Deciding upon the relevancy of intonational features for the marking of focus: a statistical approach

Anton Batliner

Angaben zur Veröffentlichung / Publication details:

Batliner, Anton. 1991. "Deciding upon the relevancy of intonational features for the marking of focus: a statistical approach." *Journal of Semantics* 8 (3): 171–89. https://doi.org/10.1093/jos/8.3.171.

Deciding upon the Relevancy of Intonational Features for the Marking of Focus: a Statistical Approach

ANTON BATLINER
University of Munich

Abstract

We present results on how focus is marked intonationally in German. Six untrained speakers produced a corpus of 360 sentences. The corpus was constructed in such a way that sentence modality and place of focus could only be differentiated by intonational means. Acoustic features representing the parameters pitch, duration, and intensity were extracted manually or automatically. The relevancy of these features and the effect of several transformations were tested with statistical methods (discriminant analysis). Perceptual experiments where the listeners had to decide upon the place of the focal accent and to judge the naturalness and categories of the utterances were performed as well. By calculating average values for the (appropriately transformed) relevant features we found 'normal', prototypical cases; by looking at utterances where all listeners agreed on the naturalness and (intended) categories we arrived at coinciding results. At the same time we found 'unusual' but regular productions. Finally, the speaker-specific use of the different parameters is discussed and the question is addressed as to whether the parameters can be classified as relevant or irrelevant for the intonational marking of focus.

MATERIAL AND PROCEDURES

This paper is concerned with the prediction of focus; focus is the part of an utterance which is semantically most important. On the phonetic surface focus is marked by the focal accent (*Fa*). To be more exact, we will try to predict the phrase that carries the *Fa*.

Our material consists of 360 utterances, spoken by six untrained speakers (three male, three female). Three different sentences with a similar syntactic structure were each put in different contexts that determined sentence modality as well as place and manner of focus (simple focus, focus projection, or multiple focus); for a detailed description of the corpus and the intended focal structures, cf. Batliner & Oppenrieder (1989) and Oppenrieder (1989). In each of the sentences the last two phrases could be stressed, depending on the surrounding context. Based on the sentence modality system according to Altmann (1987), the sentences formed minimal pairs that could only be differentiated by their intonational form: *focus in final vs. focus in prefinal position* on the one hand, and *questions vs. non-questions* on the other hand. Table 1 shows an

example of a context sentence, the pertinent test sentence, and the induced sentence modality and place of focus. Table 2 shows the three test sentences, an (awkward) word-by-word translation into English, an appropriate translation, and a finer description of the induced sentence modalities question/non-question (*Q/NQ*), *NQ* being either assertion, imperative, or adhortative.

The only instruction given to the speakers was to produce the context and the test sentence. We did not instruct the speakers to produce the *Fa* or *Qs/NQs* in a certain way: by instructing the speakers, one can eliminate certain variabilities and facilitate the analysis. On the other hand one loses the chance to find regular and interesting deviations and merely receives several realizations of representative cases where representativeness is based on the intuition of the researcher.

By evaluating a relatively large number of cases we expected to find both representative cases (which we will call **central types**) and rarer but acceptable cases (which we will call **marginal types**). The data were evaluated in two ways that proved to be converging:

Table 1 Examples of context and test sentence, induced sentence modality and place of focus

Constellation of sentence modality and focus: Assertion, focus on 'linen'

Context: Mother: 'What does the master make Nina weave at the moment?'

Sentence: Employee: 'She makes Nina weave the linen.'

Table 2 Test sentences, translation, and induced sentence modalities

Sie läßt die Nina das Leinen weben?/.
She makes the Nina the linen weave
She makes Nina weave the linen
assertive question vs. **assertion**

Lassen Sie den Manni die Bohnen schneiden?/!
Make the Manni the beans cut
Make Manni cut the beans
polar question vs. **imperative**

Lassen wir den Leo die Blumen düngen?/!
Let us make the Leo the flowers fertilize
Let us make Leo fertilize the flowers
polar question vs. **adhortative**

(i) We extracted acoustic feature values that represent the prosodic parameters pitch, duration, and intensity. Using a statistical classifier we tested the relevancy of the features with respect to the place of the *Fa*. By calculating average values for the relevant features we found the central type of each *Q/NQ-Fa* constellation.

(ii) We presented the utterances to a forum of listeners who judged the naturalness, category, and place of *Fa*. Category roughly means sentence modality; as for the differences, cf. Oppenrieder (1988). By selecting the utterances that were judged to be the 'best' ones and by comparing the feature values of those utterances with the average values from (i) we found the central type as well as marginal types.

EXTRACTION OF FEATURES

For each utterance we calculated the following features:

(i) *For the whole utterance*: the fundamental frequency (Fo) at the end of the utterance (*off*); the all-point regression line of the Fo values (*reg*); the duration in centiseconds.

(ii) *For the 2nd and 3rd phrase*: the maximal and minimal Fo value; the difference of the position on the time axis of the maximal and minimal Fo value in centiseconds; the duration in centiseconds; the average and maximal logarithmic energy.

The parameter values were extracted 'by hand' on mingograms and automatically from the digitized versions of the utterances (cf. Nöth 1989 for details on the Fo algorithm and the computation of the energy values). In Batliner *et al.* (1989) we showed that automatically extracted Fo values produced recognition rates comparable to those from mingogram values. An automatic extraction of the durational values, however, would pose a problem (cf. Batliner & Nöth 1989: 212 f.).

PERCEPTION EXPERIMENTS

An average of twelve listeners participated in three different perception experiments:

(i) Context and test sentence were presented by earphone and at the same time in a written version. On a rating scale from 1 (test sentence matches very well with context) to 5 (test sentence does not match at all), the

listeners had to judge the naturalness of the production. We will name the average rating of the listeners *NAT*.

(ii) The test sentence was presented in isolation. The listeners had to classify the sentence as question, assertion, imperative, exclamation, or optative. We will name the percentage of classifications as question *MOD*.

(iii) The test sentence was again presented in isolation. The listeners had to decide which of the phrases carried the *Fa*. If *fai* 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 2nd phrase as stressed) and −1 (all listeners perceived the 3rd phrase as stressed).

STATISTICAL EVALUATION OF THE EXTRACTED FEATURES

'Best' transformations

Each of the intonational features was used as a predictor variable in the discriminant analysis to predict sentence modality (*Q/NQ*) and (position of the) *Fa*. Because of the combinatorial explosion the optimal feature combination had to be determined heuristically: the predictors entered the analysis separately and (if the feature was calculated for the 2nd and 3rd phrase) together with the corresponding variable for the other phrase. Several transformations for each variable were tested. In order to reduce the necessary amount of computation all cases were used both for learning and testing with learn = test ($l = t$). Throughout this paper, the analyses are based on this constellation, if not explicitly another constellation ($l\,5\,t1$ or $l1\,t\,5$, cf. below) is referred to. The relevant variables under the best transformation were put into multivariate discriminant analyses. We can only present the most important results; for a more detailed discussion see Batliner (1989a). The statistical method is fully described in Klecka (1980) and Norusis (1986). Further applications of this method with respect to the prediction of sentence modality can be found e.g. in Batliner (1988) and Batliner *et al.* (1989).

Fo

The transformation of the Hz values into semitones did not improve the classification results. A possible explanation could be that semitone transformation 'over' normalizes the different voice **ranges** of male and female

speakers (cf. Batliner *et al.* 1989). A normalization of the voice **register** by subtracting a reference value for either the speaker or the utterance resulted in significant improvements in the prediction. In the final analyses we used semitone values and subtracted the basic value of the speaker, i.e. the lowest Fo value produced by the speaker. The transformed maximal and minimal values for the 2nd and 3rd phrase are called max_2, max_3, min_2, and min_3.

The relative position of the maximal and minimal values on the time axis for the 2nd and 3rd phrase are called pos_2 and pos_3. These values are positive, if the minimal value comes later than the maximal value; they are negative, if it is the other way round.

Duration

Best prediction was achieved after a normalization of the speaking rate that took into consideration average duration of that phrase for each speaker (*avduri*) and the average duration of the syllables in the utterance (*dur/number of syllables*):

$$\frac{dur}{avduri} \cdot \frac{duri}{dur/number\ of\ syllables}$$

The transformed duration values for the 2nd and 3rd phrase are called dur_2 and dur_3. We tested several other formulas. The results did not differ much—as long as the actual duration value was put into relation to some reasonable reference value.

Intensity

The best results were achieved with the maximal energy in the 0–5000 Hz band. Average values, 'sonorant' energy sub-bands, and normalizations with respect to the average energy level of the utterance, or with respect to the different intrinsic energy values of the vowels, produced worse results. The intensity values for the 2nd and 3rd phrase are called int_2 and int_3.

Discarded transformations

Declination

The phenomenon of declination—the lowering of the Fo curve along the time axis—is well known. Often accents are described as excursions from this

(hypothetical base-) line. In that case, a Fo peak later in the utterance must not have the same excursion height as an earlier peak to indicate an accent and/or the *Fa*.

It could be possible for our material as well to base the analysis not on (properly transformed) absolute parameter values but on values that are put into relation to a falling declination line. We computed therefore both an **abstract** ('neutral') speaker-specific declination line based on *NQ*s with an 'unmarked' declination and a **concrete** declination line for each utterance as an all point regression line. The prediction of the *Fa* based on these values was inconsistent and generally not as good as a prediction based on the values described in the previous section. The reason might be that our computation of the declination line is not the best one. Anyway, there seems to be virtually no agreement on adequate computation (cf. Lieberman 1986; Lieberman *et al.* 1985; t'Hart 1986; Ladd 1984; and Batliner 1989b: 72). In our opinion, a declination line is therefore still rather an object of investigation than an appropriate reference parameter. (In any case, the discriminant analysis takes into consideration the effect of declination because it is based on the **distribution** of the parameter values and not only on the **absolute** values.)

Comparison ratios

The Fo values of the 2nd and 3rd phrase can be put into the analysis separately, or they can be combined into comparison ratios; cf. Taylor & Wales (1987): for the two phrases that could be accented in their Australian English material, they computed three different comparison ratios:

Division ratio = a/u.
Subtraction ratio = a − u.
Michaelson Contrast ratio = (a − u)/(a + u).
(a = accented, u = unaccented).

In a multivariate regression analysis, they obtained much better results with the contrast ratio than with the two other ratios; the average values of R^2 ('explained variance') are:

contrast ratio	0.85
subtraction ratio:	0.15
division ratio:	0.29

Unfortunately, Taylor & Wales have not done any analyses with the raw data that could be compared with our data. We computed comparison ratios for our variables as well and put them into regression and discriminant analyses; our results can be summarized as follows:

(i) The contrast ratio was not better than the two other ratios.
(ii) The comparison ratios were not better than the absolute values.

We cannot explain the huge differences between the results of Taylor & Wales and our results in (i); as a consequence, we did not work with comparison ratios, but with the separate parameter values of the 2nd and the 3rd phrase. (Again, the extra information contained in the comparison ratios are taken into consideration by the discriminant analysis because it is based on the **joint** distribution of the predictor variables of the 2nd and the 3rd phrase.)

Results

In Figures 1 and 2, per cent correct classifications are displayed if only one variable is used as predictor variable in the (univariate) discriminant analysis. On the abscissa, the different variables are plotted; on the ordinate, the per cent correct classifications. For the *Q/NQ*-classification, duration and intensity are not included, because they always produced results near chance level. For the *Fa* classification, not *off* and *reg* were used, but duration and intensity.

For *Q/NQ* (Figure 1), most of the variables are relevant, the most relevant ones being *off*, max_3, and then *reg* and min_3. (Of course, most of these variables are more or less correlated with each other; cf. Batliner 1989a: 37 ff.). If one tries to predict the *Fa* and does not separate *Q*s and *NQ*s (*FaAll* in Figure 2), the results are not very convincing; a separation of *Q*s and *NQ*s yields better results. The most relevant variables are max_3 and dur_3 for *NQ*s, and max_2 and pos_2 for *Q*s.

Besides *l* – *t* (learn – test), multivariate analyses with two further learn and test constellations were conducted (Figure 3):

(i) Learning sample: 5 speakers; test sample: 1 speaker (simulation of speaker independence: *l5t1*). This is the most relevant constellation for a speaker-independent automatic speech understanding system.

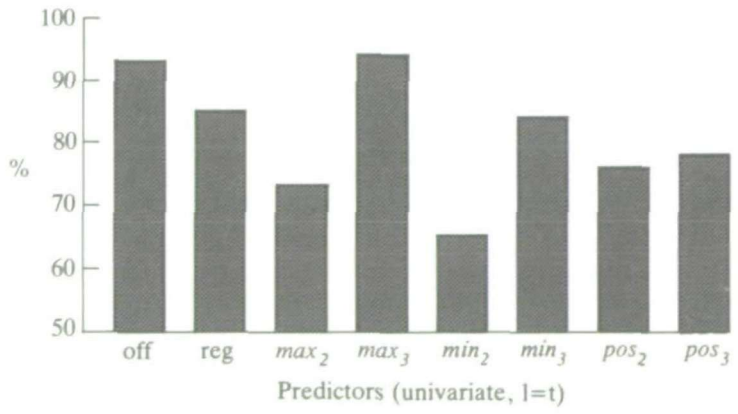

Figure 1 Per cent correct classifications: questions/non-questions

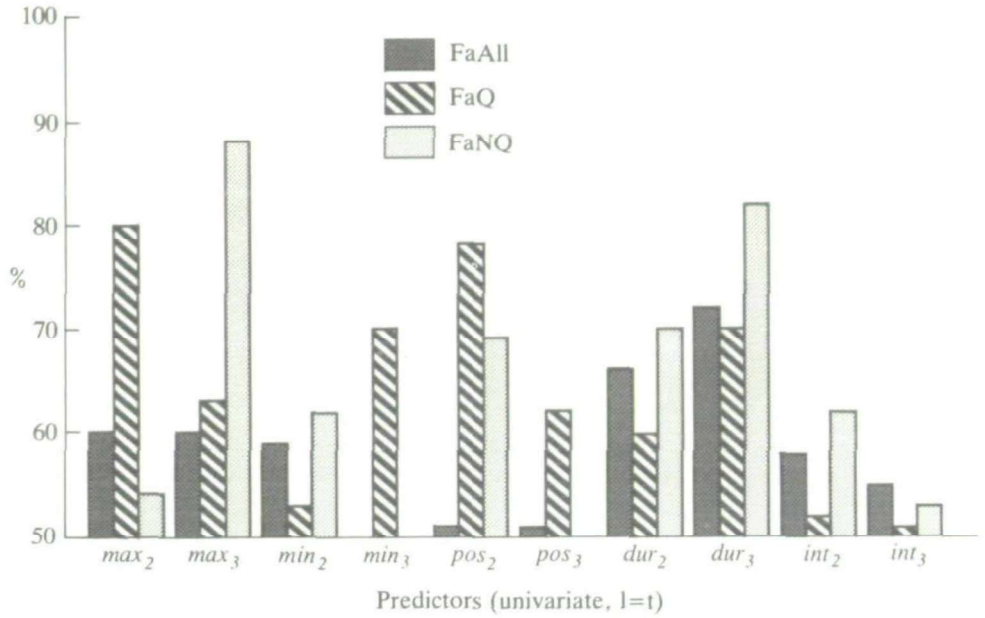

Figure 2 Per cent correct classifications: *Fa*

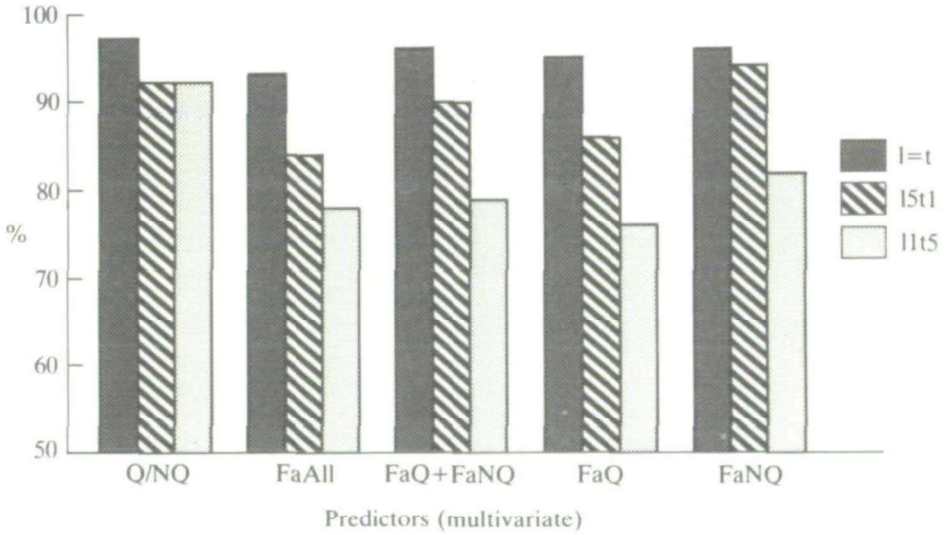

Figure 3 Per cent correct classifications

(ii) Learning sample: 1 speaker; test sample: 5 speakers (generalization from a single speaker to the other speakers: *l*1 *t* 5).

All the univariate discriminant analyses were done with $l - t$. If we look at the corresponding multivariate analysis (all the variables are put at the same time into the analysis; $l - t$ in Figure 3), the classification is very good (always well above 90 per cent), best for *Q*/*NQ*; as for the *Fa*, the separation of *Q*s and *NQ*s

(*FaQ* and *FaNQ* and the weighted mean of these two groups *FaQ* + *FaNQ* in Figure 3) produces better results than an analysis with no separation of *Q*s and *NQ*s (*FaAll*), especially for *l5t1*.

Figure 4 shows the correlation of the predictors with the discriminant function in a multivariate analysis for $l - t$. The greater the correlation, the more relevant is the predictor. For the impact of the predictor on the assignment of the *Fa*, the signs are irrelevant. *Ceteris paribus*, a positive value indicates rather *Fa* on the 2nd phrase, and a negative value rather *Fa* on the 3rd phrase. (In our case, this procedure is more appropriate than the discriminant function of the predictors, as some of the variables are correlated with each other; cf. Klecka 1980: 33 f.). The different relevancy of e.g. max_2, max_3, min_3, pos_2, and pos_3 for *Q*s and *NQ*s shows up clearly.

Generally, the results indicate that in *Q*s, other intonational parameters are used to mark the *Fa* or the same parameters are used in a different way than in *NQ*s. The prediction is worse if *Q*s and *NQ*s are analysed together than if they are treated separately.

Fa is classified better in *NQ*s than in *Q*s. The explanation might be that in *Q*s the same parameters are used to indicate sentence modality as well as place of *Fa*; cf. especially the variable height of the Fo offset. There are therefore more degrees of freedom in *Q*s and consequently more possible confusions.

The results under *FaQ* and *FaNQ* were achieved with a grouping into *Q*s and *NQ*s 'by hand'. For $l - t$ the grouping of the *Q*/*NQ*-classifier was used as an input to the *FaQ*- and *FaNQ*-classifier as well. The classification errors of the first step even improved the results (cf. the error analysis below).

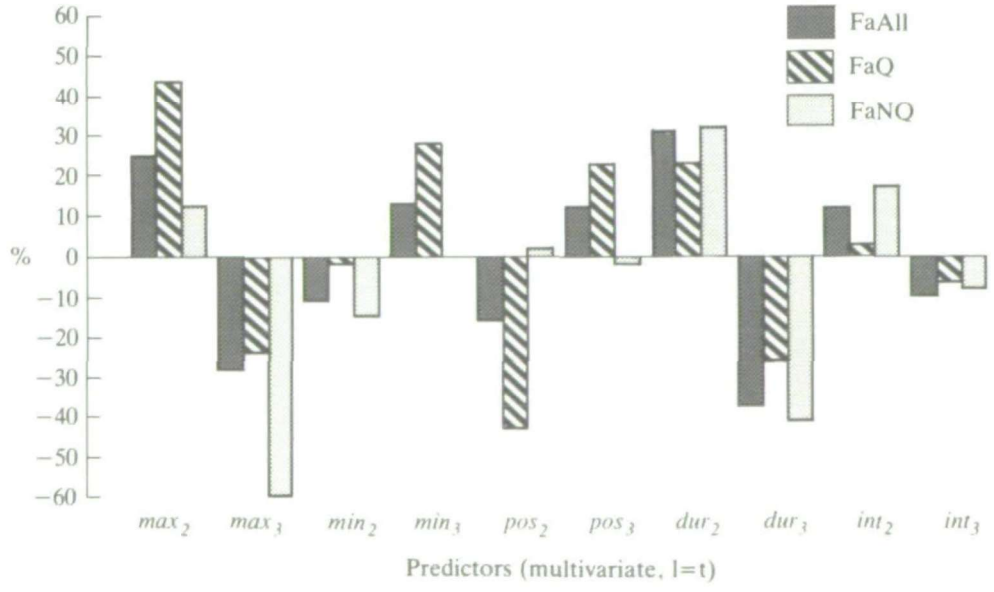

Figure 4 Correlation: predictors with discriminant function

CENTRAL AND MARGINAL TYPES

We will now show the two converging strategies (cf. the first section) as to how to find the central types:

(i) Each of the 4 *Q/NQ – Fa* constelllations has one central type that is characterized by the average values of the predictors.

(ii) We inspected those cases where a strong agreement among the listeners could be observed: practically all the listeners agreed upon the intended *Q/NQ* grouping, the place of the *Fa*, and the naturalness of the production ($MOD \geq 80$ for *Q*s and $MOD \leq 20$ for *NQ*s, $|FOK| = 1$, $NAT \leq 2$). Twenty-four out of the 360 cases passed these strict criteria. Nineteen cases could be identified as representatives of the central types.

For the four central types, Figures 5–8 show the average feature values as well as the Fo contour of a typical production (four out of the nineteen cases): the dashed vertical line marks the border between the 2nd and the 3rd phrase of the actual production. For the 2nd and 3rd phrase, each of the filled squares shows averages for max_2, min_2, max_3 and min_3. The position on the abscissa corresponds to the average position on the time axis in centiseconds starting from the beginning of the utterance; the position on the ordinate corresponds to the average Fo values in semitones above the speaker-specific basic value (st_{bas}). On the top of each figure average beginning point and duration of the 2nd and 3rd phrases is displayed. In the following characterization, the terms 'High',

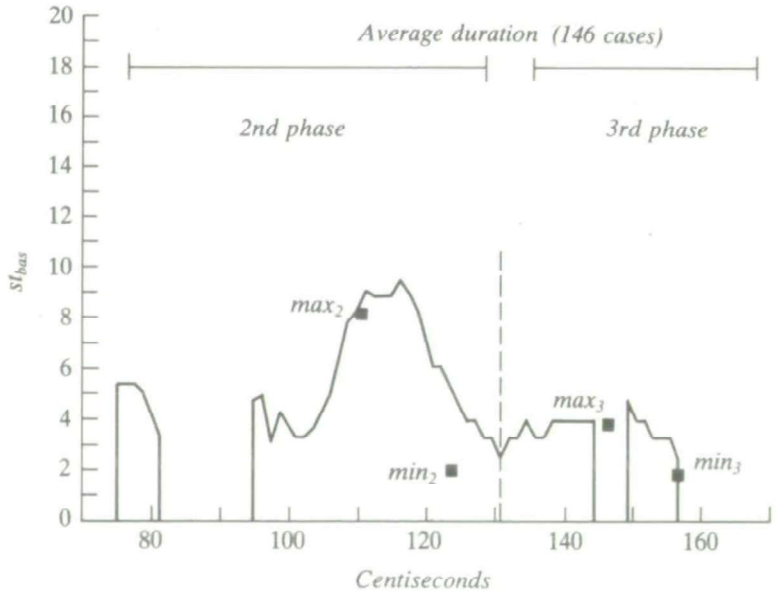

Figure 5 Focus on 2nd phrase, non-question, central type

'Low', and 'boundary tone' (cf. the tone sequence model, e.g. in Pierrehumbert 1980) are used interchangeably with the terms 'rising'/'falling' contour.

(1) *Focus on 2nd phrase, non-question* (Figure 5): the contour is falling in both phrases (High Low). Max_2 is markedly higher than max_3; min_2 and min_3 do not differ.

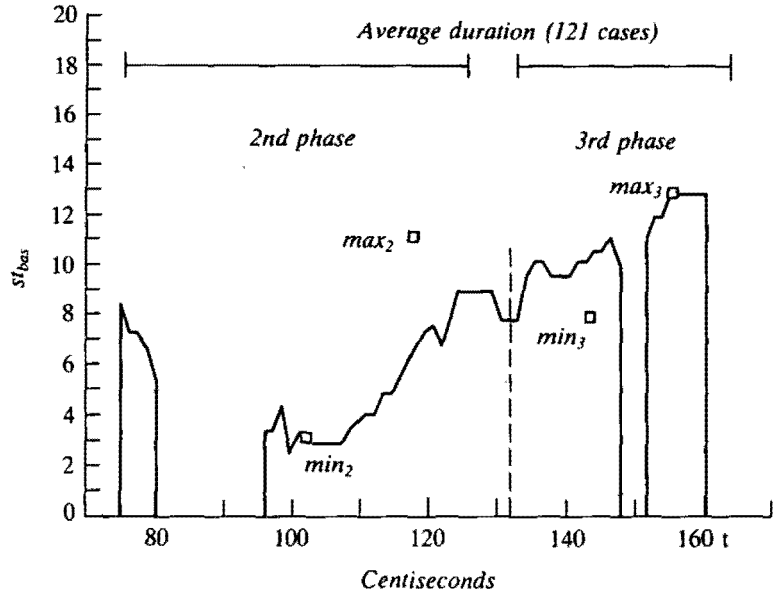

Figure 6 Focus on 2nd phrase, question, central type

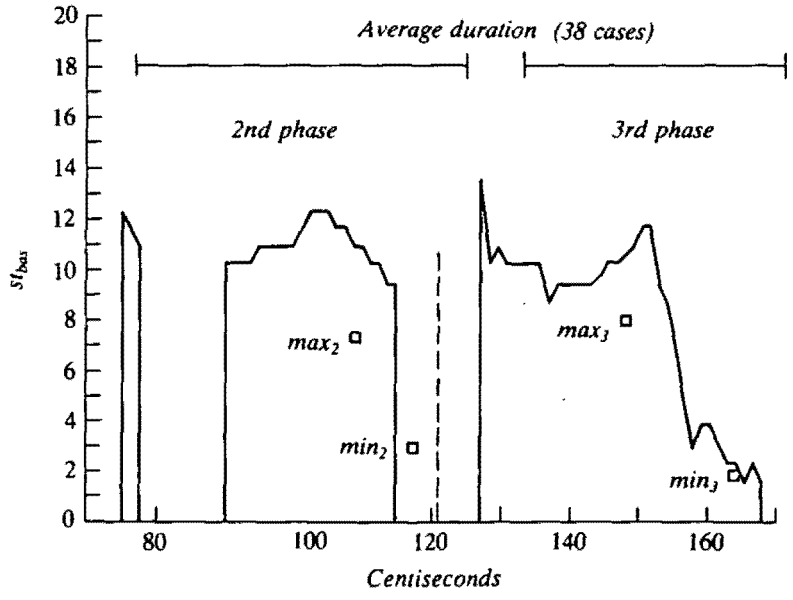

Figure 7 Focus on 3rd phrase, non-question, central type

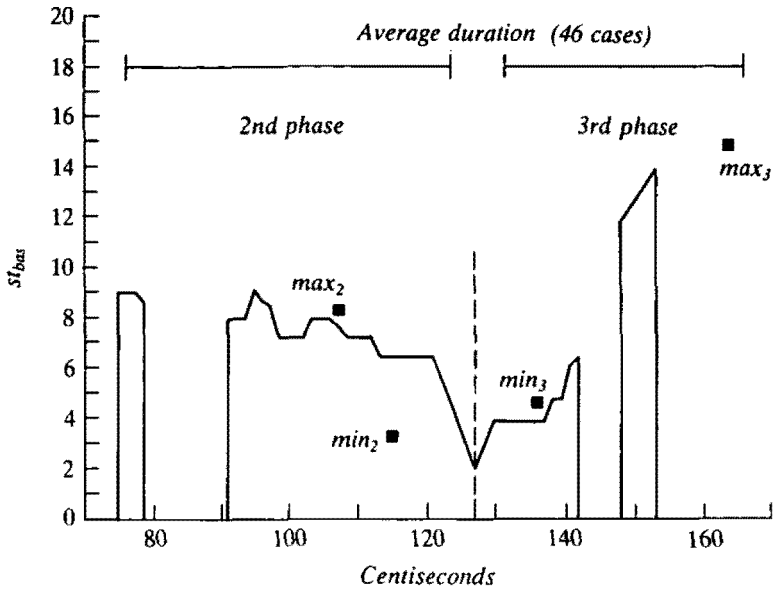

Figure 8 Focus on 3rd phrase, question, central type

(2) *Focus on 3rd phrase, non-question* (Figure 7): the contour is again falling in both phrases (High Low). Max_3 is about as high as max_2; min_2 and min_3 do not differ.

Comparing the two types, we can say that the absolute values for the features of the 2nd phrase in Figures 5 and 7 do not differ remarkably. It is rather the relative values of the features in comparison with the respective values of the 3rd phrase that marks the *Fa*.

(3) *Focus on 2nd phrase, question* (Figure 6): the contour is rising in both phrases (Low High).
(4) *Focus on 3rd phrase, question* (Figure 8): in the 2nd phrase, this type has a falling contour comparable to the *NQ*s, whereas in the 3rd phrase, the contour is rising (Low High)

Comparing these two types, we can say that the Fo range of the phrase with the *Fa* is markedly greater than that of the other phrase. In the final phrase, a rising contour (high boundary tone) is used for both types to mark sentence modality.

The remaining five cases can be grouped into three marginal types which are displayed in Figures 9–11. To demonstrate the deviations from the central types, the respective average values are projected into the contours of the marginal types:

(1) One speaker typically marked *Fa* in prefinal position with a falling contour (High Low), even in *Q*s. If one looks at the average feature values for all speakers and for this specific speaker, one could say that this marginal type across speakers is a central type for this speaker (Figure 9).

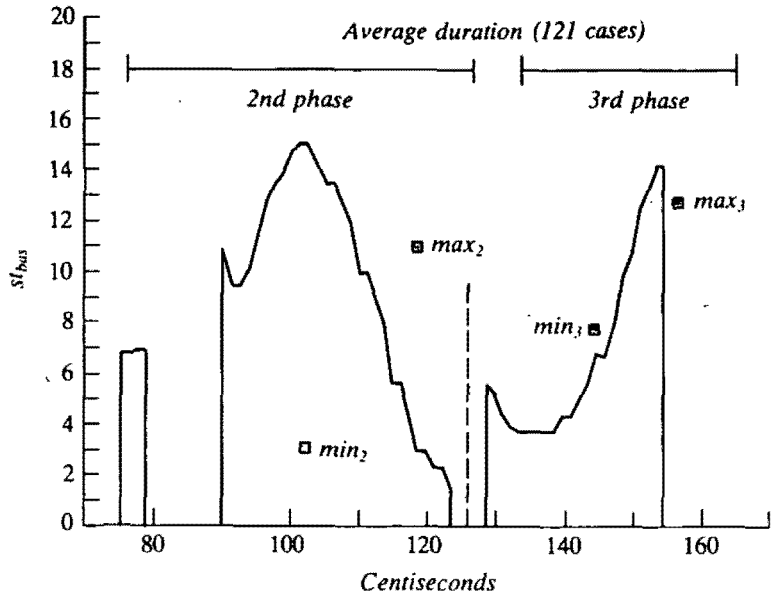

Figure 9 Focus on 2nd phrase, question, marginal type

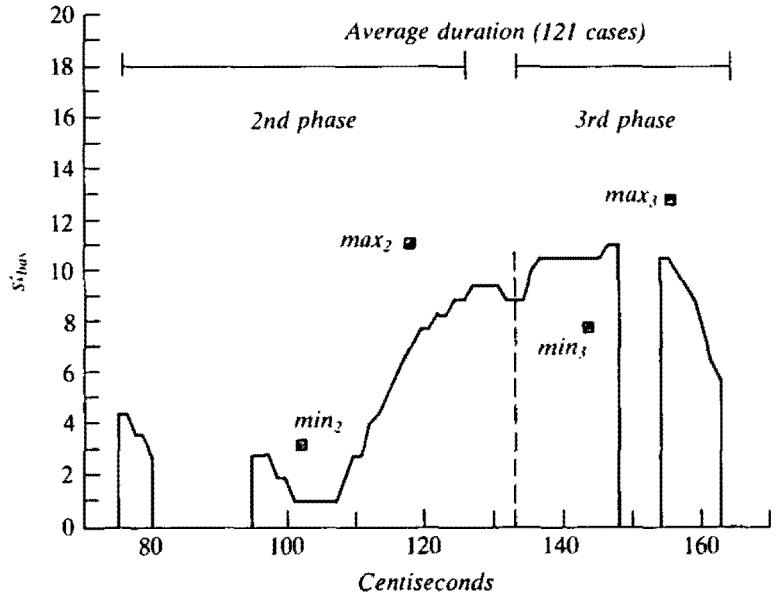

Figure 10 Focus on 2nd phrase, question, marginal type

(2) Another speaker typically marked Qs only in the phrase with the *Fa*; i.e. with *Fa* in prefinal position, the final phrase showed a falling contour comparable to *NQ*s (Figure 10).

(3) The last marginal type, an *NQ* with *Fa* on 3rd phrase, could approximately be described as a 'hat-contour' (cf. Cohen & t'Hart 1967), i.e. a concatention of the two Fo-peaks on the 2nd and 3rd phrase and a low Fo-value at the end of the utterance (Figure 11).

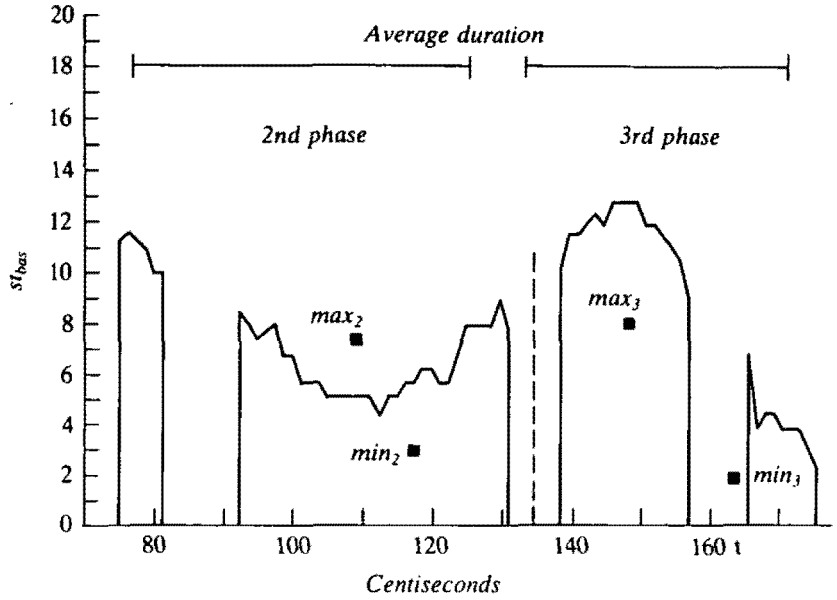

Figure 11 Focus on 3rd phrase, non-question, marginal type

ERROR ANALYSIS

For $l = t$, there are 27 misclassifications, 10 for *Q/NQ* and 19 for *Fa* (i.e. two double misclassifications).

Question/non-question

In all the 10 cases, *Q*s are misclassified as *NQ*s. Eight cases are clearly misproductions, as they are not classified as *Q*s in the perception experiment (cf. (ii) in the section 'Perception experiments' above), and 9 got very low *NAT*-scores, i.e. they were judged as unnatural productions as well.

This also explains the fact mentioned above that the classification errors of *Q/NQ* improved the results of the *Fa* classification: the items under question were misproduced as *NQ*s, and the position of the *Fa* could therefore be classified correctly because the *Fa* had the intonational shape of an *NQ*.

Focal accent

In all but 2 cases, there are indications that the *Fa*-assignment in production and/or perception is not clear-cut: in 11 cases, there is **no agreement** between the results of the perception experiment *FOK* (cf. (iii) in 'Perception experiments' above) and another experiment, where listeners only had to decide upon the place of the sentence accent (cf. Batliner, 1989a: 30, 65 ff.) In 12 cases,

there is a very **weak agreement** between the subjects in the perception experiment ($|FOK| < .4$). In 7 cases, the probability of group membership in the discriminant analysis is near 50 per cent (possibly because of a violation of a necessary assumption).

To sum up the error analysis with respect to the placement of the *Fa*: this is not an easy task (cf. e.g. Lieberman 1965 and Lickey & Waibel 1985), neither for the native speaker/listener nor for the discriminant analysis. We are playing safe when we conclude that the misclassifications did not occur because our statistical model was inadequate, but because of the inherent difficulty of placing the *Fa*.

SPEAKER-SPECIFIC USE OF THE VARIABLES

In Figures 5–11, we have seen that the production of the four different *Q/NQ-Fa* constellations is not uniform across speakers. In Figure 3, it can be seen that *l1 t5*, i.e. the generalization of one speaker to the other five speakers, yields considerably worse results for the prediction of focus than *l5t1* (not for the prediction of *Q/NQ*, by the way). It is therefore very likely that different speakers use the predictor variables in a different way. This fact is illustrated in Figures 12–14, where the correlation between each predictor variable and the discriminant function are plotted for each speaker (*S1–S6*) separately. The higher the correlation, the more important is the variable; the signs are irrelevant. A positive value indicates rather *Q* (Figure 12) or *Fa* on the 2nd phrase (Figures 13, 14), and a negative value rather *NQ* (Figure 12) or *Fa* on the 3rd phrase (Figures 13, 14). If the bars had roughly the same height, all the speakers would use the parameter under consideration in the same way. Of course, a certain variability is normal; some of the differences might as well be traced back to automatic (physiological) processes or to co-variation with another variable. A clear-cut difference, however, can indicate an active process: the speaker uses different parameters or the same parameters in a different way.

A more detailed discussion of the speaker-specific use of the parameters can be found in Batliner (1989a: 55 ff.). We will just mention some of the most striking differences:

(i) For *S2–S6*, *off* is very relevant for the marking of *Q/NQ*, but not for *S1* (Figure 12). *S1* produces *Q*s with *Fa* on the 2nd phrase regularly with a falling contour (cf. the marginal type in Figure 10).

(ii) For the *Fa* assignment in *Q*s, pos_3 is much more important for *S1* than for the other speakers (Figure 13). In that case, pos_3 co-varies with the height of the offset, cf. Figure 10.

(iii) *S6* uses pos_2 for the *Fa* assignment in *NQ*s, but not *S1–S5*, cf. Figure 14 and the 'hat-contour' in figure 11 that was produced by *S6*.

(iv) Duration and intensity are used differently by the speakers, cf. Figures 13 and 14. Although these differences might be caused by automatic processes to a certain extent, we will show in the following that these two parameters contribute to the marking of focus in their own way.

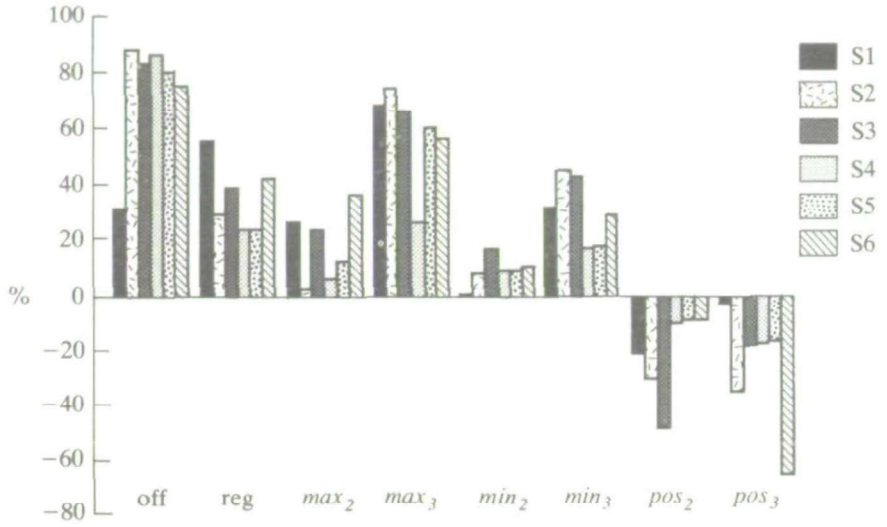

Figure 12 Intra-speaker correlations: predictors with discriminant function: *Q/NQ*

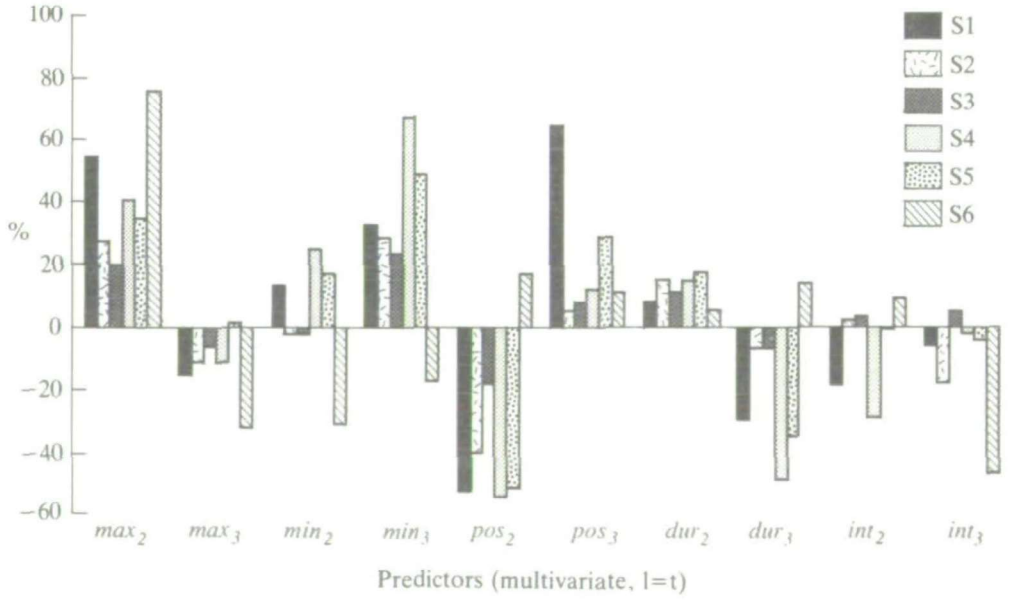

Figure 13 Intra-speaker correlations: predictors with discriminant function: *FaQ*

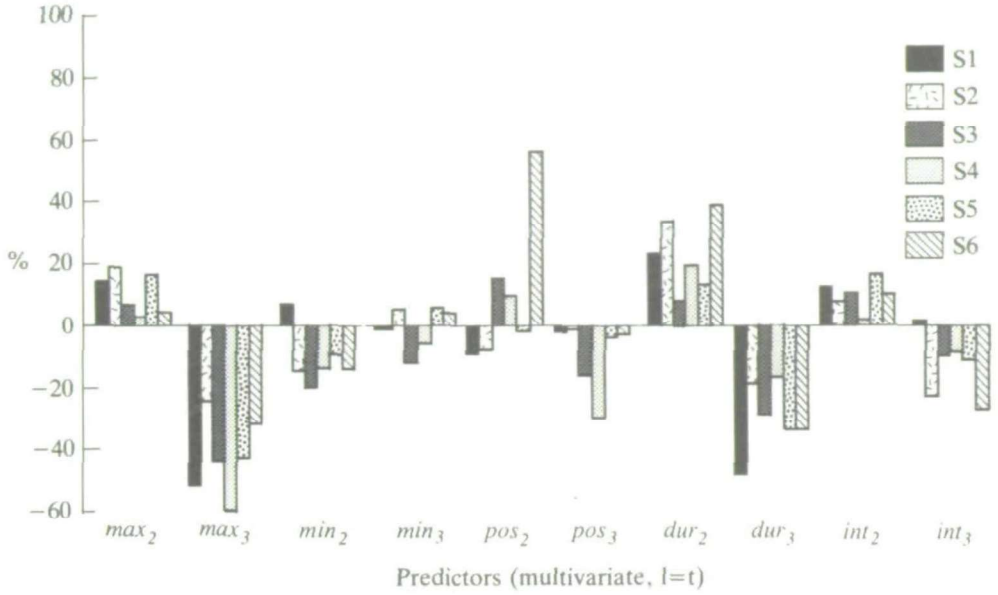

Figure 14 Intra-speaker correlations: predictors with discriminant function: *FaNQ*

WHICH VARIABLES ARE THE RELEVANT ONES?

Coming back to the title of this paper, 'Deciding upon the relevancy . . .', it turned out that some transformations of the variables considerably improved the classification. We have not shown yet whether some variables might be irrelevant—candidates are of course intensity and/or duration. In Figure 15, per cent correct classification are plotted for $l = t$ and $l5t1$, if we—stepwise—exclude (i) intensity, (ii) intensity and duration, and (iii) intensity, duration and

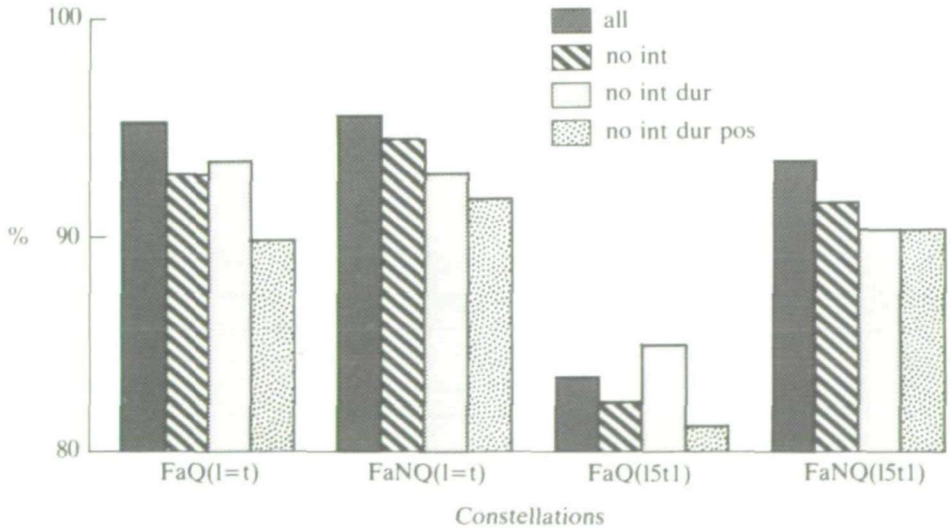

Figure 15 Per cent correct classifications ($l5t1$)

position. It can be seen that the prediction gets worse. (The only exception of this step function is *FaQ* with no intensity and no duration. The reason might be that int_2, int_3, and dur_2 are rather irrelevant for Qs, cf. Figure 2.) In this range, a difference of 2 per cent for example—about 7 cases out of 360—is not a small difference if one considers the (informal) '80/20-rule': that it costs 20 per cent expenses to get 80 per cent of the results, but for the remaining 20 per cent one needs 80 per cent expenses.

Note that, generally, the classification gets worse if an additional and irrelevant predictor variable is put into the analysis. In our case, the classification gets better if more variables are added—therefore, duration and intensity might be of minor importance but they cannot be irrelevant. In other words, if only a tonal model is used that does not take into consideration these two parameters, quite a lot about the placement of the *Fa* can be said, but it is not exactly the whole story of the marking of focus by intonational means.

CONCLUSION

The purpose of this study was to find out how focus is marked intonationally in German. We have shown that all three intonational parameters are used for this task (in order of importance: Fo, duration, and intensity). Speaker-specific or utterance-specific transformations of the features improved their relevancy. Using two different approaches, a statistical and a 'psychological' one (average values and perception experiments), we arrived at central (mostly used) and marginal (rare but acceptable) types. The results indicate that the focal accent is marked differently in questions and non-questions. Speaker-specific ways to use the intonational parameters for the marking of focus were observed. Generally, the focus could be predicted with a high probability (up to 96 per cent), depending on the chosen constellation and/or transformation.

Acknowledgements

This research was financed by the *Deutsche Forschungsgemeinschaft* (DFG). It was carried out in close cooperation with E. Nöth (University of Erlangen). Parts of this paper were published in Batliner & Nöth (1989); a more detailed presentation can be found in Batliner (1989a); cf. also Batliner *et al.* (1990).

ANTON BATLINER
Institut für Deutsche Philologie
Universität München
Schellingstr. 3
8 München 40
FRG

REFERENCES

Altmann, H. (1987), 'Zur Problematik der Konstitution von Satzmodi als Formtypen', in J. Meibauer (ed.) *Satzmodus zwischen Grammatik und Pragmatik*, Niemeyer, Tübingen, 22–56.

Altmann, H. (ed.) (1988), *Intonationsforschungen*, Niemeyer, Tübingen.

Altmann, H., A. Batliner & W. Oppenrieder (eds) (1989), *Zur Intonation von Modus und Fokus im Deutschen*, Niemeyer, Tübingen.

Batliner, A. (1988), 'Produktion und Prädiktion: Die Rolle intonatorischer und anderer Merkmale bei der Bestimmung des Satzmodus', in H. Altmann (ed.), 207–21.

Batliner, A. (1989a). 'Fokus, Modus und die große Zahl: Zur intonatorischen Indizierung des Fokus im Deutschen, in Altmann, Batliner & Oppenrieder (eds), 21–70.

Batliner, A. (1989b), 'Fokus, Deklination und Wendepunkt', in Altmann, Batliner & Oppenrieder (eds), 71–85.

Batiner, A. & E. Nöth (1989), 'The prediction of focus', Proceedings of the European Conference on Speech Communication and Technology, Paris, 26–28 September 1989, 210–13.

Batliner, A., E. Nöth, R. Lang, & G. Stallwitz, (1989), 'Zur Klassifikation von Fragen und Nicht-Fragen anhand intonatorischer Merkmale', *Fortschritte der Akustik–DAGA '89*, Bad Honnef: DPG-GmbH, 335–8.

Batliner, A. & W. Oppenrieder (1989), 'Korpora und Auswertung', in Altmann, Batliner & Oppenrieder (eds), 281–331.

Batliner, A., W. Oppenrieder, E. Nöth & G. Stallwitz (1990), '"Neue Information" im Sprachsignal. Die prosodische Markierung der Fokusstruktur', *Fortschritte der Akustik–DAGA '90* DPG GmbH, Bad Honnef, 1059–62.

Cohen, A. & J. 't Hart (1967), 'On the anatomy of intonation, *Lingua* **19**: 177–92.

't Hart, J. (1986), 'Declination has not been defeated: a reply to Lieberman *et al*', *J. Acoust. Soc. Am*, **80**: 1838–40.

Klecka, W. R. (1980), 'Discriminant analysis', Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–019, Sage Publications, Beverly Hills and London.

Ladd, D. R. (1984), 'Declination: a review and some hypotheses', in J. C. Ewen & J. M. Anderson (eds), *Phonology Yearbook*, **1**, 53–74.

Lickey, S. A. & A. Waibel (1985), 'Perceptual stress assignments', in A. Waibel (1986), *Prosody and Speech Recognition*, Carnegie Mellon University, Computer Science Department, 192–8.

Lieberman, P. (1965), 'On the acoustic basis of the perception of intonation by linguists', *Word*, **21**: 40–54.

Lieberman, P. (1986), 'Alice in declination-land: a reply to Johan 't Hart', *J. Acoust. Soc. Am.*, **80**: 1840–2.

Lieberman, P., W. Katz., A. Jongman, R. Zimmerman & M. Miller (1985), 'Measures of the sentence intonation of red and spontaneous speech in American English', *J. Acoust. Soc. Am.*, **77**, 649–57.

Norusis, M. J. (1986), *SPSSPC+Advanced Statistics*, Chicago: SPSS Inc.

Nöth, E. (1991), 'Prosodische Information in der automatischen Spracherkennung-Berechnung und Anwendung, Niemeyer, Tübingen.

Oppenrieder, W. (1988), 'Intonation and Identifikation: Kategorisierungstests zur kontextfreien Identifikation von Satzmodi', in H. Altmann (ed.), 153–67.

Oppenrieder, W. (1989), 'Fokus, Fokusprojektion und ihre intonatorische Kennzeichnung', in Altmann, Batliner & Oppenrieder (eds), 267–80.

Pierrehumbert, J. B. (1980), 'The phonology and phonetics of English intonation', Ph.D. Dissertation, MIT.

Taylor, S. & R. Wales (1987), 'Primitive mechanisms of accent perception', *Journal of Phonetics*, **15**: 235–46.