-answer pair on a scale between 1 and 6 (forced choice), with 1 being a very good and 6 being a very bad score (German school grades). Due to a (now fixed) mistake in the algorithm, one stimulus type (sentence 2, intended focus on *einen*) did not properly reflect the intended conditions and was thus eliminated from all further studies.

3.2 Results

Contexts matching the intended impression were rated significantly more acceptable than nonmatching ones (Kolmogorov-Smirnov, $p \leq 0.001$). Besides, there was a substantial negative correlation between the judgements and the matching vs. non-matching question-answer pairs ($\rho = -0,49$, $p \leq 0.01$) without taking into account any specific strategy of prominence manipulation. Figure 1 illustrates the distribution of judgements for matching vs. non-matching question-answer pairs.

general lower in percentage than the word before. This is in agreement with other results, for example [7].The differences in the percentages get even higher when we combine two or three score classes. In the sections before we defined the classes 'III - V' as 'marked' and the classes '0 - II' as 'unmarked'. In the last two lines of the table we see clear acoustic differences between the two classes. This holds true especially for the word after a marked word.

Streefkerk et al. [8] got similar results: They had 10 listeners to judge 500 read aloud sentences (in Dutch). They found a linear relation between cumulative subject scores (here equated with prominence) and F_0 range per syllable and loudness per vowel, respectively. The correlation with syllable duration was smaller, probably influenced by speaking rate and final lengthening factors

8. CONCLUSION

People apply special sitatuation knowledge when producing or listening to speech. So, in our domain of negotiation dialogues, time adverbs like 'before' and 'after' are more important than the noun 'clock' which is often omitted in the dialogues. For efficient communication, unimportant information has to be deaccented, but in this domain, time and date have to be articulated very clearly.

So, in the prediction test, subjects judged the day of the month to be more important than the name of the month. In the spontaneous dialogues however, the focal accents were realized as a broad focus on the name of the month. This can be explained with economical reasons: narrow or contrastive focus demands a higher articulatory effort [9]. This is usually avoided. Only for absolutely necessary contrasts, the narrow focus is employed.

Another possible reason for the difference between prediction and perception is that subjects didn't know the context. In general, words which have been mentioned before, tend to be deaccented in the following sentences. Without context, there was more 'new' information for the subjects, so that they predicted more focal accents than were realized in the spontaneous dialogues.

Results in general show, that an automatic labelling procedure as used in [4] can be useful even for spontaneous speech databases, because contrastive accents are somewhat rare. Nevertheless, it would be a strong improvement, if a kind of 'dialogue memory' would store the already mentioned items. As a consequence, in the following sentences, some accents could be excluded when the word was already mentioned.

Another aim of this study was to compare the labelling of several subjects against to the author's. The comparison of the subject scores and the author's labels showed a high agreement. So it is possible to use the labels from one expert as a reliable base for a training of detection algorithms.

The additional examination of the scores and the acoustic salience of the marked words showed a high correlation, too. Words which are marked by 3 - 5 subjects are acoustically more salient than the other words. This holds especially true for the words after the actual marked word.

However, since the study was based on a small corpus, no general rules can be concluded now for the marking of time and date in this special domain. Nevertheless it is possible to define probabilities in which one rule is applied more often than the other.

ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the Verbmobil Project under Grant 01 IV 101 G. The responsibility for the contents of this study lies with the author.

REFERENCES

[1] D. Bolinger (1972): Accent is predictable (if you're a mind-reader). Language 48, 633 - 644

[2] R. Ladd (1980): The structure of intonational meaning. Indiana University Press

[3] A. Cruttenden (1986): Intonation. Cambridge University Press

[4] A. Batliner, V. Warnke, E. Nöth, J. Buckow, R. Huber, M. Nutt (1998): How to label accent position in spontaneous speech automatically with the help of syntactic-prosodic boundary labels. Verbmobil-Report 228

[5] A. Elsner (1997): Focus detection with additional information of phrase boundaries and sentence mode. Proceedings EUROSPEECH '97, Vol. 1, 227 – 230, Rhodes

[6] R. Ladd (1996): Intonational Phonology. Cambridge University Press

[7] S. Eady und W. Cooper (1986): Speech intonation and focus location in matched statements and questions. J. Acoust. Soc. Am., 80, 402 – 415

[8] B. Streefkerk, L. Pols, L. ten Bosch (1998): Automatic Detection of Prominence (as defined by listeners judgements) in read aloud sentences. Proceedings of ICSLP '98, Sydney

[9] D. Erickson (1998): Effects of Contrastive Emphasis on Jaw Opening. Phonetica, 55, 147 – 169