THE INTONATIONAL MARKING OF FOCAL STRUCTURE: WISHFUL THINKING OR HARD FACT?

A. Batliner*, W. Oppenrieder*, E. Nöth#, and G. Stallwitz#

*** Institut für Deutsche Philologie, München, F.R.G.**
 # Lehrstuhl für Informatik 5 (Mustererkennung), Erlangen, F.R.G.

ABSTRACT

German is considered to be a language where the sentences normally have one prominent part (place of focal accent) which is marked by intonational means. In this paper, we will address the question whether more special structures like double focus vs. single focus or narrow focus vs. broad focus (focus projection) are marked by intonational means as well or whether they simply have to be extracted out of the linguistic context.

1. INTRODUCTION

The linguistic concept of focussing concerns an aspect of the semantic structure of sentences. Sentences can be divided into a focus component and a complementary background component. The focussed information is selected from a set of alternatives which are 'under discussion' in the given context, whereas the background is not related to such alternatives. The focus-background structure of a sentence is a function of the linguistic and non-linguistic context of the sentence. A suitable means of defining the focus in declarative sentences is the so-called question test: Only those expressions are focussed (i.e. selected from a set of alternatives 'under discussion') that replace the *wh*-expression in a *wh*-interrogative sentence (cf. table 1). Focussing is signalled by the focal accent (FA), i.e. the accentuation of at least one syllable in the focussed constituent, which is selected according to rules specifying the 'focus exponent' [8: 223ff].

2. MATERIAL AND PROCEDURE

In our study we examined the intonational form of four types of focus structure realized in two question (Q) types (declarative and inversion Q) and in two non-question (NQ) types (declarative and imperative sentence). The material consists of 3 different A.c.I.-constructions with a dependent transitive verb. Six untrained speakers (3 male, 3 female) produced a total of 360 sentences together with context sentences, which induced sentence modality, focal structure and thereby the FA. The intended FA in the embedded sentence can be on the 2nd phrase (2PHR), the 3rd phrase (3PHR) or on both phrases (2/3PHR); cf. table 1.

In this paper we want to address the question whether the focal structures 'double focus' and 'broad focus' are really indicated by intonational means or whether they have to be extracted out of the linguistic/situational context.

For each utterance the following features for the 2PHR and the 3PHR were extracted and normalized; for details see [2,4]:

- The maximal and minimal fundamental frequency (Fo) values MAX and MIN, transformed into semitones and normalized with respect to voice register by subtracting the lowest Fo value of the speakers;
- The difference DIF of the position on the time axis of MAX and MIN in centisec.;

Table 1: Examples of focal structure (focus underlined) and intended FA (capitalized); declarative sentence *She makes Nina weave the linen*; context sentences in English translation.

(Narrow) object focus FA on 2PHR:	What does the master make Nina weave? Sie läßt die Nina <u>das LEInen</u> weben.
(Broad) object/verb focus (focus projection),FA on 2PHR:	What does the master make Nina do? Sie läßt die Nina <u>das LEInen weben</u>.
Double focus FA on 2/3PHR:	What does the master make Nina do with which material? Sie läßt die Nina <u>das LEInen</u> <u>WEben</u>.
(Narrow) verb focus FA on 3PHR:	What does the master make Nina do with the linen? Sie läßt die Nina das Leinen <u>WEben</u>.

- The duration DUR in centisec. The normalization of the speaking rate took into consideration the average duration of that phrase for each speaker and the average duration of the syllables in the utterance [2:33].
- The maximal energy in the 0-5000 Hz band.

The parameter values were extracted 'by hand' on mingograms and automatically from the digitized versions of the utterances [2,4,6]. An average of 12 listeners participated in the following perception experiments: The test sentence was presented in isolation. The listeners had to decide which of the phrases carried the FA. If FA_i is the number of listeners who perceived the ith phrase as most stressed then

$FOK = (FA_2-FA_3)/(FA_1+FA_2+FA_3)$

takes on values between 1 (all listeners perceived the 2PHR as stressed) and -1 (all listeners perceived the 3PHR as stressed). FA on the 2PHR takes on values above 0.5 and FA on the 3PHR below -0.5. Double focus on the 2/3PHR is defined operationally as $|FOK| < 0.5$, i.e. those items - about 25% of the whole corpus - where the subjects are rather uncertain about the place of the FA. Note that this value is in a way arbitrary, and that it is not a strict definition, but rather an "in these cases it is likely that..."-way of defining the focal structure.

The results of a statistical classification procedure (discriminant analysis) will be reported for two different learn and test constellations:

l=t: All utterances were used for learning and testing with learn=test. This is the "best possible" constellation, i.e. it provides an upper limit for the predictive power of the variables, but over-adaption is likely.

l5t1: As a training sample we used 5 speakers, and the remaining speaker as the test sample (leave one out). This simulates speaker independence and avoids over-adaption.

Since a separate treatment of Qs and NQs [4:211] yields better results than when analyzed together, only these results will be discussed.

3. RESULTS AND DISCUSSION

In experiments like ours, the linguist defines the intended focal structure and thereby place and (possibly special) form of the FA. The subjects must comprehend the given focal structure and produce the FA 'in the right way'. The produced FA should be judged with perception experiments as described in part 2, because only then can we be sure that misproductions are filtered out. The acoustic parameter values can be used to predict (PREDFA) the perceived FA (PERCFA) as well as the intended FA (INTFA). The mapping from one step to another is never optimal. Table 2 shows for l=t and separated into Qs and NQs, 3 different crosstabulations. All variables were used as predictors; as for their respective relevance, cf. [2,4]. To give an example, the first 3 numbers in table 2b) read as follows: 81 Qs had the INTFA on the 2PHR; 42 out of the 81 had a PREDFA on the 2PHR, the rest on the 2/3PHR.

The following points shall be discussed briefly:

Table 2: Crosstabulations; for b) and c), the sum of the NQs is only 180, because 8 items could not be predicted for technical reasons.

	Qs (n=172)			NQs (n=188)		
a)	PERCFA			PERCFA		
INTFA	2	2/3	3	2	2/3	3
2	**49**	**31**	1	**69**	**3**	0
2/3	**21**	**26**	3	**57**	**16**	2
3	4	9	28	6	9	26

	Qs (n=172)			NQs (n=180)		
b)	PREDFA			PREDFA		
INTFA	2	2/3	3	2	2/3	3
2	**42**	**39**	0	**62**	**8**	0
2/3	**22**	**22**	6	**54**	**11**	5
3	4	5	32	5	13	22

	Qs (n=172)			NQs (n=180)		
c)	PREDFA			PREDFA		
PERCFA	2	2/3	3	2	2/3	3
2	**43**	**30**	1	**115**	**12**	0
2/3	**23**	**33**	10	**6**	**14**	6
3	2	3	27	0	6	21

- Double focus (INTFA on 2/3PHR) is not marked very often intonationally, cf. the bold numbers in table 2. It follows that subjects do not necessarily indicate double focus by intonational means. At least for the rhythmical structure and the linguistic and non-linguistic context of our test sentences, the two ways of expressing double focus might be free variants, in the case of the FA on the 2PHR a sort of pseudo projection [7: 274f]. Of course, the subjects might simply not have understood the intended focal structure. However, this is not very likely because in other perception experiments, where listeners had to judge the naturalness of the items [2:29f], double focus items with the FA on the 2PHR did not get worse scores than those with the FA on the 2/3PHR [7:275].
- There is a greater confusion between 2PHR and 2/3PHR for Qs than for NQs. The reason might be that in Qs, the Fo offset is mostly high, and that intensity covaries to a certain extent with rising Fo [2:38]. On the one hand, intensity is not relevant for Qs [2:41ff], on the other hand, this covariation might puzzle the listeners that much that their judgments are more uncertain and fall below the limit of FOK=0.5.
- The mapping PERCFA-PREDFA is best as expected, because here, perception is directly related via the acoustic features with the classification.
- Separation of verb focus vs. the other foci is best, i.e. separation of 'clear' single foci is very good, cf. table 3. In this table, percentages of errors are given for 3 different constellations:
a) Prediction of single foci in [2,4] with the border between FA on the 2PHR and FA on the 3PHR at FOK=0.0;
b) Prediction of 'triple foci' (FA on 2PHR, 2/3PHR, 3PHR) as in table 2c);
c) Prediction of 'clear' single foci, i.e. the confusion rate between FA on the 2PHR and on the 3PHR in table 2c).

Table 3: Classification errors in %

	Qs		NQs	
	l=t	l5t1	l=t	l5t1
single foci	5	14	4	6
triple foci	40	46	17	27
clear foci	2	2	0	2

As for the Qs, the above mentioned covariation of intensity and Fo might be the reason for the marked difference of 12% between 'single foci' and 'clear foci' for l5t1 in table 3. It follows from this table, that for automatic speech recognition, it might be suitable not only to predict the FA, but also to try to predict clear FA in order to eliminate wrong hypotheses with a high probability.

4. NARROW VS. BROAD FOCUS

Here, we will only report results for the NQs, because in Qs, the simultaneous marking of sentence modality and FA renders the discussion of the (poor) classification rate even more difficult.
It can be seen in table 4, that the 'realistic' recognition rate, 63% for l5t1 with no over-adaption, is rather low. That does not necessarily mean that narrow vs. broad focus is not indicated at all by intonational means: Because the sample

size is rather small (n=72), a few misproductions can influence the result markedly. Besides that, a close inspection of the individual speakers indicates a speaker specific use of the variables. Nevertheless, the mean difference of the 3 most relevant features DUR on the 2PHR and DIF on the 2PHR and the 3PHR can be interpreted. In Fig. 1, the mean values of MAX, MIN and DIF are given (x-axis: time in centisec. from the beginning of the utterances, y-axis: semitones normalized with respect to the speakerspecific lowest value). The values of MAX and MIN are almost identical. For narrow focus however, DIF is greater on the 2PHR and smaller on the 3PHR than for broad focus, i.e. the slope is less steep on the 2PHR and steeper on the 3PHR. In [5] it was shown that long Fo inflections are generally judged to be of greater impact than short ones of similar rate. If this holds true for our data as well, for narrow focus, the 2PHR is marked more clearly than the 3PHR. The same applies to DUR on the 2PHR (mean value 3.47 for narrow and 3.22 for broad focus). Note that in another perception experiment [2:30], the position of the FA was equally distributed on 2PHR (80%) and on 3PHR (20%) for narrow and for broad focus. If these two structures are marked differently at all, it may be by features (as DIF) that are rather irrelevant for the marking of the FA in NQs [2:41].

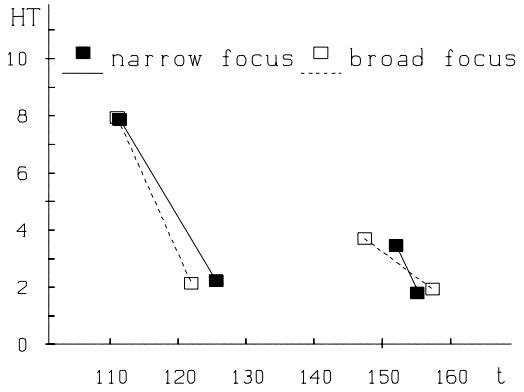
Table 4: Recognition rates for narrow vs. broad focus

	all features, stepwise selection	three most relevant features
l=t	73	65
l5t1	63	63

5. FINAL DISCUSSION

As far as our results can be generalized to other speakers and to spontaneous speech as well, the intonational marking of the special focal structures double vs. single focus and narrow vs. broad focus is neither a hard fact nor only wishful thinking: Double focus was marked intonationally sometimes, but not very often. It is not yet clear whether the differences between broad and narrow focus shown in Fig. 1 are stable across subjects and other material.

Fig. 1: Narrow vs. broad focus



REFERENCES

[1] ALTMANN, H./BATLINER, A./OPPENRIEDER, W. (eds.) (1989), *"Zur Intonation von Modus und Fokus im Deutschen",* Tübingen: Niemeyer.

[2] BATLINER, A. (1989), "Fokus, Modus und die große Zahl. Zur intonatorischen Indizierung des Fokus im Deutschen", In: [1], 21-70.

[3] BATLINER, A./OPPENRIEDER, W./NÖTH, E./STALLWITZ, G. (1990), "'Neue Information' im Sprachsignal. Die prosodische Markierung der Fokusstruktur", *Fortschritte der Akustik - DAGA'90,* 1059-1062.

[4] BATLINER, A./NÖTH, E. (1989), "The Prediction of Focus", *Proc. ECSCT, Vol.1,* Paris, 210-213.

[5] BLACK, J. (1970), "The Magnitude of Pitch Inflections", *Proceedings ICPhS, Prag 1967,* München: Hueber, 177-181.

[6] NÖTH, E. (1991), *"Prosodische Information in der automatischen Spracherkennung - Berechnung und Anwendung",* Tübingen: Niemeyer.

[7] OPPENRIEDER, W. (1989), "Fokus, Fokusprojektion und ihre intonatorische Kennzeichnung", In: [1], 267-280.

[8] UHMANN, S. (1987), *"Fokussierung und Intonation",* Phil. Diss., Konstanz. To appear under the title *"Fokusprojektion",* Tübingen: Niemeyer, 1991.