GRADIENT PERCEPTION OF INTONATION[1]

Caroline Féry and Ruben Stoel, Potsdam

Abstract

This paper addresses the question of the association of pragmatic meanings with intonational contours from the point of view of perception. The central domain of investigation is the graded acceptability of tonal patterns in contexts for which they were not originally intended. If pragmatic meanings and tunes are matching each other unambiguously, it is expected that categorical results in judgments about wellformedness of tonal contours are the rule. But in fact, acceptability judgments about tonal contours are found to be gradient. Even though speakers agree on the location of accents and on the place of boundaries, the implementation of these in terms of the realization of tones (pitch accents, phrase accents, boundary tones) may be subject to variation. Speakers are sensitive to the topic-focus structure of utterances as opposed to a single focus one. Narrow vs wide focus, which may tragger a pitch accent on the same word are also distinguished from each other, but to a lesser extent.

1. Background

Prosody plays a crucial role in communication. First, we partition our utterances in prosodic chunks, like phonological phrases and intonation phrases, which correspond to sense units (Selkirk 1984), syntactic constituents (Nespor & Vogel 1986, Truckenbrodt 1999) or information structural blocks (Vallduví 1992). These phrases, which help both speaker and hearer to structure the discourse, are signalled phonetically by boundary tones, segmental lengthening or some other phonological cues. A second factor playing a role in the phonological pattern is the distribution and form of pitch accents, associated with prominent syllables. A syllable may be prominent if it is the bearer of the lexical stress of a word or of a larger consitituent which is itself prominent. Prominence, as a consequence of focus or topic,

---

is assigned on discourse structural grounds, but the primary reason for focusing or topicalizing is not purely linguistic. A speaker may decide to speak about some object in her surrounding or an object she knows about, and decide to focus on one property of this object. Or she may answer a question asked by a protagonist because she feels she has to deliver some bit of information. In other words, prominence may be assigned to some linguistic constituents because of contextual or cognitive reasons (Bolinger 1972). The other reason to assign a pitch accent to a syllable is purely grammatical. An internal argument of a German predicate + argument complex, for example, may receive a pitch accent, and the verb may be unaccented. Still the whole phrase may be prominent (see Bierwisch 1968, Schmerling 1976, Selkirk 1984, 1995, von Stechow & Uhmann 1986, Cinque 1993 among others).

A pitch accent, indicated with an asterisk, following Pierrehumbert (1980), may be realized in different ways. In Standard German, nuclear accents (last accents of an intonation phrase) are bitonally falling, HL, or rising, LH, whereas prenuclear accents are also rising or falling or monotonally high (H) or low (L) (see Féry 1993, Grabe 1998, Grice, Baumann and Benzmüller 2003 for phonological studies of intonation of standard German). Prosodic phrases may be terminated with a boundary tone, which is written with a subscripted $_P$ for a phonological phrase, and a subscripted $_I$ for an intonation phrase (following Hayes & Lahiri's 1991 notation). For the sake of illustration, two pitch tracks of a sentence used in the experiments described below are shown with their phonological tone structure.
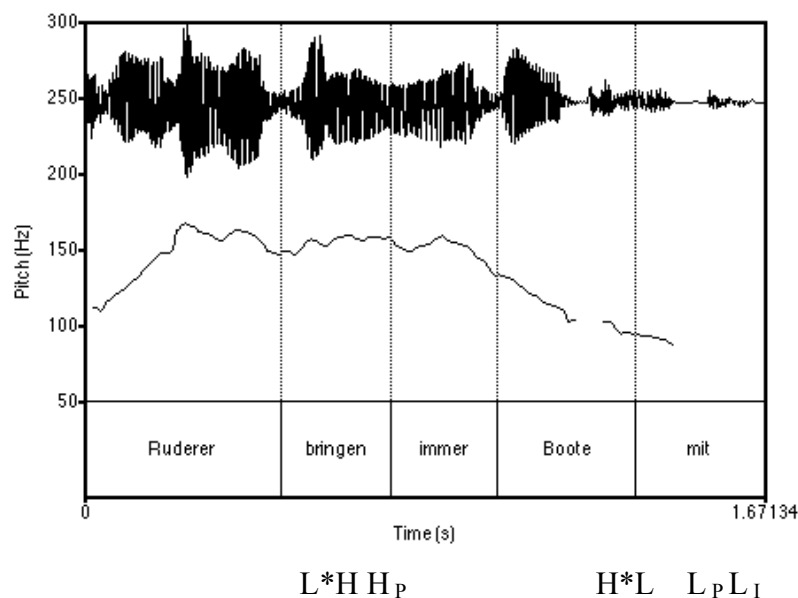

L*H H$_P$ H*L   L$_P$L$_I$

Fig.1 Pitch track of *RUDERER bringen immer BOOTE mit*

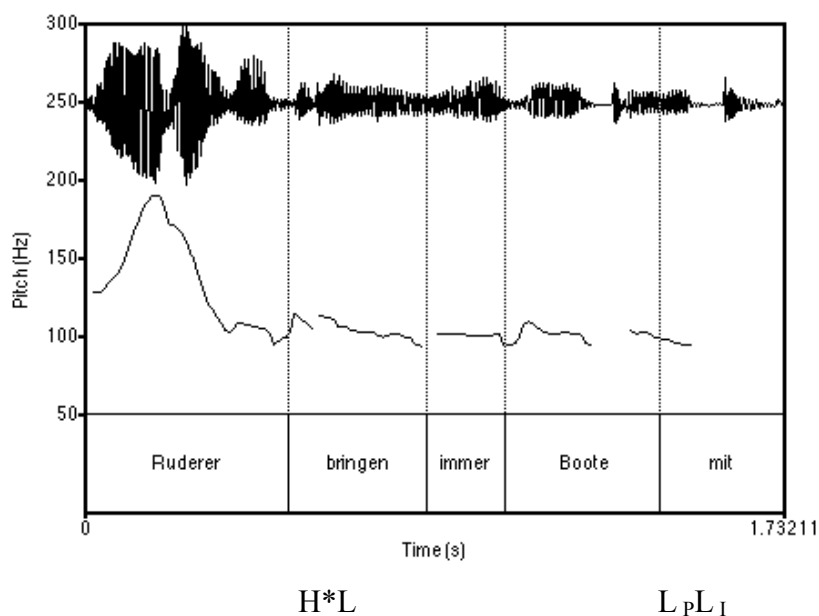H*L   L $_P$L $_I$

Fig.2 Pitch track of *RUDERER bringen immer Boote mit*

A number of studies have addressed the perception of tonal structures, both from the point of view of the location of pitch accents in different pragmatic contexts and of their phonetic realization. The first issue, the location of pitch accents and their role for the focus structure, has been investigated for English by Gussenhoven (1983), Birch & Clifton (1995), Jannedy (2003) and Welby (2003) among others, all of whom examine the role of prenuclear accents on the verb in a VP consisting of a verb plus an argument or an adjunct in English. Gussenhoven's (1983) sentence accent assignment rules SAAR predict that in a focused predicate argument complex, only the argument needs to be stressed, but that a prenuclear accent can be added freely on a verb without impairing processing. In a verbal phrase, by contrast, both the verb and the adjunct need to be stressed. Gussenhoven himself finds full confirmation of this prediction in experimental work. In mini-dialogues such as in (1), there is a difference between the focus structure of (1a) and (1b). In (1a), the whole VP *share a flat* is focused, whereas in (1b) only the direct object is focused, the difference being elicited by the preceding question. The same kind of contrast is obtained in the dialogues in (2) which contain a verb followed by an adjunct.

(1) Verb and argument

    a.   C: Do you live by yourself?

          U: I [share a flat]$_F$ (the whole VP is focused)

3

b. C: I hate sharing things, don't you?

U: I share [a flat]_F   (the argument NP is focused)

(2) Verb and adjunct

a. C: Where will you be in January?

U: We will be [skiing in Scotland]_F   (the whole VP is focused)

b. C: Where will you be skiing?

U: We will be skiing [in Scotland]_F   (the adjunct PP is focused)

In a task which consisted in deciding which of two answers is the most appropriate as a response to a preceding question, Gussenhoven presented the two accent structures, spoken by naive speakers. He cross-spliced questions and answers, so as to obtain both answers in both contexts, and presented the dialogues to other English speakers. He found that the presence of an accent on the verb in addition to the expected accent on the object in (1) does not change the acceptability of the pitch accent structure, and that in both contexts. The speakers did not do better than by chance when required to choose between the two contexts on the basis of such an accent pattern. But in (2), the absence of a stress on the verb in (2a) was an indicator that the verb had to be given (and thus not focused), so that the speakers did better than in the predicate-argument condition in the same task. Gussenhoven also tested a third group of question-answer-pairs in which the verb had a narrow focus. These obtained still a better score.

Birch & Clifton (1995) conducted similar experiments, but obtained slightly different results. They also prepared adequate and less adequate pairs of questions and answers. An example of a dialogue set is reproduced in (3). Only the pairs QA/R1 and QB/R3 are perfectly matching, all others are predicted to be more or less deviant along the same lines as those just explained.

(3)

a. Questions   QA: Isn't Kerry pretty smart?

QB: Isn't Kerry good at math?

b. Responses   R1: Yes, she TEACHES MATH.

R2: Yes, she teaches MATH.

R3: Yes, she TEACHES math.

4

And indeed, Birch & Clifton found in judgement and decision tasks that as an answer to question QA, speakers prefer R1, with two accents, over R2, with just one accent on the argument NP. The difference was small but significant. And unsurprisingly, R3 was by far the preferred answer to QB. All other pairs obtained poorer scores. In a second experiment, speakers had to decide of how well the pairs made sense. In this case, the results for QA were similar to those of Gussenhoven: There was no difference between a sentence with two accents (R1) and a sentence with just one accent on the argument (R2).

These results speak in favour of a theory of accent assignment in which prenuclear pitch accents - but not postnuclear ones - can be assigned more or less freely, though Birch & Clifton's results were even more gradient than those of Gussenhoven.

Perception experiments bearing on the location of pitch accents conducted for Dutch, (Nooteboom & Kruyt 1987, Krahmer & Swerts 2001), and for German (Hruska, Alter, Steinhauer & Friederici 2001) show for these languages as well, that a prenuclear accent is readily acceptable, but that a postnuclear one is out of question and that accents on narrowly focused items in an otherwise nonnuclear position are more readily perceived than accents on words accented per default in their unmarked accent pattern. Nooteboom & Kruyt (1987) rightly explain the acceptability of a prenuclear accent in terms of topicalizing or thematizing the bearer of such an accent, and observe that a sentence with a supplementary prenuclear accent can get an interpretation in which the prenuclear accent is information structurally prominent.

A second group of experimental studies examine the perception of tonal structures in relation to the form or excursion of the pitch accents in different pragmatic contexts. Krahmer & Swerts (2001) were interested in the question whether contrastive accents are phonologically different from other accents in terms of shape and prominence. Comparing contrastive accents with normal nuclear accents in Dutch, they did not find any difference in shape, if one excepts the fact that contrastive accents are nuclear accents and that nuclear accents differ from prenuclear accents, but they found a difference in prominence, although this difference can only be perceived in context - and not on isolated words - confirming the fact that accent is a relational notion that can only be appreciated with comparison with syllables which are more or less accented than the syllable under consideration.

Jannedy (2002) looked at the well-formedness of four tonal patterns in contexts requiring different focus structures. The method used in her study consisted in letting the speakers hear one of four contours in a sentence like (4), with a subject, a predicate and an object. The speakers had to decide which of the following questions suits best the contour.

(4)   *Benjamin heard shots*

    1) What happened?

    2) Who heard shots?

    3) What did Benjamin hear?

    4) Who heard what?

The target sentence was preceded by a short description (one or two sentences) providing a context for the target sentence, after which the sentence itself was uttered in one of the four tonal contours: a unique pitch accent on the subject (H*L L$_I$), a hat contour with a pitch accent on the subject and another pitch accent on the object (H* H*L L$_I$), a rising pitch accent on the object preceded by a prenuclear pitch accent on the subject (H* LH* L$_I$), and a double peak on the subject and the object (H*L L$_P$ H*L L$_I$). Since the context and the wording were identical, if speakers picked up different questions, the difference had to be due to the tonal contour. Jannedy's results showed some variation and gradience. As expected, a single early accent was interpreted as narrow focus on the subject, and a late rising tone was mostly interpreted as a late focus. But the other contours, the hat pattern and the double peak, did not show clear preferences, though an interpretation in term of a broad focus seems to be chosen by most speakers per default. Jannedy interprets some of her results as an effect of the difference in prosodic phrasing, and speculates that it is this difference which speakers reacted to rather than the difference in pitch accents. She assumes that each prosodic phrase is accompanied by exactly one nuclear accent, which means that the hat pattern, the early peak and the late peak are all compatible with a single prosodic phrase, whereas a double peak contour is an indicator of a double prosodic phrase.

The role played by stress and prosodic phrasing for disambiguating syntactic structures is altogether much better understood than the role of the form of the pitch accents, since prosodic correlates in ambiguous syntactic contexts have been addressed and tested in quite a number of papers on language processing. Syntactic phrasing correlates strongly with prosodic phrasing, that is accent location and type, but also, and probably even more, with phonetic boundary cues.

The general hypothesis expressed in most of the papers on the subject is that specific prosodic parsing mechanisms can have a direct impact on syntactic processing (see for instance Beckman 1996, Schafer 1997, Fodor 1998 and Kitagawa & Fodor, this volume). Schafer, Carlson, Clifton & Frazier (2000) show that the presence of a pitch accent on a *wh-*

word introducing an embedded clause facilitates the interpretation of the clause as an interrogative, as in *I asked the pretty little girl who is cold,* as opposed to a relative clause in which the pitch accent is preferably on *cold*. It seems that the distribution of pitch accents can disambiguate the syntactic structure. Similar results have been obtained by Warren, Grabe & Nolan (1995) who were interested in the distinction between early and late closure sentences (whether a noun phrase is attached to the preceding or the following clause), and the way speakers phrase and stress the phonetic material in both cases. In particular, they found that stress shift in expressions like *Hong Kong problems*, with stress on *Hong*, favour a late closure reading, as the one found in a sentence like (5a). In the early closure reading of this expression, as in (5b), no stress shift is expected, because *Hong Kong* is final in its prosodic phrase, and as a result, *Kong* is stressed. In sum, they found a well-founded strong correlation between stress shift and preference for late closure reading, as well as between the absence of stress shift and early closure reading. See Speer et al (2003) as well as Schafer et al (2000) for good summaries of this line of research.

(5)   a. Whenever parliament discusses Hong Kong problems, they are solved instantly.
       b. Whenever parliament discusses Hong Kong, problems are solved instantly.

Turning to studies on German, Bader (1996, 1998) found what he calls 'a covert activation of suprasegmental phonology' in silent reading. He found that a syntactic ambiguity leads to strong garden path effects only if focus particles were present that added a prosodic ambiguity. In his view, garden path effects are thus enhanced if prosodic representations are also affected.

Steinhauer (2000) used for German the method of event-related potentials to show that prosodic cues can determine whether a noun phrase, like *Anna* in (6), is parsed as the object of the first verb (6a) or of the second verb (6b). If a boundary appears after the main verb *verspricht* 'promises', the reading in which *Anna* is the object of *entlasten* 'to support' is favoured, and the normally easy to process sentence (6a), according to both the Late Closure and the Minimal Attachment principles of the garden path model (Frazier, 1978, 1987), becomes more difficult. The resulting severe processing difficulties in (6a) can be viewed as a prosody induced "reversed garden path effect", since sentence (6a) should be usually preferred over (6b).

(6) a. [Peter verspricht Anna] zu arbeiten und das Büro zu putzen

7

Peter promises Anna to work . . . and to clean the office

    b. Peter verspricht [Anna zu entlasten] und das Büro zu putzen

Peter promises to support Anna . . . and to clean the office

The result of this study show that the effect of a strong prosodic indicator, like a pitch accent or a boundary tone, can override the unmarked processing preferences otherwise found in reading experiments. All the works summed up in this section indicate clearly that prosody plays a major role in the processing of spoken and read language.

The experiments reported in the remaining of this paper aim at testing for German intonational structures in particular contexts. The study uses pitch accents and prosodic phrasing, though not in their function of disambiguating a syntactic ambiguity, but rather in their function as bearers of information structure and disambiguators of semantic ambiguities. It is shown that speakers are aware of minimal differences in tonal structure, but that the use speakers make of this awareness is utterly gradient.

2. Experimental design

2.1 Materials

The experiments reported in this paper were so conceived as to test the sensibility of German native speakers for the tonal make-up of sentences in relation with information structure. The informants had to judge the naturalness of spoken utterances in different contexts induced by preliminary sentences. Three experiments were conducted, in which different kinds of target sentences were used. These were either short sentences (section 3), long sentences (section 4), or sentences containing a negation and a quantifier, each having variable scope (section 5). (7) shows an example of a short sentence.

(7) Short sentences

<u>Maler bringen immer Bilder mit.</u>

Painters bring always pictures with

a. Narrow focus on the subject (NFS): Tom told me that photographs always bring pictures to our neighbours. But this is not true.

b. Narrow focus on the object (NFO): It is said that painters always bring books to our neighbour. But this is not true.

c. Topic-focus (TF): My neighbour often organizes big parties, and for this reason she also gets lots of presents. Movie directors give her movies, writers give her books and ...

All other sentences used for the experiments are listed in the appendix. Contexts and stimuli sentences were spoken by a trained speaker in a sound-proof booth and recorded on a DAT recorder (Sony 100). The speaker was instructed to speak naturally, in a normal tempo. He read the context-target pairs at once, first the context and then the stimulus sentence. There were 48 pairs for the three experiments together (18 short sentences, 18 long sentences and 12 quantifier-negation sentences). Among the 48 sentences, only 15 (6, 6, and 3) were truly different, since the other realizations were intonational variants of these 15 sentences. The sentences were evaluated by three independent trained phonologists as to their naturalness, as well as to their adequacy and were found to be natural and adequate. Context sentences and stimulus sentences were digitized into individual sound files, ready to be cross-spliced. No manipulation whatever was undertaken in order to not endanger the naturalness. However, a large number of trials were necessary, as well as a first complete run of the experiment, in which the sentences and their contexts turned out to be unsatisfactory for our aims.

2.2 Procedure

A set-up was conceived in which the subjects were in a quiet room with a presentation using the DMDX experiment generator software developed by K. and J. Forster at the University of Arizona. The experimenter left the subject alone in the room after brief initial instructions as to beginning and ending the session. The subjects worked through the DMDX presentation in a self-paced manner. It lead them through a set of carefully worded instructions, practice utterances, and finally the experiment itself, consisting of 102 target sentences. No fillers were inserted, but three practice sentences started the experiment. This experiment was itself included in a set of experiments in which the subjects performed different tasks: production of read material, and dialogues. The instructions made it clear that the aim of the experiment was to test the intonation and stress structure of the sentences, and not their meaning or syntax. The experiment was under the subjects' control, who had to press the return key in order to start and continue the experiment. The stimuli were presented auditorily only: pairs of context and stimulus sentence were presented sequentially. The subject heard first a context, and after hitting the return key, the test sentence. The task consisted in judging the adequacy of the intonation of the sentence in the given context. Every recorded sentence of the groups of short and long sentences was presented 9 times, in three different intonational

and stress patterns, and each of these patterns in three different contexts. The scope sentences were presented 16 times each. One group of subjects had to judge the sentences on a scale of 1 (very bad) to 8 (perfect). The entire scale was presented on the screen together with every auditory presentation of the stimuli, because it contradicted the school scoring system, where 1 is the best note and 6 the worst. After having made up her mind as to how well or how badly the intonation of the sentence matched the context, the subject entered her judgment into the computer with the appropriate number, and went on to the next stimulus. Another group of subjects had to judge the same sentences, but in a categorical way: j for "yes" meant that the sentences were adequate in the offered context, and n for "no" meant that they were not adequate.

The sentences, together with the scope sentences (discussed in section 4), were presented in a randomized order, different for each subject. The set-up and the instructions included the option of repeating the context-stimulus pair for a particular sentence. Most subjects made occasional use of this possibility.

2.3 Subjects

Two groups of 15 subjects took part in the first experiment. They were native speakers of Standard German, students at the University of Potsdam, and had no known hearing of speech deficit. All were paid or acquired credit points for their participation in the experiment.

3. Experiment 1: Short sentences

3.1 Materials

In the first experiment, there were six short sentences consisting of a simple subject (an animate noun in plural), a verb (*mitbringen* 'bring'), an adverb (*immer* 'always') and a simple object (an inanimate noun in plural). These are shown in (7). The separable but unstressed particle *mit* was located at the end of the sentence, resulting in a non-final object. The sentences were inserted in three different contexts inducing the following information structures: narrow contrastive focus on the subject (NFS), narrow contrastive focus on the object (NFO), and topic-focus (TF). The sentences with narrow focus were elicited by contrasting a pre-mentioned element with another one which it replaced in the following sentence. A topic-focus was always elicited by pre-mentioning some pairs of elements with the same structure as the tested sentence.

3.2 Hypotheses

The hypotheses for the short sentences are based on the experimental literature discussed above where it is shown that the location of the accents, as well as their relative prominence are the crucial factors for judgments of pairs of contexts and target sentences. In German, a prenuclear accent is often rising, but the rise is more important in the topic-focus structure than in the narrow focus on the object, as can be seen from a comparison between Figs. 1 and 3. Fig. 3 displays a narrow focus on the object. The subject *Ruderer* has a pitch accent but its excursion is much smaller than in the topic-focus configuration. The difference in excursion is an indicator of a difference in prosodic phrasing, as Liberman & Pierrehumbert (1984) assumes for English and Büring (1997) and Féry (1993) for German. The TF sentences consist if two phrases with the rising tone on the subject defining both a pitch accent and a boundary tone (H_P), whereas all other realizations consist of only one phrase with one (NFS) or two (NFO) pitch accents.

L*H   H*L $L_P L_I$
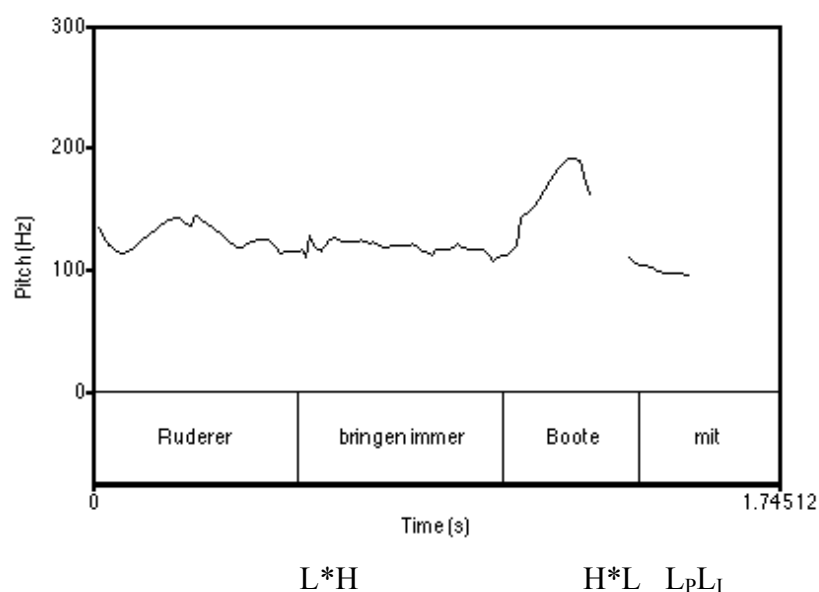
Fig. 3 Pitch track of *Ruderer bringen immer Boote mit*

We propose the following hypotheses (see section 1):

1. Matching pairs: The mean acceptability scores for matching context-intonation pairs should be higher than for non-matching pairs. This follows from the fact that German is a true intonation language in which pitch accents are the most important cues for information structure. More specifically, we expect that, in a NFS context, NFS intonation is more acceptable than both NFO intonation and TF intonation. Likewise, in a NFO context, NFO

11

intonation should be more acceptable than NFS and TF intonation, and, in a TF context, TF intonation should be more acceptable than NFS and NFO intonation.

2. Missing nuclear accent: In a context eliciting an accent on the object (i.e. both in NFO and TF context), sentences with NFS intonation should obtain lower scores than sentences with NFO or TF intonation, because in this case the expected late nuclear accent is missing.

3. Added nuclear accent: In a context asking for an early accent (i.e. a NFS context), sentences realized with a late nuclear accent (i.e. NFO or TF intonation) are worse than sentences without an added accent (i.e. NFS intonation).

4. Missing prenuclear accent: In a TF context, sentences with just one accent on the object (i.e. NFO intonation) should get intermediate scores, i.e. higher than sentences with NFS intonation, but lower than sentences with TF intonation.

5. Added prenuclear accent: In a context in which just a nuclear accent on the object is elicited (i.e. NFO context), sentences with an accent on both the object and the subject (i.e. TF intonation) should get intermediate scores, i.e. higher than sentences with NFS intonation, but lower than sentences with NFO intonation.

3.3 Results

As mentioned above, there were two groups of 15 subjects each. One group judged the sentences on a scale of 1 (very bad) to 8 (perfect). In table 1, the data for these subjects consist of the mean score of the six sentences having the same context-intonation pair. The second group of subjects judged the same sentences in a categorical way. For these subjects, table 2 gives the number of sentences (out of 6) that were considered adequate for each context-intonation pair. The correlation between the mean scores in table 1 and 2 is almost perfect (Pearson's product-moment correlation = 0.984, p = 0.000). The interaction between context and intonation is displayed graphically in figure 4. It presents the results of only the first group (i.e. scale answers), but a graph of the second group would look very similar due to the strong association between the two groups.

Table 1. Short sentences: mean judgements scores (on a scale from 1 to 8).

context / intonation	NFS	NFO	TF
NFS	7.7	1.5	2.0
NFO	2.0	7.2	5.9
TF	2.0	3.7	6.8

Table 2. Short sentences: mean number of sentences judged adequate (out of 6).

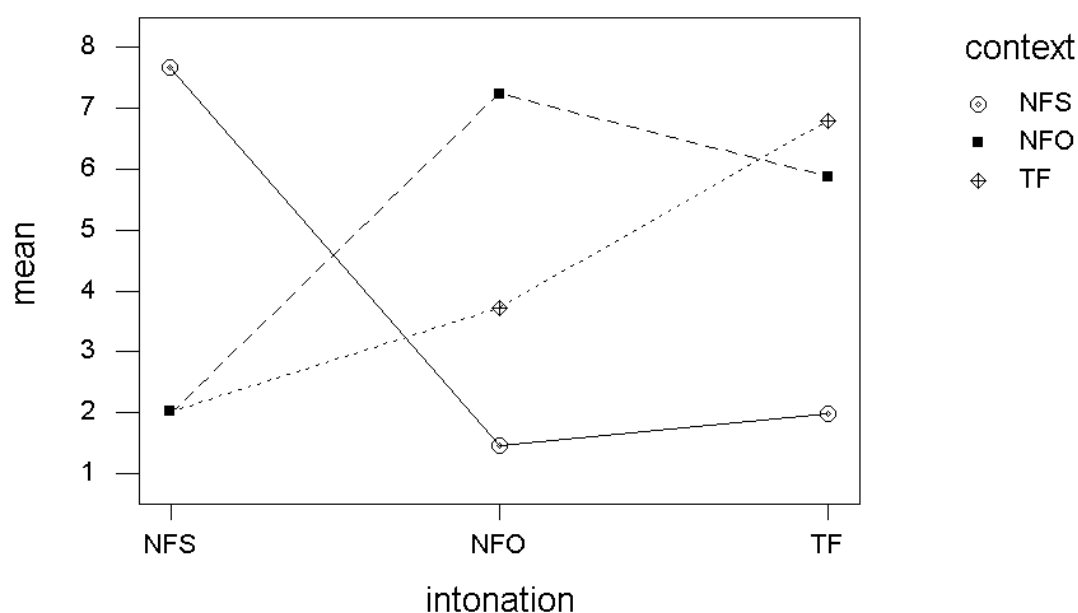| context / intonation	NFS	NFO	TF
NFS	5.5	1.1	0.7
NFO	1.3	5.3	3.9
TF	0.4	1.9	5.2

Figure 4. Mean acceptability scores for short sentences (scale anwers).

The five hypotheses discussed in section 2.4 were translated into the 8 statistical hypotheses presented in table 3. These include: 1. Matching pairs (nr. 1, 2, 3, 4, 5, and 6), 2. Missing nuclear accent (nr. 3, 5, 7, and 8), 3. Added nuclear accent (nr. 1 and 2), 4. Missing prenuclear accent (nr. 6 and 8), and 5. Added prenuclear accent (nr. 4 and 7). Table 3 mentions context before intonation, e.g. NFS-NFO is a sentence presented in a NFS context with a NFO intonation pattern. Paired t-tests were used to make the planned comparisons. After applying the Bonferroni adjustment, all differences in means were found to be significant for both groups of subjects, except for the difference in means between NFO-NFO and NFO-TF sentences in case of the first group.

13

Table 3. Planned comparisons for short sentences (context-intonation pairs).

nr.	hypothesis	scale answers			yes-no answers		
		t	p[1]		t	p[1]	
1	NFS-NFS > NFS-NFO	25.1818	0	*	7.8649	0	*
2	NFS-NFS > NFS-TF	14.3525	0	*	9.8052	0	*
3	NFO-NFO > NFO-NFS	8.6676	0	*	8.3666	0	*
4	NFO-NFO > NFO-TF	2.2659	0.0199		2.8807	0.0060	*
5	TF-TF > TF-NFS	9.5092	0	*	9.9983	0	*
6	TF-TF > TF-NFO	10.4457	0	*	8.5114	0	*
7	NFO-NFS < NFO-TF	-6.4770	0	*	-4.1710	0.0005	*
8	TF-NFS < TF-NFO	-3.3765	0.0023	*	-3.6173	0.0014	*

[1]df = 14; * = significant at the 0.05 level after correcting for multiple comparisons (required significance level: 0.05 / 8 = 0.00625)

3.4 Discussion

All predictions were borne out. First, the scores for matching context-intonation pairs were higher than for non-matching pairs (hypothesis 1). Second, a missing nuclear accent (hypothesis 2) and an added nuclear accent (hypothesis 3) triggered lower scores than sentences with the expected accentuation. The same was true for both a missing prenuclear accent (hypothesis 4) and an added prenuclear accent (hypothesis 5).

Gradient judgments were obtained in two different ways, either directly, by letting the informants give their own gradient results, or indirectly, by counting categorical results. The very high correlation between the two groups of means suggests that it does not matter which method is used, as both methods give very similar results.

4. Experiment 2: Long sentences

4.1 Materials

The second experiment that was conducted was very similar to the first experiment, except that the target sentences were six longer sentences in which the subject and the object were more complex. The subjects denoted a group of travellers with a certain destination, the verb

nehmen 'take' was accompanied by the adverb *meistens* 'most of the time' and a prepositional object denoting a way of transportation (article + adjective + noun). These sentences were inserted in the same information structural contexts as the short sentences.

(8) Long sentences

<u>Passagiere nach Rom nehmen meistens den späten Flug.</u>[2]

Passengers to Rome take most the late flight

a. Narrow focus on the subject (NFS): It is said that the people to Athen mostly take the late flight, but this is not true.

a. Narrow focus on the object (NFO): Mona says that passengers to Rome mostly take the early flight, but this is not true.

c. Topic-focus (TF): Commuters who work far away from home have often similar habits. Business people who go to Paris often take their car, travellers to London take the train from Calais and ...

Besides their length, the long sentences had a more complex syntactic structure than the short ones. The subject always had the structure 'N to place' (*Passagiere nach Rom* 'passengers to Rome', *Schiffe nach Sardinien* 'boats to Sardaigne', etc) and the context elicited contrastive stress on the place name, which is the default position for accent in such a DP. The objects contained an adjective and a noun (*schnellen Zug* 'rapid train', *späten Flug* 'late plane'), and the accent was always on the adjective, the marked position for accent. Contrastive stress on the adjective rendered the focus structure unambiguously narrow. Our decision to use a contrastive narrow focus was driven by the intention to have a very clear accentual structure. The same procedure was used as in the first experiment. The hypotheses were also the same as those formulated for the short sentences.

4.2 Subjects

Two groups of 15 subjects, who had not participated in the first experiment, took part in the experiment. One group of subjects judged the sentences on a scale of 1 to 8, while the other group judged the same sentences in a categorical way. All subjects were students at the University of Potsdam.

---

[2] As Ede Zimmermann (p.c.) observes, it is not undisputed whether there is a structural ambiguity between the temporal and the quantificational reading of *meistens*. We suspect that, even if confirmed, this ambiguity played no role in the experimental results.

4.3 Results

For each of the subjects who rated the sentences on a scale from 1 to 8, a mean score of the six sentences with identical context and intonation was calculated. The mean values of all subjects are given in table 4. For the subjects who gave categorical judgments, the number of sentences (out of 6) that were considered adequate were counted, and the mean values are given in table 5. The correlation between the two groups of mean scores in table 4 and 5 is again almost perfect (Pearson's product-moment correlation = 0.967, p = 0.000). A graph of the interaction between context and intonation for the first group of subjects is displayed in figure 5.

Table 4. Mean values for long sentences (acceptability scores on a scale from 1 to 8).

context / intonation	NFS	NFO	TF
NFS	7.5	1.5	1.7
NFO	1.6	7.4	5.6
TF	3.6	5.1	6.9

Table 5. Mean values for long sentences (sum of 6 categorical acceptability scores).

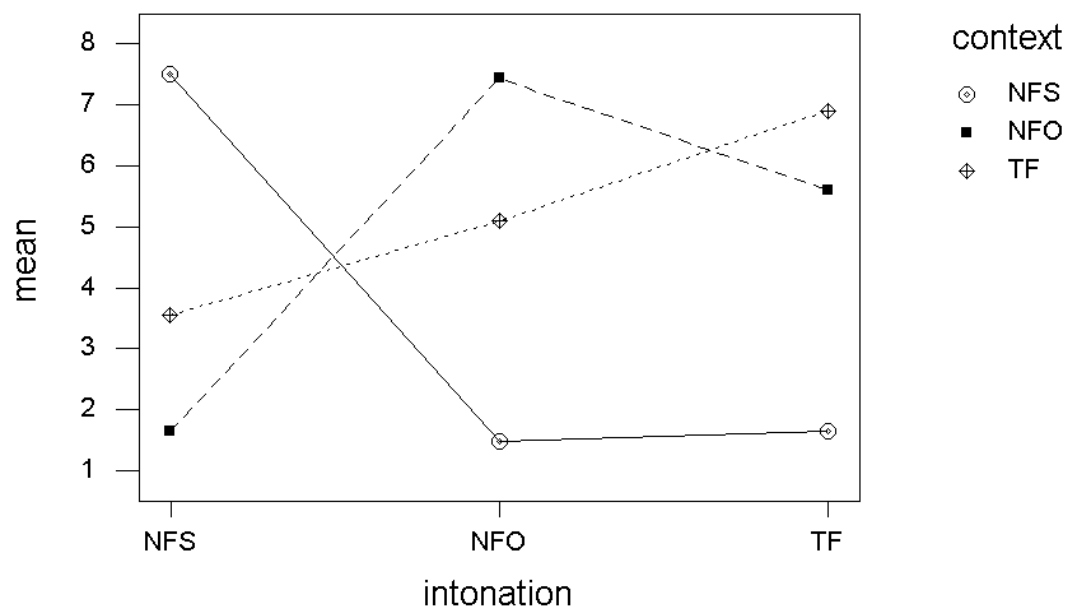| context / intonation	NFS	NFO	TF
NFS	6.0	0.3	0.4
NFO	0.8	5.8	3.4
TF	0.7	2.7	5.3

Figure 5. Mean acceptability scores for long sentences (scale anwers).

The same hypotheses were tested as for the short sentences. Paired t-tests were used to make the planned comparisons. After applying the Bonferroni adjustment, all differences in means were found to be significant for both group of subjects. The results are presented in table 6.

Table 6. Planned comparisons for long sentences (context-intonation pairs).

nr.	hypothesis	scale answers		yes-no answers	
		t	p^{1}	t	p^{1}
1	NFS-NFS > NFS-NFO	17.2602	0 *	35.5581	0 *
2	NFS-NFS > NFS-TF	14.9014	0 *	29.4368	0 *
3	NFO-NFO > NFO-NFS	12.6571	0 *	12.4263	0 *
4	NFO-NFO > NFO-TF	5.4442	0 *	6.0406	0 *
5	TF-TF > TF-NFS	6.7591	0 *	13.4360	0 *
6	TF-TF > TF-NFO	4.3347	0.0003 *	4.6519	0.0002 *
7	NFO-NFS < NFO-TF	-8.7558	0 *	-5.5018	0 *
8	TF-NFS < TF-NFO	-2.9515	0.0053 *	-4.6810	0.0002 *

[1]df = 14; * = significant at the 0.05 level after correcting for multiple comparisons (required significance level: 0.05 / 8 = 0.00625)

4.4 Discussion

Again a very high correlation between the two groups of subjects was found, suggesting that both methods are equally good for obtaining gradient judgments.

All hypotheses were confirmed by the experiment. Thus, matching context-intonation pairs scored better than non-matching pairs, a missing or an added nuclear accent resulted in lower scores than the expected accentuation, and the same was true for a missing or added prenuclear accent.

5. Experiment 3: Scope sentences

5.1 Materials

The sentences in the third experiment consisted of three utterances allowing a variable scope of negation and a variable scope of quantifier. These sentences consist of a subject made up of a quantifier and a noun, an auxiliary, the negation *nicht*, and a past participle or an adjective (below called the predicate). Four contexts were constructed, as illustrated in (9). First a context eliciting two accents: one on the quantifier and one on the negation. The second context elicits a narrow focus on the quantifier, the third context a narrow focus on the negation, and the last context was a topic-focus one, eliciting two accents again.

(9) Ambiguous scope of negation:

Beide Autos sind nicht beschädigt worden.

Both cars were not damaged

1. Two foci ('two'): It would be too bad if Charles would have lost both his Jaguar and his Porsche because of the bad weather, but fortunately it was not the case.

2. Narrow focus on the quantifier (FQ): Is only Peter's car not been damaged? No, ...

3. Narrow focus on the negation (FN): I have seen that both your cars have been staying in the garage for ages. Were they damaged at the accident? - No, I already told you, ...

4. Topic-focus (TF): Several things happened at the accident. Three bikes are now ruined, a passenger is at the hospital, but nothing dramatic happened to the involved cars.

5.2 Subjects and procedure

Four groups of 15 subjects (all students at the University of Potsdam) took part in the experiment. The first two groups judged the sentences on a scale of 1 (very bad) to 8

(perfect), while the third and fourth group judged the same sentences in a categorical way. In addition to the scope sentences, the first and the third group also judged the short sentences of experiment 1, while the second and fourth group judged the long sentences of experiment 2.

5.3 Hypotheses

The syntactic structure of the sentences in this experiment is simple, but their semantic structure is not. The negation can have scope over the quantifier or vice-versa. In the experiment, one context called unambiguously for wide scope of the negation (not Q), and one unambiguously for wide scope of the quantifier (for Q, it is not the case that…). The first case is triggered by a double accentuation on the quantifier and the negation ('two' context in 9), and the second by a single accent on the quantifier (FQ context in 9).[3] It is assumed here that the scope inversion reading (not Q), elicited by the 'two' context can be explained by general properties of topicalization, visible in languages with resumptive pronouns. The topicalized quantifier in the sentences under consideration is in a position of extraposition to the left, but is nevertheless interpreted as in the scope of the negation (see also Höhle 1991). All authors who have studied the scope inversion phenomenon in German (Höhle 1991, Jacobs 1997, Büring 1997, Krifka 1998) have insisted on the necessity of a rise-fall contour to get the interpretation aimed at, and this is the contour which has been produced by our speaker as well. Crucially, an independent phonological phrase is formed which contains the topicalized constituent, separate from the main clause. In a realization with only one accent on the quantifier, by contrast, the quantifier is interpreted in situ, before the negation, and has consequently wide scope over the negation.[4] Prosodically, the quantifier cannot be interpreted as being topicalized because it has the focal accent of the sentence. In our experiment, the context eliciting this accent pattern was one in which the quantifier was contrastively accented.

The other two patterns, a single accent on the negation (FN) and a double accent on the quantifier and on the predicate (TF) do not evoke clear scopal relationships. A unique accent on the negation contradicts the preceding sentence. In the experimental sentences, the

---

[3] As a generalization, the negation may have wider scope when both the quantifier and the negation (or the negated constituent) are accented. This generalization holds only for this type of constructions, but not for other sentences with inverted scope, like those with two quantifiers discussed in Krifka (1998).

[4] Krifka (1998) explains scope inversion of sentences with two quantifiers by allowing movement of accented constituents at the syntactic component of the grammar. Both

predicate had been stressed in the preceding matching context. However, it was not possible to unambiguously reconstruct the context from the negated sentence only. An accent on the quantifier, the noun or the predicate changes the pragmatics of the sentence, but in the realization with a single accent on the negation, these differences are cancelled. The hypothesis was thus that an accent on the negation would be tolerated in a variety of contexts (hypothesis 3 below), and not only in a verum/falsum reading (Höhle 1992), though the context eliciting a narrow focus on the quantifier (FQ) or the one calling for a topic-focus (TF) intonation, were not entirely adequate for a realization with a single accent on the negation.

The TF context with accents both on the NP containing the quantifier and on the predicate can also have different readings, one being that the predicate is contrasted. Inverted scope is also not impossible in this case.

Our hypotheses can be formulated as follows:

1. Matching pairs should get higher scores on average than non-matching ones.

2. In this case, too, missing or added nuclear accents are expected to get lower scores on average than sentences with proper nuclear accents, and sentences with missing or added prenuclear accents should again get intermediate scores.

3. A unique accent on the negation gets on average higher scores than any other intonation pattern, since the context denied by such a sentence is not unambiguous.

5.4 Results

A mean score of the three sentences of a context-intonation pair was calculated for each subject who rated the sentences on a scale from 1 to 8. The mean values for these subjects are given in table 7, and a graph is displayed in figure 6. Table 8 presents the mean number of sentences (out of 3) that were considered adequate by the subjects who gave categorical judgments. Once again, the correlation between the two groups of means is almost perfect (Pearson's product-moment correlation = 0.973, p = 0.000).

---

topicalized and focused constituents have to be pre-verbal at some stages of the derivation in

Table 7. Mean values for scope sentences (acceptability scores on a scale from 1 to 8).

context / intonation	two	FQ	FN	TF
two	6.1	3.6	5.1	6.1
FQ	3.7	7.0	3.2	3.4
FN	5.4	3.1	6.5	5.3
TF	5.4	3.6	4.7	5.8
all	5.1	4.3	4.9	5.1

Table 8. Mean values for scope sentences (sum of 3 categorical acceptability scores).

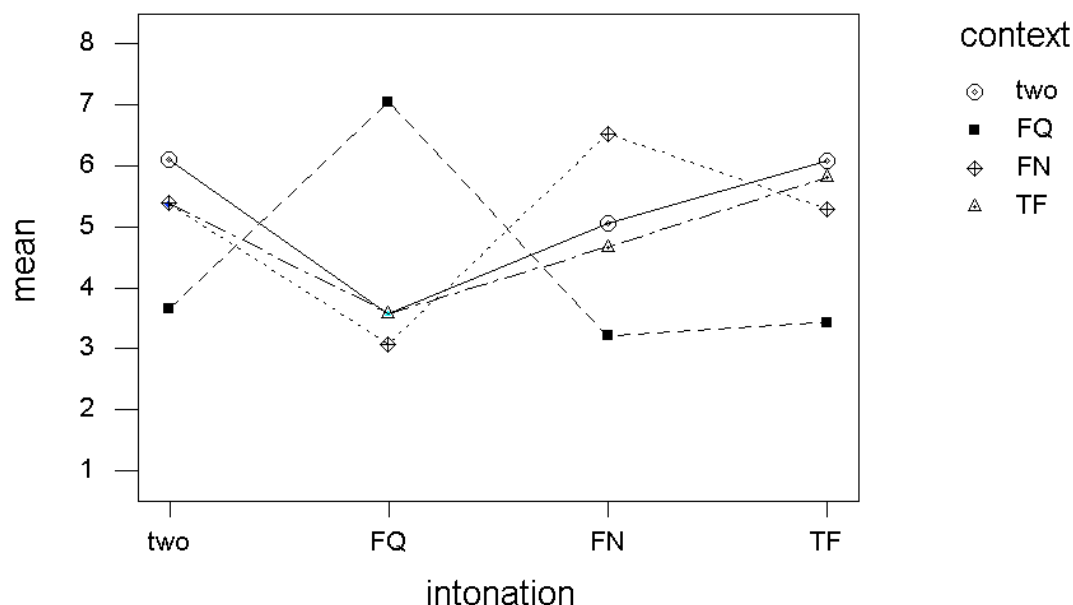| context / intonation	two	FQ	FN	TF
two	2.2	0.8	1.9	2.1
FQ	1.0	2.7	0.8	1.1
FN	1.9	0.5	2.7	1.7
TF	2.3	1.2	1.6	2.2
all	1.8	1.3	1.8	1.8

order to get stress.

Figure 6. Mean acceptability scores for scope sentences (scale answers).

The hypotheses discussed in section 4.3 were translated into 27 statistical hypotheses, which are presented in table 9. Paired t-tests were used to make the planned comparisons. The mean scores of matching pairs (nr. 1 to 12 in table 9) were significantly higher in 10 out of 12 comparisons in case of the scale group, and in 8 out of 12 comparisons for the yes-no group. Missing nuclear accent (nr. 8, 13, and 14) got significantly lower scores in 3 out of 3 comparisons for both groups. An added nuclear accent (nr. 5, 15, and 16) got a significantly lower scores in only 1 out of 3 comparisons in case of both groups. Missing prenuclear accents (nr. 17 to 20) and added prenuclear accent (nr. 21 to 24) all had significantly higher scores in 8 out of 8 comparisons for both groups. However, a unique accent on the negation (nr. 25 to 27) did not get higher scores than other intonation patterns.

Table 9. Planned comparisons for scope sentences (context-intonation pairs).

nr.	hypothesis	scale answers		yes-no answers		
		t	p[1]	t	p[1]	
1	two-two > two-FQ	8.0673	0 *	7.3928	0	*
2	two-two > two-FN	4.9107	0 *	1.3516	0.0935	
3	two-two > two-TF	0.0859	0.4661	0.4412	0.3312	
4	FQ-FQ > FQ-two	10.231	0 *	10.4652	0	*

#	Comparison						
5	FQ-FQ > FQ-FN	9.6816	0	*	11.2123	0	*
6	FQ-FQ > FQ-TF	9.8416	0	*	8.6574	0	*
7	FN-FN > FN-two	4.1624	0.0001	*	5.0000	0	*
8	FN-FN > FN-FQ	11.0510	0	*	14.9939	0	*
9	FN-FN > FN-TF	4.6951	0	*	6.9522	0	*
10	TF-TF > TF-two	1.4435	0.0798		false		
11	TF-TF > TF-FQ	5.5603	0	*	4.6637	0	*
12	TF-TF > TF-FN	3.3860	0.0010	*	3.0016	0.0027	
13	FN-FQ < FN-two	-7.1812	0	*	-8.2605	0	*
14	FN-FQ < FN-TF	-6.9657	0	*	-7.0000	0	*
15	FQ-FN < FQ-two	-2.1456	0.0202		-1.1395	0.1319	
16	FQ-FN < FQ-TF	-0.8536	0.2002		-1.7951	0.0415	
17	two-FN > FN-FQ	5.1973	0	*	8.2260	0	*
18	two-FN > FQ-FN	5.5223	0	*	6.4839	0	*
19	TF-FN > FN-FQ	4.1986	0.0001	*	7.1313	0	*
20	TF-FN > FQ-FN	4.1766	0.0001	*	6.1166	0	*
21	FN-two > FN-FQ	7.1812	0	*	8.2605	0	*
22	FN-two > FQ-FN	6.2430	0	*	9.1044	0	*
23	FN-TF > FN-FQ	6.9657	0	*	7.0000	0	*
24	FN-TF > FQ-FN	5.4904	0	*	6.1764	0	*
25	all-FN > all-two [2]	false			false		
26	all-FN > all-FQ [2]	1.8278	0.0350		2.7741	0.0032	
27	all-FN > all-TF [2]	false			false		

[1]df = 29; * = significant at the 0.05 level after correcting for multiple comparisons (required significance level: 0.05 / 27 = 0.00185)

[2]df = 119

5.5 Discussion

The results are more ambiguous than in case of the short and long sentences, as the scores are less extreme. First, hypothesis 1 is confirmed in only 75% of the comparisons that were made. Also, an added nuclear accent was not as bad as expected, and it was simply false that a unique accent on the negation would do better on average than other intonation patterns.

The more diffused results in this type of sentences can perhaps be explained by the greater cognitive cost that these sentences need in order to be processed by the speakers. That this is indeed the case is shown in the next section.

The second explanation for the more gradient results comes from the fact that some context-sentence pairs which were not originally intended to be well-formed were nevertheless possible under a slightly different interpretation of the context. This was explained in 4.3 with the observation that a stressed negation may negate different contexts. But it is also true in other cases as well. The 'two' and TF pairs were also felt to be interchangeable, to cite just one additional example.

A remarkable result is that in all three experiments the scores obtained for the two groups of subjects (scale and yes-no answers) were very similar. In other words, the same gradient results can be obtained by using either gradient or not-gradient judgments. This is remarkable since the cognitive task executed in both groups is very different. It could have been the case that in a sentence with a high score of acceptability the rating by scale would have been gradient, but the yes-no judgment categorical. However, if the groups of informants is large enough, 'intolerant' subjects compensate the degree of insecurity that remains in subjects asked to give a judgment on a scale.

6. Reaction times

Additional information on the cognitive cost of the task was gathered by the measure of reaction times. Table 10 shows that it took more time to process the long sentences than the short ones, and the scope sentences took more time than the long sentences. The reaction times not only increase progressively for the kind of sentences, but it can also be observed that taking a decision on a scale needs more time than taking a categorical decision.

Table 10. Mean reaction times in seconds

	Short sentences	Long sentences	Scope sentences
Scale	4.2 s	5.4 s	5.6 s
Yes-no	3.7 s	4.9 s	5.4 s

7. Conclusion

After an overview of the relevant literature, this paper has presented results of an experiment on the perception of intonation patterns in German. Three kinds of sentences elicited in different information structural contexts were cross-spliced and informants were asked to judge the acceptability of context-target pairs. First, the short sentences elicited the clearest results, and confirmed for German that prenuclear accents can be more freely deleted and added than nuclear accents without endangering the acceptability of the tonal contour. Second, this was confirmed for the longer sentences, as well, though the added complexity of the syntactic structure was responsible for the slightly less straightforward results. Finally, the scope sentences delivered results which were more difficult to interpret. None of the sentence type was syntactically ambiguous, but the sentences with variable scope were ambiguous as to the scope relationships between quantifier and negation. The increasingly gradient results were explained by a clear increase in complexity from the short sentences to the sentences with scope ambiguities. This increasing complexity was triggered by the different factors playing a role in the processing of the sentences: accents, boundaries, quantification, negation. Because more factors entered the well-formedness judgment of these sentences, there was an increase in variance and grammaticality judgments were less stable.

The results of the short and long sentences, and, to a less extent, those of the scope sentences, point to a very high correlation between context and prosodic structure. Speakers and hearers do use prosodic information such as presence vs. absence of pitch accents, their form and the phrasing to assess the well-formedness of context-target sentence pairs, and they do so consistently.

Though we offer no analysis of how our gradient data can be accounted for in a formal grammar, we conclude with a remark that a categorical grammar will not do. A gradient grammar, like stochastic OT, which uses overlapping constraints, can account much better for the observed variability. This is, however, the object of a future research.

Appendix

Short sentences (three contexts)
1. Maler bringen immer Bilder mit. 'Painters always bring pictures.'
2. Lehrer bringen immer Hefte mit. 'Teacher always bring notebooks.'
3. Sänger bringen immer Trommeln mit. 'Singers always bring drums.'
4. Ruderer bringen immer Boote mit. 'Oarsmen always bring boats.'

5. Geiger bringen immer Platten mit. 'Violonists always bring records.'

6. Schüler bringen immer Stifte mit. 'Students always bring pens.'

Long sentences (three contexts)

7. Passagiere nach Rom nehmen meistens den späten Flug. 'Passengers to Rome always take the late flight.'

8. Reisende nach Mailand fahren oft mit dem schnellen Bus.

'Travellers to Milan often drive with the express bus.'

9. Autofahrer nach Griechenland nehmen immer den kürzesten Weg.

'Car drivers always take the shortest road.'

10. Schiffe nach Sardinien fahren meistens mit voller Ladung.

'Ships to Sardinia mostly sail with a full cargo.'

11. Züge nach England fahren oft mit rasantem Tempo

'Trains to England often ride with full speed.'

12. Trekker nach Katmandu reisen meistens mit vollem Rucksack.

'Treckers to Katmandou mostly travel with a full backpack.'

Variable scope sentences (four contexts)

13. Alle Generäle sind nicht loyal. 'All generals are not loyal.'

14. Beide Autos sind nicht beschädigt worden. 'Both cars have not been damaged.'

15. Viele Gäste sind nicht gekommen. 'Many guests did not come.'

References

Bader, M. (1996). *Sprachverstehen. Syntax und Prosodie beim Lesen.* Westdeutscher Verlag.

---- (1998). 'Prosodic influences on reading syntactically ambiguous sentences', in J. D. Fodor and F. Ferreira (eds.), *Reanalysis in sentence Processing.* Dordrecht: Kluwer, 1-46.

Beckman, M. (1996). The parsing of prosody. *Language and Cognitive Processes* 11: 17-67.

Bierwisch, M. (1968). Two critical problems in accent rules. *Journal of Linguistics* 4: 173-178.

Birch, S. & Clifton, C. (1995). Focus, Accent, and Argument Structure: Effects on Language Comprehension. *Language and Speech* 38: 365-391.

Bolinger, D. (1972). Accent is predictable (If You're a Mind-Reader). *Language* 48: 633-644.

Büring, D. (1997). *The 49th bridge accent*. Mouton de Gruyter. Berlin.

Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic Inquiry* 24: 239-297.

Féry, C. (1993). *German intonational patterns*. Niemeyer, Tübingen.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research* 27: 285-319.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation. University of Connecticut.

---- (1987) Syntactic Processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5: 519-559.

Grabe, E. (1998). *Comparative intonational phonology: English and German*. Doctoral dissertation, Universiteit Nijmegen.

Grice, M., Baumann, S., and Benzmüller, R. (2003). German Intonation in Autosegmental Phonology. In Jun, S.-A. (ed.) *Prosodic typology*. OUP.

Gussenhoven, C. (1983). Testing the reality of focus domains. *Language and Speech* 26. 61-80.

Hayes, B. and Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory* 9. 47-96.

Höhle, T. (1991). On reconstruction and coordination. In: Haider, H. & K. Netter (eds.) *Representation and derivation in the theory of grammar*. Dordrecht: Reidel.

---- (1992). Über Verum-Fokus im Deutschen. In: Jacobs, Joachim (eds.) *Informationsstruktur und Grammatik* (= Linguistische Berichte, Sonderheft 4): 112-141.

Hruska, C., Alter, K., Steinhauer, K. and Steube, A. (2001). Misleading dialogues: Human's brain reaction to prosodic information. In Cave, C., I. Guaitella & S. Santi (Eds.) Orality and Gesture. Interactions et comportements multimodaux dans la communication. Paris. L'Harmattan. 425-430.

Jannedy, S. (2003). *Hat patterns and double peaks: the phonetics and psycholinguistics of broad versus late narrow versus double focus intonations*. Unpublished doctoral dissertation. The Ohio State University.

Jacobs, J. (1997). I-Topikalisierung. *Linguistische Berichte* 168: 91-133.

Krahmer, E. and Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication* 34: 391-405.

Krifka, M. (1998). Scope inversion under the rise-fall contour in German. *Linguistic Inquiry* 29: 75-112.

Liberman, M. and Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In Aronoff, M. & R.T. Oehrle (eds.) *Language sound structure*. Cambridge: MIT Press, 157-233.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nooteboom, S.G. and Kruyt, J.G. (1987): Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America* 82: 1512-1524.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. MIT Diss.

Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. Unpublished doctoral dissertation, Amherst, MA: University of Massachusetts.

Schafer, A. J., Carlson, K., Clifton, C. and Frazier, L. (2000). Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and Speech* 43: 75-105.

Selkirk, E. O. (1984). *Phonology and syntax. The relation between sound and structure*. Cambridge: MIT Press.

---- (1995). Sentence Prosody: Intonation, Stress and Phrasing. In: Goldsmith, John (ed.) *Handbook of Phonological Theory*. Blackwell, 550-569.

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

Speer, S. Warren, P. and Schafer, A. (2003). Intonation and sentence processing. In: *Proceedings of the International Congress of Phonetic Sciences 15*. Barcelona, 95-106.

Stechow, A.v. and Uhmann, S. (1986). Some remarks on focus projection. In Abraham, W. & S. de Meij (eds.) *Topic, focus and configurationality*. Amsterdam, 295-320.

Steinhauer, K. (2000). *Hirnphysiologische Korrelate prosodischer Satzverarbeitung bei gesprochener und geschriebener Sprache*. MPI series in cognitive neuroscience 18.

Truckenbrodt, H. (1999). On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30: 219-255.

Vallduví, E. (1992). *The Informational Component*. New York. Garland.

Warren, P., Grabe, E. and Nolan, F. (1995). Prosody, Phonology and Parsing in Closure Ambiguities. *Language and Cognitive Processes 10*: 457-486.

Welby, P. (2003). *Effects of pitch accent position, type and status on focus projection*. Language and Speech 46. 53-58.