Production and Perception of Thematic Contrast in German

Bettina Braun

Max Planck Institute for Psycholinguistics,
Nijmegen, The Netherlands

Running title: Thematic contrast in German

Mailing address:
Dr Bettina Braun
Max Planck Institute for Psycholinguistics
Postbus 310, 6500 AH Nijmegen
The Netherlands

Email: bettina.braun@mpi.nl

Acknowledgements: This research was supported by the German Research Council (DFG) within the International Research Training Group 'Language Technology and Cognitive Systems' (715). I would particularly like to thank Bob Ladd, Bill Barry, and Mark Steedman for valuable discussion on the design and evaluation of the experiments, as well as Anne Cutler and Bob Ladd for their comments on earlier versions of the paper.

ABSTRACT

Recent results on segmental anchoring suggest that rising prenuclear accents have two fairly invariant tonal targets (Arvaniti et al., 1998; Ladd et al., 1999). For German, however, there is introspective evidence that the rising topic accent (or theme accent) is realised differently when signalling contrast than when not. In this article, the acoustic basis for these reported intuitive differences is investigated in terms of the scaling (height) and alignment (positioning) of tonal targets.

Subjects read target sentences in a contrastive and a non-contrastive context (experiment 1). Statistical comparison showed that themes in contrastive context exhibited a higher and later peak, and segmental durations were longer compared to the identical sentence in non-contrastive context. The positioning and scaling of accents can hence be controlled in a linguistically meaningful way.

In experiment 2, non-linguists' perception of a subset of the production data was assessed. They chose whether in a contrastive context the presumed contrastive or non-contrastive realisation of the sentences was more appropriate. For some sentence pairs only, subjects performed better than chance, emphasising the importance of the acoustic differences.

A group of linguists then annotated the thematic accents of the contrastive and non-contrastive versions of the same data (from experiment 2). For half of the cases, the same accent type was assigned to the contrastive and non-contrastive version. There was a high degree of variability and overlap in annotating the data, which questions a categorical distinction in accent type for contrastive and non-contrastive themes in German.

Keywords: Contrast, Intonation, Annotation, Autosegmental-metrical, German

INTRODUCTION

The same sentence (i.e. the same string of words) can be produced with different speech melodies to express different meanings. A strong interpretation difference is achieved by associating a pitch movement with different words of an utterance (Krahmer & Swerts, 1999, 2001; Weber et al., 2004). In example 1a, for instance, the noun is accented (accented syllables are highlighted by small capitals), which leads to a neutral statement. It could be used as an answer to an information-seeking question, such as 'Can you think of a good present for Peter?'. If the adjective is accented, however, the sentence sounds contrastive, or even corrective (example 1b). It would be appropriate in a context in which the uncle brings a red bicycle along.

(1) a) He wanted a green BIcycle.

b) He wanted a GREEN bicycle.

Naturally, there have been a number of proposals on how to relate intonation to linguistic and paralinguistic meaning, on how to formalise intonational meaning and on how to incorporate intonational information into semantic formalisms (O'Connor & Arnold, 1961; von Essen, 1964; Crystal, 1969; Ladd, 1980; Gussenhoven, 1984; Brazil, 1985; Couper-Kuhlen, 1986; Cruttenden, 1986; Bolinger, 1989; Pierrehumbert & Hirschberg, 1990; Féry, 1993; Büring, 1997b; Steedman, 2000; Gussenhoven, 2002; Grice et al., 2005, among others). In attempts to describe the intonational system of a certain languages, a partial or full description of the possible pitch contours is presented together with the contexts in which these are appropriate (O'Connor & Arnold, 1961; von Essen, 1964; Crystal, 1969; Couper-Kuhlen, 1986; Cruttenden, 1986; Féry, 1993; Grice et al., 2005). Most semantic formalisms make use of intonational categories of some sort (e.g. A and B accents in Jackendoff (1972) and Büring (1997b), or autosegmental-metrical accent types in Steedman (2000)). Therefore, the adequacy of the formalisms depend on the kind of intonational categories referred to; especially it is important that they are sufficiently distinct prosodically and that they can be reliably annotated. Since prosodic realisations were shown to vary with regional background (Grabe, 2004; Gilles & Peters, 2004; Peters, 2004; Atterer & Ladd, 2004) and from speaker to speaker, semantically meaningful accent categories should be broad enough to include intra- and interspeaker variability, but at the same time narrow enough so that meaningful phonetic variation is captured. This demand may be difficult

to fulfil, especially in the light of the high proportion of inter-transcriber disagreement on the identity of accent types in intonational annotation (Lieberman, 1965; Pitrelli et al., 1994; Grice et al., 1996). Interestingly, semantic theories often use accent type distinctions that are highly overlapping to express important semantic distinctions. Steedman (2000), for instance, uses the accent L+H* for thematic contrast and H* for rhematic contrast (the terms 'theme' and 'rheme' refer to information structure primitives; themes link the sentence to the previous context, while rhemes provide new information). In the reliability study conducted by Grice et al. (1996) for GToBI (Grice et al., 2005), this accentual distinction resulted in the highest inner-annotator disagreement (33%). In the study of Pitrelli et al. (1994) these accent types have even been merged (p. 125).

In this paper, a semantic distinction that has received some attention in linguistics is analysed: the distinction between contrastive and non-contrastive themes (or topics[1]). Contrastive themes are said to point to alternatives (to the current theme) the speaker wants to talk about (Krifka, 1999; Umbach, 2001), to evoke a scope inversion if two quantifiers or a quantifier and a negation particle are involved (Wunderlich, 1991; Büring, 1997a; Krifka, 1998). Also, sentences with a contrastive theme can give rise to contrastive implicatures (Büring, 1997b; Lee, 1999). For instance, if a possible suspect in a murder trial is asked by the police constable 'Where were you at the time of the murder?' and the suspect answers with 'I was at home', he can do this both neutrally and in a way as to implicate that he knows of another person who might have been at the site of crime. In German, this effect is most easily achieved by producing a hat pattern[2] (i.e. a pitch rise on the personal pronoun (theme), a sustained high pitch, and a fall on the last word (rheme); this contour was first described by Cohen & t'Hart (1967) for Dutch). Consequently, the presence of a hat pattern is often related to contrastive themes (Wunderlich, 1991; Büring, 1997b; Mehlhorn, 2001).

The presence of a hat pattern is not always indicative of contrastive themes, however. Kohler (1991a), for instance, identified this contour as 'ideally suited for matter-of-fact reading in German' (p. 328). More and more intuitive prosodic descriptions on contrast have emphasised that it is especially the initial thematic pitch rise that differs in contrastive and non-contrastive contexts (Féry, 1993; Büring, 1997b; Jacobs, 1997). Büring (1997b), for instance, described the difference between contrastive and non-contrastive themes as a difference in pitch excursion of the thematic rise (p. 52). Jacobs (1997), on the other hand,

discussed that it is not the pitch rise *per se* that signals a contrastive theme but rather the fact that the rise is preceded by a noticeable trough (which is why he uses the notion *root contour*). Féry (1993) distinguished between two different accent types for thematic material in German, H* and L*H. According to her, however, the difference between these two contours 'is not always phonologically clear-cut' (p. 151).

It appears that this important semantic distinction has not been sufficiently analysed in prosodic terms yet. Except for a related production experiment conducted by Mehlhorn (2001) there is only introspective evidence for a prosodic difference in contrastive and non-contrastive contexts. Mehlhorn's experiment mainly aimed at eliciting hat patterns and she hence collected more material in contrastive contexts. Furthermore, a great proportion of the data consisted of sentences from the literature that have been said to be realised with hat patterns (such as sentences with particular syntactic structures, e.g. containing split-NPs, sentences with quantifiers, and complex sentences with more than one topic). She compared the realisations in the two contexts by overlying the f_0-contours and found that contrastive themes had a steeper rise, a higher f_0-range, and a longer syllable duration. Further, she reported that sentences with contrastive topics were generally produced by a hat pattern. This last finding, however, might be influenced by her choice of experimental material.

To avoid any bias, in the present production experiment, speakers read identical sentences in contrastive and non-contrastive contexts. As will be explained in more detail in the Method section of experiment 1, the target sentences had neutral word orders that do not demand particular intonational configurations. The data were analysed both acoustically (experiment 1), measuring the positioning and scaling of tonal targets and phonologically by a group of linguists (experiment 3). Also, the sentences' perception and interpretation was assessed (experiment 2).

Recent results in intonational phonology suggest strong regularities in intonational structure, both in terms of the scaling of tonal targets as well as concerning their position. Maeda (1976), for instance, reported that speakers vary little in the low pitch at the end of utterances. Liberman & Pierrehumbert (1984) showed that the height of accents is highly predictable from the context they appear in. Increasing the pitch height of an accent has been reported to increase its perceived prominence (Rietveld & Gussenhoven, 1985; Terken, 1991; Kohler & Gartenberg, 1991) and emphasis (Nolan, 1995; Ladd & Morton, 1997).

With respect to the positioning of accents, various researchers have reported stable segmental anchors for low tonal targets before an accentual rise (Caspers & van Heuven, 1993; Prieto et al., 1995; Arvaniti et al., 1998). Accentual peaks were reported to be influenced by the proximity of an upcoming prosodic boundary or accent, phonological vowel length, and speech rate (Steele, 1986; Silverman & Pierrehumbert, 1990; Caspers & van Heuven, 1993; Prieto et al., 1995; Arvaniti et al., 1998). Recently, Ladd and various colleagues argued that accentual peaks are also consistently aligned with the segmental structure. For instance, Modern Greek prenuclear accents were shown to have high targets at the end of the rise that were aligned at a fixed distance from the beginning of the first post-stressed vowel (Arvaniti et al., 1998). The presence of two fairly stable segmental anchor points was replicated for English rising accents by Ladd et al. (1999). They reported that there are only 'small and insignificant effects of rate on alignment' (p. 1543)[3].

If there is a 'norm' or a 'standard' in the scaling and positioning of tonal targets, deviations from that norm can probably be used to signal different functions. If the norm is tightly defined (which is suggested by the findings on segmental anchoring reviewed above), deviations will in principle not have to be large in order to express a different meaning. Hence, fine-grained variations in the scaling and alignment of accents could be linguistically meaningful and that is why an acoustic analysis is important. However, such fine-grained phonetic differences might not be noticeable enough to be interpreted differently or to be annotated by a different intonational category.

Consequently, the present paper is divided into three parts. First, in experiment 1, an acoustic analysis of actual productions of syntactically unmarked sentences in contrastive and non-contrast is described. To date there is no data available to support or undermine (introspective) proposals on the realisation of contrastive themes. In experiment 2, the perception and interpretation of sentences produced in contrastive and non-contrastive context is investigated. Linguistically naïve subjects choose which of two productions is more appropriate in a contrastive context. Mehlhorn (2001) also conducted two perception experiments on contrastive and non-contrastive themes in German but they demanded meta-linguistic judgements (relative prominence, direction of pitch movement). The findings of these experiments indicate that themes in contrastive contexts are perceived differently than themes in non-contrastive contexts. However, they do inform us about the

communicative function of different productions, i.e. whether they are *interpreted* differently. In experiment 3, finally, results from a labelling experiment are described. Productions in contrastive and non-contrastive contexts are taken out of context and annotated by a group of linguists. This experiment will show whether there is an accentual distinction between contrastive and non-contrastive themes. If there is an accentual distinction we will learn which intonational distinctions mark the semantic distinction between contrastive and non-contrastive themes. This information will help to improve the definition of a proper prosody–semantics interface.

EXPERIMENT 1

In experiment 1 the focus lies on the question whether the precontext (contrastive vs. non-contrastive) has a significant effect on the acoustic realisation of prenuclear accents in German (in terms of scaling and alignment). More specifically, two questions are addressed:

1. Are syntactically unmarked German sentences with contrastive themes consistently produced with a lower trough before the rise (Jacobs, 1997) and a larger pitch-excursion (Büring, 1997b)?

2. Are syntactically unmarked German sentences with contrastive themes produced with hat patterns? Are hat patterns only found in contrastive contexts and not in non-contrastive contexts (Büring, 1997b; Mehlhorn, 2001)? Or are hat patterns not related to contrastiveness (Kohler, 1991a)?

Method

Provided that differences in precontext are indeed expressed prosodically, these differences should hold across speakers and across differences in phonological and metrical structure. Therefore, a fair amount of variation was included in the experimental items. The stimuli varied in the number of syllables in the target word, phonological vowel length of the stressed syllable, and the position of the word stress in the thematic constituent. Furthermore, subjects were not chosen from a particular variety of German, but came from different regions.

Short paragraphs were designed and the target sentences were embedded in these larger contexts that controlled both the information-structure and the contrastiveness. Identical

sentences were hence produced in a contrastive and non-contrastive context. Analysing alignment in continuous read text differs from most previous studies on alignment; these mostly had lists of isolated sentences to be read aloud (cf. Arvaniti et al., 1998; Ladd et al., 1999; Atterer & Ladd, 2004, among others).

Participants. Twelve native German speakers, seven female and five male, voluntarily participated in the recording. They were between 23 and 36 years old. At the time of recording they were graduate and postgraduate students, or staff members of Edinburgh University; they had been in Edinburgh for periods ranging from a few months to four years. The participants all spoke Standard German but originated from different parts of Germany (eight northern German speakers and four southern German speakers, for more detail see Table 8 in the Appendix). None of them had any known speech or hearing problems and they were naïve with respect to the purpose of the experiment.

Materials. Target sentences started either with a subject noun-phrase or a prepositional phrase. These word orders are very frequent in German (Weber & Müller, 2004) and are not inherently contrastive. Target sentences consisted of a simple theme–rheme structure, with the theme identifying a topic (an animated subject, a location or a time) and the rheme making a proposition about it. Sentences were constructed so that the rheme accent (focus accent) would fall on the same constituent in contrastive and non-contrastive contexts. This reduces the influence of the rhematic accent on the phonetic implementation of the thematic accent (the upcoming prosodic context was shown to influence the alignment of high targets, see e.g. Silverman & Pierrehumbert, 1990; Prieto et al., 1995; Caspers & van Heuven, 1993).

Target constituents were chosen to be maximally sonorant and to be groupable with other items in a 'set-of-alternatives', e.g. the noun-phrase 'the Romans' (which can be contrasted with 'the Teutons'). In total, twelve target sentences were constructed; they are listed in Table 9 in the Appendix.

Two conditions were investigated, contrastive and non-contrastive precontext. Twelve short paragraphs were constructed for both conditions with five to six sentences each (on average 5.5 sentences in non-contrastive contexts and 5.6 in contrastive contexts). To reduce the influence of paragraph intonation (Sluijter & Terken, 1993), the position of the target sentence in the paragraph was matched across contexts. Target sentences did not

non-contrastive context	contrastive context
Malaysien ist vielen Europäern weitgehend unbekannt. Das Land besteht aus zwei Inseln. Zur Vereinfachung der Kommunikation zwischen den beiden Inselteilen besitzt fast jeder Haushalt einen Computer mit Internetverbindung. Trotzdem ist Malaysien kein hochtechnologisches Land. **Die Malayen leben von der Landwirtschaft**. Sie sind zwar nicht reich, aber auch nicht besonders arm.	Malaysia und Indonesien sind Nachbarländer im Südchinesischen Meer. Trotz dieser geographischen Nähe unterscheiden sich die Lebensgewohnheiten der Malayen und Indonesier stark. In Indonesien spielt der Tourismus eine große Rolle. Dieser Sektor bietet viele Arbeitsplätze. **Die Malayen leben von der Landwirtschaft**. Vor allem malaysischer Reis ist in ganz Asien bekannt.

Table 1: Sample non-contrastive and contrastive paragraph. The target sentences are highlighted in bold face. For an English translation see Table 7 in the Appendix.

appear in paragraph-final position to avoid paragraph-final lowering of f_0 and creaky voice.

Two sample paragraphs are shown in Table 1, for an English translation see Table 7 in the Appendix. For the non-contrastive context, the target constituent or a hypernym was introduced as a topic early in the paragraph and resumed in the target sentence (e.g. Malaysia, the Malaysians). In contrastive paragraphs, a proposition about an alternative element occurred in the precontext (e.g. the Indonesians, the Malaysians). To link the two elements to one another, the precontext also contained a sentence stating the commonality between the two alternatives. The contrast is always established by a semantic parallelism (Prevost, 1995; Theune, 1999). To avoid exaggerated productions, the sentences in question were not syntactically parallel.

In addition, twelve distractor paragraphs on the same topic as the target paragraphs were constructed to prevent subjects from expecting identical sentences in paragraphs about certain topics. Sixteen other filler paragraphs, which were unrelated to the experi-

mental items, were constructed. This resulted in 52 paragraphs.

Procedure. Participants received written instructions to read the paragraphs at normal speed as fluently as possible. They were given a pile of 52 A5 cards that contained the paragraphs. They were told to silently scan the texts before reading them aloud. If paragraphs contained too many hesitations or false starts they were asked to read them again. Participants were recorded in a sound-proof room in the Department of Theoretical and Applied Linguistics at Edinburgh University. Data was stored simultaneously on DAT-tapes and on a PC with a sampling rate of 44.1kHz. The presentation of the paragraphs was block-wise randomised, separating the contrastive and non-contrastive versions of a given target sentence by a minimum of five other paragraphs.

Analysis

The recordings of three speakers were discarded before analysis. Two of them read too fast and monotonously, which made the identification of f_0-minima and maxima almost impossible and therefore unreliable[4]. One speaker had a very emphatic reading style and inserted an intermediate phrase break after all the thematic elements. One sentence ('In America besitzen viele eine Waffe', Engl. 'In America, many people own a weapon') had to be discarded from analysis for most of the speakers because the focal (rheme) accent was realised on different constituents in the two conditions. Various other sentences had to be excluded because of hesitations and mispronounciations in the target sentence for one of the two conditions. For one speaker, there were recording errors, so that half of her sentence-pairs had to be discarded. For each of the remaining speakers, two sentence pairs were excluded. The unequal number of items per speaker is not crucial since the experiment is designed for within-subject comparisons. These comparisons are based on 86 utterance pairs.

Data were analysed using Xwaves; f_0-tracking was conducted with the in-built pitch-tracking algorithm (get_f0), which is based on normalised cross correlation function and dynamic programming (cf. Talkin, 1995). We used the default values of 10 ms frame shift and 7.5 ms autocorrelation window. Artefacts introduced by the pitch-tracking algorithm (pitch doubling or halving) were manually corrected. Up to a sequence of 5 missing f_0-values were linearly interpolated. Then, the pitch-contour was smoothed using a 7-frame

window (7.5 ms each) with mean smoothing.

Labelling. Data annotation was done on the segmental and suprasegmental level, concentrating around the area of the f_0-rise. Label points are summarised in Figure 1, including suprasegmental, segmental, and lexical labels. Standard segmentation criteria (cf. Peterson & Lehiste, 1960) were followed for segmental annotation. Following Ladd et al. (2000); Atterer & Ladd (2004), four segmental landmarks situated in the area around the stressed and post-stressed syllable were marked:

C0: Consonantal onset of the stressed syllable

V0: Start of the stressed vowel

C1: Consonantal onset of the post-stressed syllable

V1: Start of the post-stressed vowel

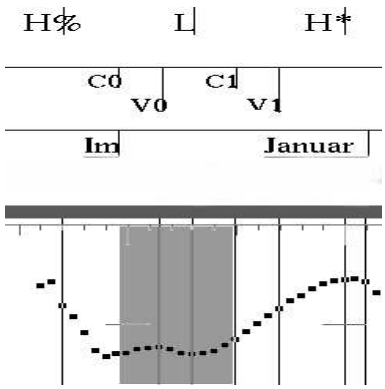

Figure 1: Suprasegmental and segmental labels, together with lexical information ('im Januar'). The stressed syllable is marked by a shaded area. Lexical labels mark the end of words. Segmental labels mark the start of the segment, suprasegmental labels mark the time point of the event.

On the suprasegmental level, the following events in or before the test words were marked. Note that these labels are purely acoustic landmarks and are not meant to correspond to ToBI labels:

H%: High point before the fall. In most cases this value was found in the middle of the vowel of the first unstressed syllable of the prosodic word. If this value was not

reliable (e.g. because of creaky voice, devoicing), the value in the following sonorant was taken (often the case in PPs beginning with 'in' or 'im').

L : Local minimum preceding the rise. In cases with a low plateau, the elbow point was manually detected and marked. The elbow represents a considerable change in the slope (from a flat plateau to a steep rise). Plateaus were defined to consist of at least four pitch points where two consecutive f_0-values must not differ more than 1 Hz.

H*: First local maximum after the stressed syllable. If there was a high plateau following the rise, the point with a considerable change in slope was marked. As in the case of low plateaus, a high plateau was defined as more than four subsequent pitch points where two consecutive f_0-values must not differ more than 1 Hz.

Dependent Variables. Four groups of acoustic variables were investigated: f_0-variables, temporal variables, alignment variables, and the slope of the rise.

Overall, five f_0 variables were analysed, the absolute f_0-value in Hz for the three suprasegmental events (f_0(H%), f_0(L), and f_0(H*)) and two dynamic measures, signalling the magnitude of the f_0-fall and the magnitude of the subsequent rise, both in Hz and in semitones (Δf_0(fall), Δf_0(rise)).

The temporal variables consisted of the duration of the whole utterance (utt_dur), of the thematic constituent (theme_dur), the stressed syllable (syll_dur), and the stressed vowel (vowel_dur). Further, the duration of the f_0-fall and the f_0-rise (Δt(fall) and Δt(rise)) were measured.

Alignment variables represent a link between the segmental and suprasegmental tier. It is still an open methodological issue whether alignment should be measured in absolute terms (referring to some given anchor point) or in proportion to the duration of the segment or syllable the event appears in (see discussions in Silverman & Pierrehumbert, 1990; Prieto et al., 1995; Atterer & Ladd, 2004; Schepman et al., submitted). Therefore, alignment was measured both absolutely and proportional to the duration of the stressed syllable. In German (as in Greek, cf. Arvaniti et al., 1998), the peak in prenuclear accents is reached in the post-stressed syllable only(Atterer & Ladd, 2004). Peak alignment was calculated with respect to the start of the stressed vowel (al(H*,V0), with respect to the start of the post-stressed syllable (al(H*,V0end)) and to start of the post-stressed vowel al(H*,V1)).

The alignment of the trough was calculated to the start of the stressed syllable and to the start of the stressed vowel: al(L,C0), al(L,V0). Following Silverman & Pierrehumbert (1990), all the alignment variables were also expressed proportionally, with respect to the duration of the stressed syllable.

Xu (2002) made the case for the importance of the speed of pitch changes. Finally, therefore, the slope of the rise was also calculated, by dividing the f_0-range by its duration (slope(rise)).

Since it has been claimed that contrastive themes are produced with a hat pattern (Jacobs, 1996; Büring, 1997b; Mehlhorn, 2001), the number of hat patterns were counted in contrastive and non-contrastive contexts. Hat patterns were defined as cases in which the pitch stays high between the two accents. Therefore, only sentences where the pitch did not drop down to a low tone between the prenuclear and nuclear accent were counted as having a hat pattern.

Further, the frequency distribution of the nuclear accent types was analysed in more detail. Nuclear accents have been shown to be categorically distinct in German (Kohler, 1991b); early peaks (comparable to ToBI (!)H+L*) contrast with medial peaks (H*). The number of high nuclear accents (collapsing the autosegmental-metrical accent types H*, !H*, L+H*, L+!H*) and low nuclear accents (collapsing !H+L* and H+L*) was compared in contrastive and non-contrastive contexts.

Results and Discussion

In this section, the results of the production experiment will be presented and discussed. First, the realisation of the theme accent in contrastive and non-contrastive contexts is described. Then, correlation analyses are discussed that explore the interaction between the acoustic variables peak height, peak position, slope of the rise and so on. This is important to understand the actual phonetic implementation of rising thematic pitch accent. Since hat patterns have been repeatedly related to contrastive themes, This is followed by an investigation of the rheme accent type and the presence of hat patterns. Finally, the interaction between rheme accent type and theme accent realisation is described.

Theme accent realisation in contrastive and non-contrastive context. Contrast is a within-subjects factor, so paired t-tests were calculated for the acoustic variables. Because of

multiple t-tests, the standard significance level of p = 0.05 was adjusted to p = 0.004 (Bonferroni correction[5]). From the 23 variables analysed, 12 differed significantly in contrastive and non-contrastive context. The results of the paired t-tests for the different groups of variables are presented in Tables 2, 3, and 4.

	Non-contrastive			Contrastive			
Variable	mean	sd	N	mean	sd	N	p
f_0(H%) in Hz	168.1	51.4	86	167.1	49.7	86	.5
f_0(L) in Hz	154.9	47.7	86	151.9	47.8	86	.01
f_0(H*) in Hz	218.0	78.9	86	227.7	79.3	86	$< .001$
Δf_0(fall) in Hz	13.2	9.8	86	15.2	11.7	86	.19
Δf_0(fall) in st	1.4	1.0	86	1.8	1.4	86	.06
Δf_0(rise) in Hz	63.1	40.4	86	75.8	38.9	86	$< .001$
Δf_0(rise) in st	5.4	2.2	86	6.5	1.9	86	$< .001$

Table 2: Means and standard deviations of f_0-variables in contrastive and non-contrastive contexts, and significance value of a two-tailed paired t-test. Significance level $p = 0.004$. Significant differences are highlighted in bold face.

For the static f_0-variables, only the height of the peak differed significantly in contrastive and non-contrastive contexts. On average, prenuclear peaks in contrastive contexts were 9.7±2.7 Hz higher than peaks in non-contrastive contexts[6]. This was also reflected in the magnitude of the f_0-rise, which was 12.7±2.4Hz larger for contrastive contexts (equivalient to 1.1±0.2 st). The magnitude of the f_0-fall, however, did not differ significantly.

Overall utterance duration did not differ significantly in the two contrast conditions. The duration of the thematic constituent was significantly longer in contrastive contexts than in non-contrastive ones, on average 19.6±4.6 ms. Since the overall utterance duration was the same in the two conditions, the thematic lengthening was at the expense of the rhematic part of the utterance. The duration of the stressed syllable was also significantly longer in contrastive contexts than in non-contrastive ones (8.3±2.7 ms) and so was the

duration of the stressed vowel (5.8±1.9 ms). There was no significant difference in the duration of the f_0-fall prior to the accentual rise but the duration of the pitch-rise took significantly longer in contrastive contexts than in non-contrastive context (20.5±5.4 ms).

	Non-contrastive			Contrastive			
Variable	mean	sd	N	mean	sd	N	p
utt_dur in ms	1611.6	351.4	86	1622.4	233.2	86	.32
theme_dur in ms	436.6	88.2	86	457.1	88.4	86	< .001
syll_dur in ms	152.7	36.1	86	161.0	39.2	86	.003
vowel_dur in ms	96.4	29.3	86	102.2	29.9	86	.004
Δt(fall) in ms	115.2	59.0	86	125.5	64.2	86	.11
Δt(rise) in ms	178.8	43.8	86	199.3	42.7	86	< .001

Table 3: Means and standard deviations of temporal variables in contrastive and non-contrastive contexts, and significance value of a two-tailed paired t-test. Significance level $p = 0.004$. Significant differences are highlighted in bold face.

The peak was significantly later in contrastive than in non-contrastive contexts. The peak calculated with respect to the stressed vowel was 25.8±4.0 ms later in contrastive contexts, calculated with respect to the start of the post-stressed syllable 20.0±3.7 ms later. The alignment of the trough, however, did not differ significantly. The proportional alignment variables showed the same effect as the absolute alignment variables.

Detailed acoustic comparison revealed that themes in contrastive contexts were realised with a later and higher peak, and longer segmental durations. Also, the duration of the f_0-rise was longer in contrastive contexts and there was a larger rise-excursion. The data hence confirm Büring's impressionistic description of a larger pitch excursion. There was, however, no lower trough for contrastive contexts, as discussed by Jacobs (1997). In contrastive contexts, peaks were also realised later than in non-contrastive ones. Gussenhoven (2002) has recently claimed that 'higher peaks will tend to be later than lower peaks' (p. 52), arguing that it takes longer to reach a higher peak. Also, he regards peak delay as a

	Non-contrastive			Contrastive			
Variable	mean	sd	N	mean	sd	N	p
al(H*,V0) in ms	169.5	41.9	86	195.3	44.2	86	< .001
al(H*,V0end) in ms	73.0	28.7	86	93.1	37.8	86	< .001
al(H*,V1) in ms	20.9	30.2	86	42.4	35.0	86	< .001
al(L,C0) in ms	47.0	43.7	86	54.9	49.0	86	.16
al(L,V0) in ms	-9.3	41.8	86	-4.0	42.3	86	.30
al(H*,V0)_prop	1.13	.26	86	1.26	.35	86	< .001
al(H*,V0end)_prop	.50	.24	86	.63	.34	86	< .001
al(H*,V1)_prop	.14	.24	86	.30	.28	86	< .001
al(L,C0)_prop	.29	.28	86	.31	.28	86	.48
al(L,V0)_prop	.08	.30	86	.05	.28	86	.36
slope(rise)	.35	.19	86	.38	.18	86	.02

Table 4: Means and standard deviation of alignment variables and the slope of the rise in contrastive and non-contrastive contexts.

substitute for peak height (Gussenhoven, 2002, 2004). Xu (2002) and Arvaniti et al. (1998), on the other hand, argued that higher targets need not be later because the articulatory velocity can be increased to reach higher targets.

Correlation between dependent variables. I analysed whether an increase in pitch height was correlated with a peak delay (which would be in line with Gussenhoven's view that these are adjusted together) or not. Also, I tested how well an increase in peak height is correlated with a steeper slope (which would be consistent with Xu's and Arvaniti et al.'s position).

Pearson correlations were calculated for the raw variables, i.e. contrastive and non-contrastive versions were pooled. Peak height correlated significantly with the slope of the rise ($r = 0.82$, $p < 0.001$, $N = 172$) but not with the position of the peak (for none of the alignment variables measured).

Since the raw data are subject to much unwanted variation, such as different f_0-level or speech rate, another set of correlation analyses were calculated, using individual comparisons between the contrastive and non-contrastive realisation of each utterance pair for every speaker. In other words, the non-contrastive version was seen as a baseline against which the deviation found in the contrastive context was calculated. Peak height ratio was computed by dividing the absolute peak height in each contrastive version by the absolute peak height of each corresponding non-contrastive version. Similarly, slope ratio and peak alignment difference were calculated.

Peak height ratio did not correlate with peak alignment difference. It only correlated slightly with the slope ratio (Pearson $r = 0.52$, $p = 0.47$, $N = 86$). Increased peak height is hence achieved by an increased slope, but not very consistently; otherwise the correlation would be higher. Later peaks were not mere phonetic concomitants of higher peaks. Figure 2 shows a scatterplot of peak height ratio and peak alignment difference. It shows that there is a high degree of variation in conveying a contrastive theme (compared to a non-contrastive theme). Most frequent was the use of both a later *and* a higher peak in contrastive contexts compared to non-contrastive ones (46.5% of the utterance pairs). Figure 2 hence illustrates Gussenhoven's claim that peak delay can be a substitute for peak height (see especially Gussenhoven, 2004, p. 90-92).

Atterer & Ladd (2004) investigated the alignment properties for northern and southern

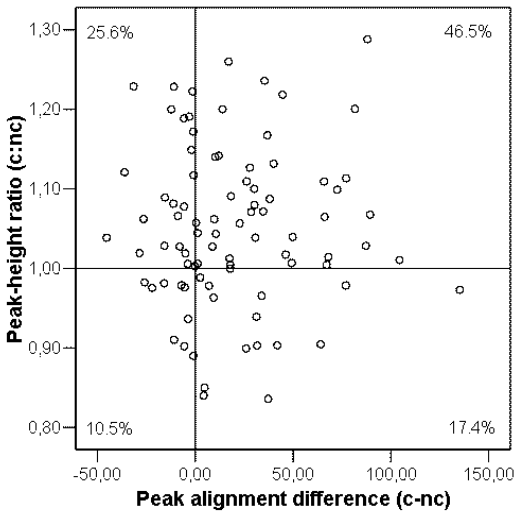

Figure 2: Scatterplot of peak height ratio and peak alignment difference (of the alignment variable al(H*,V0end) in ms). To the right to the vertical line, peaks are later in the contrastive versions than in the respective non-contrastive ones; above the horizontal line, peaks are higher for contrastive contexts. Percentages indicate what proportion of the data falls into the respective quadrants.

German speakers and reported that the trough before the prenuclear rise was significantly later for southern German speakers than for Northerners (this study came to my eyes only after having finished the recordings). A similar trend was observed for the high target at the end of the rise but that difference was not statistically significant. Given their results, it would be likely that southern Germans (whose rises are generally later) would employ peak height more extensively than northern German speakers (who have more freedom to delay the peak). A non-parametric Mann-Whitney test with speaker-origin (north vs. south) as independent variable and peak-height ratio and alignment-difference as dependent variables revealed that there was no influence of dialect on the contrast-marking strategies. Even if Northern and Southern German speakers have different alignment strategies, these do not interfer with with their contrast-marking.

Hat patterns and rheme accent type. Overall, only 10% of the sentences were realised with a hat pattern, and hat patterns were not more frequent in contrastive than in non-contrastive contexts ($\chi^2 < 2$). That is, in syntactically neutral, read German sentences with a contrast

in both the thematic and rhematic element, speakers do not tend to produce hat patterns. This is apparently different in sentences with particular syntactic constructions (Steube, 2001; Mehlhorn, 2001) or in sentences in which pronouns or adjectives are marked with the rising theme accent (see examples in Büring, 1997b).

The frequency distribution of the rhematic (nuclear) accent is shown in Table 5. There were significantly more low rheme accents in the data than high ones ($\chi^2 = 7.53$, df = 1, $p < 0.01$). There was no significant difference in accent distribution in non-contrastive contexts, but there were significantly more low rheme accents than high ones in contrastive contexts ($\chi^2 = 13.44$, df = 1, $p < 0.01$). A contrastive precontext (i.e. a contrastive theme and rheme in the precontext) hence provoked significantly more low nuclear accents[7]. This indicates that — instead of a prototypical hat pattern — a low nuclear accent could be the primary intonational realisation of a contrastive theme. Semantically, however, it is difficult to establish a direct connection between low rheme accents and contrastive contexts as these accents were also quite common in non-contrastive contexts. As there might be similar variations in rheme accent realisation as found for theme accent, more data is needed to establish a connection between contrastive contexts and rheme accent type.

	Non-contrastive context	Contrastive context	Total
High accent	42	26	68
Low accent	44	60	104
Total	86	86	172

Table 5: Frequency distribution of rheme accents in contrastive and non-contrastive contexts.

Interaction between theme and rheme accent. In contrastive contexts, speakers produced significantly more low rheme accents than high ones. Silverman & Pierrehumbert (1990) and Caspers & van Heuven (1993) showed that the position of the peak was influenced by the proximity of a following accent. It is unclear whether — in the present data —

the accent *type* of the nuclear (rhematic) accent influenced the prosodic realisation of the prenuclear (thematic) accent. Therefore, I tested whether there was an interaction between the acoustic realisation of the thematic accent and rheme accent type. A non-parametric Mann-Whithey test with rheme-accent type as independent variable and the different alignment-variables as dependent variables was calculated. Except for al(H*,V0), none of the alignment variables showed a difference. The position of the peak (with respect to the stressed vowel) was significantly later when the rheme accent was low and earlier when the rheme accent was high ($p < 0.001$). Why is this effect not reflected in the other alignment variables? One explanation is that the other alignment-variables calculate the position of the peak with respect to the start of the *post-stressed* syllable or vowel; al(H*,V0) is the only alignment variable that is linked to the (start of the *stressed syllable*. I therefore hypothesised that the rheme accent type may predominantly influence the duration of the stressed thematic syllable, which then results in an alignment difference for al(H*,V0) only. Calculating a Mann-Whithney test with syllable duration (sylldur) and duration of the stressed vowel (durV0) as dependent variables confirmed this hypothesis. The durations of the thematic vowel and syllable were significantly longer when the rheme accent was low ($p < 0.01$). The rheme accent type also influenced the peak height of the thematic accent ($p = 0.02$). These results appear to demonstrate some sort of interaction between the thematic and rhematic accent realisation but these could equally well be independent effects of the contrastive context.

The results of the production experiment have clearly shown that linguistically naïve speakers mark themes in contrastive contexts differently from themes in non-contrastive contexts, mainly by means of peak height, peak position, and segmental duration. As opposed to claims in the literature, sentences with contrastive themes were neither generally realised with a hat pattern, nor were they realised more often with a hat pattern. Even though the hat pattern was not the standard contour in contrastive contexts, rhemes in contrastive contexts were significantly more often realised with a low rhematic accent. This might point to the realisation of a 'weakened' hat pattern in contrastive contexts. However, since high and low rheme accents were equally distributed in non-contrastive contexts, it is too early to arrive at a generalisation.

The question that arises now is whether the acoustic differences found between theme

accents produced in contrastive and non-contrastive contexts are meaningful or whether they merely constitute phonetic variation in the implementation of sentence-initial rising accents? To test whether the different productions lead to interpretation differences, a perception experiment with linguistically naïve listeners was performed.

EXPERIMENT 2

In experiment 2 the linguistic interpretation of utterances produced in contrastive and non-contrastive contexts is tested. This experiment will give information about the question whether the acoustic differences between themes in contrastive and non-contrastive contexts are 'real', i.e. semantically meaningful.

Method

Subjects were visually presented the start of a semantic parallelism and had to decide between two auditorily presented continuations. The auditory stimuli were sentence pairs pseudo-randomly selected from the production data (see below). If subjects are able to hear differences between the two realisations and consistently choose a particular version, then it will be important to capture the intonational differences found. If their choice is random, then the acoustic differences are not strong, consistent, or unambiguous enough to point to a certain context. In these cases, the intonational information need not be incorporated into semantic formalisms.

Participants. Fourteen graduate and postgraduate students voluntarily took part in the experiment. They were native speakers of German with no known hearing problems. All participants were naïve with respect to the purpose of the experiment.

Materials. Ten sentence pairs were selected. The two versions of each sentence pair differed with respect to peak alignment and peak height, as shown in Figure 3. To minimise any effect of rhematic accent type, only sentence pairs with the same rheme accent type in contrastive and non-contrastive realisation were included (seven containing low accents, three high ones, see Table 10 in the Appendix).

Three further sentence-pairs were included for familiarisation at the start of the experiment and one in the end. The stimuli were presented as the second clause of a semantic

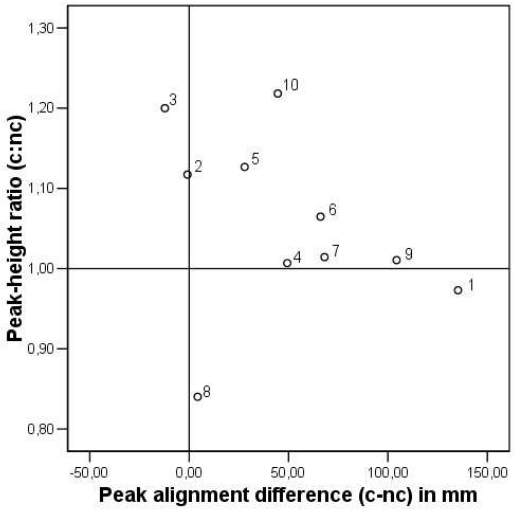

Figure 3: Peak height ratio and peak alignment difference (al(H*,V0end)) for the materials used in experiment 2 and 3. The scale in this figure is identical to the one used in Figure 2.

paralellism. Semantic parallelisms consist of two clauses and both the theme and the rheme of the two clauses are contrasted (e.g. In Indonesia, tourism is very important and the Malaysians live from agriculture). For every stimulus pair, an appropriate semantic parallelism was constructed. The full set of materials are listed in Table 11 in the Appendix.

Procedure. The experiment was carried out on a PC; the stimuli were presented via headphones (Sony MDR-V150). The first clause of the parallel construction was presented visually on the screen, followed by three dots to indicate the continuation. There were two loudspeaker symbols on the screen, one for the contrasitve version, one for the non-contrastive version. These were labelled a) and b) respectively. After clicking on a loudspeaker symbol, subjects heard the contrastive or non-contrastive version[8].

Subjects received written instruction to chose the more appropriate continuation to the semantic parallelism. The experiment was self-paced. Participants could hear the two possible continuations as often as they wished by clicking on two symbos labelled a) and b). They were asked to tick the more appropriate choice on paper. To prevent subjects from making decistions that they were not able to make, they could tick c) if they heard no difference at all between the two versions. Stimulus presentation was randomised and the order of the contrastive and non-contrastive versions was reversed for half of the subjects.

The order of filler items was fixed for all lists.

Results and Discussion

In the majority of cases (95.7%), subjects chose either the contrastive or non-contrastive version. Only 5 subjects made use of the category 'no difference', one subject twice (for pairs 1 and 8), and four subjects once (for pairs 2, 4, 5, and 6, respectively). Hence there was no single utterance pair for which the two versions were particularly hard to distinguish.

Contrary to expectation, the presumed contrastive version was not chosen more often than the presumed non-contrastive version. Only in 54±27% did subjects prefer the contrastive version. Instead, the presumed contrastive and non-contrastive versions were interpreted rather differently by different listeners. For most of the utterances pairs it is the case that what sounded appropriate for some listeners was inappropriate for others, see Figure 4.

There were a few sentence pairs (pairs 4, 7, and 9), however, for which there was a high agreement between listeres as to the more appropriate version. For sentence pairs 4 and 9, subjects significantly preferred the contrastive version ($\chi^2 = 7.2, df = 1, p < 0.05$ and $\chi^2 = 5.78, df = 1, p < 0.05$, respectively). For sentence pair 7, however, subjects significantly preferred the *non-contrastive* version ($\chi^2 = 8.64, df = 1, p < 0.05$)[9]. At least for these sentence pairs, the difference in intonational form corresponds to a functional difference.

Subjects' preference patterns are difficult to explain in terms of peak height or peak alignment differences between contrastive and non-contrastive versions (as e.g. shown in Figure 3). The pairs with constisten preference patterns (4, 7, and 9), for instance, all show a considerable difference in peak alignment between the contrastive and non-contrastive version. But so do the versions of pair 10 for which there was no significant preference for any of the versions. Also, the acoustic differences in theme marking cannot explain why listeners preferred the non-contrastive version for pair 7, but the contrastive one for pairs 4 and 9 although the alignment differences point in the same direction.

So how can the results be explained? To find out to which acoustic differences listeners were most sensitive, the number of contrastive responses to an utterance pair were cor-

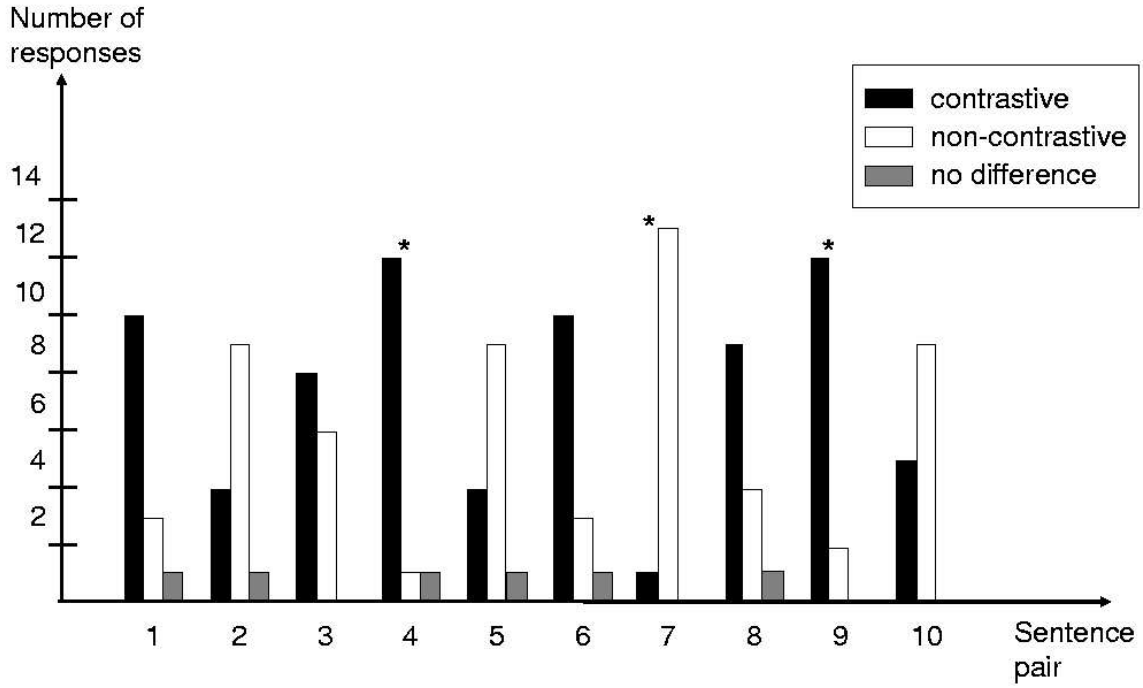

Figure 4: Number of contrastive, non-contrastive and 'no-difference' choices for the ten sentence pairs. Asterisks mark a significant difference ($p < 0.05$) between number of contrastive and non-contrastive responses ($\chi^2 \geq 5.78$).

related with several acoustic variables that describe the realisational difference between the two versions. These acoustic variables were (a) the difference in peak alignment between contrastive and non-contrastive realisation, (b) the peak height-ratio between contrastive and non-contrastive realisation, (c) the f_0-excursion-ratio between contrastive and non-contrastive realisation, and (d) the slope ratio of the rise between contrastive and non-contrastive realisation, and (e) the ratio of the duration of the thematic constituent between contrastive and non-contrastive realisation. Using a Spearman's Rho, there was no correlation between the number of contrastive responses and any of the acoustic variables analysed.

It might be that listeners were sensitive to information other than the realisation of the thematic accent, such as the rheme accent. Although the rhematic accent type of each utterance pair was controlled, there might be phonetic differences in the realisation of the rheme accent that listeners reacted to. Another possible explanation is that subject were reacting to other functions of prosody besides contrast marking. One subject, for instance,

reported afterwards that she dispreferred 'exaggerated' pitch ranges (a similar result was obtained by Chen, 2003). Also in some linguistic perception experiments, a paralinguistic interpretation of intonational phenomena was observed (Andreeva & Barry, 1999; Braun, 2004).

Given the observed differences in the realisation of theme accents in contrastive and non-contrastive contexts (experiment 1), and given that listeners could reliably classify certain versions, it seems vital to equip semantic formalisms with this intonational information. As discussed in the introduction, semantic theories already employ some sort of intonational information but this is often based on introspection. Experiment 1 provided natural productions of neutral German sentences, recorded in pragmatically different contexts. Some of them were perceptually evaluated in experiment 2. These materials constitute a perfect testbed to investigate whether the natural differences in theme accent realisation in the two contexts are sufficiently distinct to be reliably annotated with different accent types. This should be the case for those sentence pairs for which listeners had clear preferences towards a particular version.

EXPERIMENT 3

The aim of experiment 3 was to investigate whether the acoustic differences found between contrastive and non-contrastive contexts lead to different accent labelling. Different accent types for themes in contrastive and non-contrastive contexts are a necessary precondition for incorporating thematic contrast into semantic theories. Since for sentence pairs 4, 7, and 9 naïve listeners consistently preferred certain versions in the contrastive context, we should expect to obtain distinct accent types for these cases at least. If linguists were able to label these data consistently with different accent types for contrastive and non-contrastive versions, we should be confident in using information about pitch accent types for semantic theories. If, on the other hand, labelling turned out to be rather inconsistent, it would be necessary to rethink present-day intonational annotation.

Method

A group of linguists, trained in using the GToBI annotation system (German Tone and Break Indices, see Grice et al., 2005) for intonation annotation labelled the prenuclear

(thematic) accents of a subset of the data collected in experiment 1. GToBI is the German variant of the original English ToBI system (Silverman et al., 1992; Beckman & Ayers, 1997). It was described as a 'set of conventions for labelling German intonation with the aim of being easy to learn, reliable, and adaptable for different labelling purposes' (Grice et al., 2005, p. 62).

Participants. Eight linguists from different German universities voluntarily participated in the labelling experiment. They had native or near-native command of German and were naïve with respect to the purpose of the experiment. All of them were trained in using GToBI for intonational annotation.

Materials. The stimuli were the same ten sentence pairs as used in experiment 2.

Procedure. The experiment was conducted on-line. The 20 stimuli were presented in pseudo-random order, separating the two versions of a sentence pair by at least two other sentences. Labellers annotated the prenuclear accent only. They could listen to the whole utterance or to the preverbal constituent in isolation, with the option to play the parts as often as they wanted. The f_0-contour could be inspected by downloading the files. They were given the choice between the three most probable accent types (H^*, $L+H^*$ and L^*+H) as well as the category 'other accent type' that they could specify in a special text field nearby. They were further asked to briefly explain their choice. Furthermore, they had to rate the certainty of their judgement on a scale from 1 (very uncertain) to 5 (very certain)

Results and Discussion

The average certainty was 3.51, ranging from 3.00 to 4.04 (mean values per subject). There were no specific stimuli that made labellers particularly uncertain. Counting those items that received the lowest number of confidence value per subject, the most difficult stimuli seemed to be 12 and 15 (lowest certainty value obtained from 3 labellers), as well as 2, 14 and 19 (each item received lowest confidence score from two labellers). This affects the labelling of the non-contrastive version of pairs 8 and 10, as well as the contrastive version of pairs 1, 6, and 7.

The results of the labelling experiment are summarised in Table 6. To simplify the presentation of the results, the data is first discussed in terms of percentage of same or

different labels for the two versions of a sentence pair. Only then we concentrate on the identity of accent types used.

Frequency of 'same' and 'different' labels. In half of the cases (52.5±21.9%), the two versions of an utterance pair were annotated with the same accent type (note, however, that 'same accent' here is not equivalent to 'intersubjectively identical'; rather, different annotators used different intonational categories to express their percept, see Table 6). In half of the cases, hence, the two versions were not annotated with different accent types. There is, however, some variation in the sentence pairs, see Figure 5. There were three sentence pairs (pairs 5, 7, and 9) for which the majority of annotators used different accent types for the contrastive and non-contrastive version (high black bars). Note that these three sentence pairs do not differ considerably from the other sentence pairs in terms of a differene in peak height and peak position between the contrastive and non-contrastive realistion (see Figure 3).

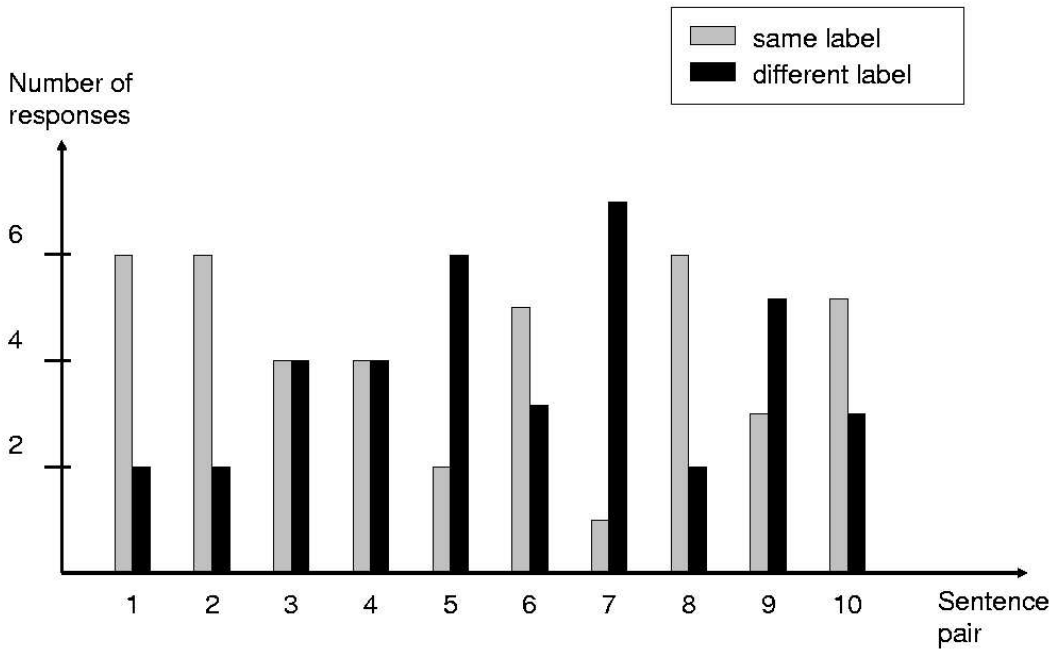

Figure 5: Number of same and different labels for the ten sentence pairs.

Strikingly, the sentence pairs that obtained most different labels are not exactly the same as the ones that naïve listeners most successfully classified. Listeners were particularly good for sentence pairs 4, 7, and 9, while we find most different labels for pairs 5, 7, and 9. Obviously, the two tasks are not directly comparable. As discussed before, listeners in

experiment 2 compared *complete* utterances, while the labellers in theory concentrated on the thematic accent only.

As in experiment 2, correlation analyses were conducted to find out to which acoustic differences labellers were most sensitive. The number of labellers who annotated the two versions of an utterance pair with a *different* accent type was correlated with (a) the difference in peak alignment between contrastive and non-contrastive realisation, (b) the peak height-ratio between contrastive and non-contrastive realisation, (c) the f_0-excursion-ratio between contrastive and non-contrastive realisation, and (d) the slope ratio of the rise between contrastive and non-contrastive realisation, and (e) the ratio of the duration of the thematic constituent between contrastive and non-contrastive realisation. There was only a significant correlation between number of 'different' labels and f_0-excursion-ratio ($r = 0.85$, $p = 0.002$, $N = 10$, using the non-parametric Spearman's Rho). Labellers obviously were most sensitive to large or small f_0-excursions in the pitch rise; the higher the f_0-excursion ratio between contrastive and non-contrastive realisation of the prenuclear accent, the more labellers annotated the two rises of a sentence pair with a different accent type. It is striking that no such correlation existed with peak alignment differences, the genuine basis for intonational contrasts in AM-approaches since Bruce (1977). It would have been expected that the larger the difference in alignment between the two versions, the more labellers would label the two versions with a different accent type.

Accent types. There was a high degree of variation in accent types. When annotators used the same accent type for both the contrastive and non-contrastive version of a sentence pair, there was a pronounced preference for L+H* (22 ×) as opposed to L*+H (15 ×), H* (4 ×), and H*+ˆH (1 ×).

When the contrastive and non-contrastive versions were labelled with different accent types, many different accentual distintions were used and a generalisation does not seem possible. Focusing on sentence pairs 4, 7, and 9, whose versions supposedly express a meaningful contrast does not clarify the picture. One could argue that labellers have their own internal 'representation' of ToBI-categories used, which could still be categorically distinct from other accent types in their particular system but need not fully coincide with the representation of other labellers. If accent type differences were indeed employed on a subjective, idiosyncratic basis, individual labellers should maintain a reliable difference

sentence pair	Ling1	Ling2	Ling3	Ling4	Ling5	Ling6	Ling7	Ling8
1 (c)	L+H*	L+H*	L+H*	L*+H	L*+H	H*	**L+H***	**L*+H**
1 (nc)							**L*+H**	**L+H***
2 (c)	L+H*	L*+H	L+H*	L*+H	L*+H	**L+H*<**	L*+H	**L*+H**
2 (nc)						**L+H***		**L+H***
3 (c)	**L+H***	L+H*	L+H*	**L*+H**	**L+H***	**H***	L*+H	L*+H
3 (nc)	**L*+H**			**L+H***	**L*+H**	**L+H***		
4 (c)	L+H*	L+H*	L+H*	L*+H	**L+H***	**L+H***	**L*+H**	**L*+H**
4 (nc)					**L*+H**	**H*<**	**L+H***	**L+H***
5 (c)	L+H*	**L*+H**	L+H*	**L*+H**	**L*+H**	**L+H***	**L*+H**	**L*+H**
5 (nc)		**L***		**H***	**H***	**H***	**L+H***	**L+H***
6 (c)	**L*+H**	L*+H	L+H*	L*+H	**L*+H**	H*+^H	**L*+H**	L*+H
6 (nc)	**L+H***				**L+H***		**L+H***	
7 (c)	**L*+H**	**L+H***	L+H*	**L*+H**	**L+H***	**H*+^H**	**L*+H**	**L*+H**
7 (nc)	**L+H***	**L***		**H***	**H***	**H***	**H***	**H***
8 (c)	L+H*	**L*+H**	L+H*	**L*+H**	H*	L+H*	H*	L+H*
8 (nc)		**L+H***		**L+H***				
9 (c)	**L*+H**	L*+H	**L*+H**	L*+H	**H***	**L+H***	L*+H	**L*+H**
9 (nc)	**L+H***		**L+H***		**L+H***	**L+H*<**		**L+H***
10 (c)	L+H*	L+H*	**L+H***	H*	L+H*	L+H*	**L*+H**	**L*+H**
10 (nc)			**H***				**H***	**H***

Table 6: Labels of eight linguists for the ten sentence pairs. Accent differentiations for the contrastive (c) and non-contrastive (nc) version of a pair are highlighted with boldface. Sentence pairs in italics resulted in clear preference patterns for non-linguists.

between the contrastive and non-contrastive versions.

Analysing the results of the labelling experiments for the individual annotators separately shows a high degree of variation both within a labeller and between different labellers. The first four labellers used the same accent type for annotating the contrastive and non-contrastive version for most of the sentence pairs, while the remaining four linguists used different accent types more often. Linguist 1 and 3 appear to have a strong preference for L+H*, whereas labeller 4 uses L*+H very frequently. Linguists 7 and 8 appear to be very consistent in annotating the contrastive versions with L*+H and the non-contrastive version with an H*-accent (L+H* or H*).

Labellers used different accent types only for half of the sentence pairs. Out of the three sentence pairs that were well classified by non-linguists only two were labelled with different accent types by the majority of labellers. There was a large variation in accent types used. This suggests that these accent categories (if they indeed constitute different categories) are either not sufficiently well defined or are interpreted differently by different annotators.

There were three utterance pairs (pairs 5, 7, and 9) for which the majority of annotators labelled the two versions with different accent types. This points to some sort of categorical difference, which is, however, not fully consistent with the results from non-expert listeners. Labellers were most sensitive to a little versus large f_0-excursion in the thematic rise. Notably, AM-contrasts are mainly based on alignment differences. ToBI and GToBI only offer a limited way to annotate differences in pitch range: these are diacritics to indicate that the pitch range is expanded (upstep ˆH) or decreased (downstep !H); however, these diacritics only describe changes in pitch range relative to an earlier high target and can hence not be used for sentence-initial accents. The high degree of disagreement in accent types might be attributed to the fact that standard GToBI categories for intonation-phrase initial accents do not take the scaling of accents into account. That pitch excursion is important becomes also clear by analsysing the annotators' comments who often mentioned extreme f_0-excursions and steep rises. It has to be emphasised that the annotator's task was extremely difficult because they only heard isolated sentences and could not adjust to the peculiarities of the speakers. Further, there was no explicit training phase involved.

Intonational contrasts used in semantic formalisms, such as L*+H vs. L+H*, are in-

deed often employed by annotators to discriminate the realisations in contrastive and non-contrastive contexts. However, the annotation was not very consistent. Further, given the high proportion of identical labels for the contrastive and non-contrastive versions and the large number of different accent pairings otherwise, a simple 1:1 mapping from present-day AM accent types to thematic contrast seems premature.

DISCUSSION

The production experiment (experiment 1) showed that sentence-initial themes in contrastive contexts are prosodically distinguished from those in non-contrastive contexts, most importantly by peak height and alignment, f_0-excursion and duration of the prenuclear rise, as well as the duration of the stressed vowel and stressed syllable. The acoustic differences are especially significant, given that the readers were not aware of reading identical sentences in pragmatically different contexts. The results support some but not all of the intuitively ascribed differences between contrastive and non-contrastive themes. It was confirmed that contrastive themes are realised with a larger pitch excursion, but the rise did not generally start from a lower trough in contrastive contexts (as described by Jacobs, 1997). Contrary to general belief, hat patterns were not very common in the production data and they were not found more often in contrastive than in non-contrastive contexts. Hence their presence constitutes neither a sufficent nor necessary condition for contrast.

The relation between contrast and prosodic features (and between the prosodic features themselves) was shown to be rather complex. Speakers could use later or higher peaks (or both) to signal a thematic contrast. The presence of different strategies to achieve the same communicative function is reminiscent of trading relations found in segmental phonetics. Repp (1982), for instance, showed that different acoustic cues resulted in the same percept. For intonation, Gussenhoven discussed various biological codes and he suggested that 'peak delay can [...] be used as an enhancement of, or even a substitute for, pitch rising' (Gussenhoven, 2002, p. 52). The present data are support this view.

It was hypothesised that deviations from a norm (in terms of the scaling and positioning of tonal targets) may be used to express certain functions. The low tonal target at the start of the pitch rise was not affected by the contrastiveness of the precontext. This is another piece of evidence for the often reported stability of low tonal targets (see e.g. Caspers &

van Heuven, 1993; Prieto et al., 1995; Arvaniti et al., 1998; Ladd et al., 1999). However, the high tonal target at the end of the rise was used for expressing contrast: the peak was moved further to the right or realised higher in contrastive contexts than in non-contrastive contexts. This effect was observed across speakers from different backgrounds, different sentence-structures, and different segmental structures. This suggests that peak height and peak position can be used (and *are* used) to express functional differences.

In experiment 2 it was tested whether the differences found in contrastive and non-contrastive contexts do relate to different semantic functions. Listeners were presented a contrastive context and had to chose between a presumed contrastive and non-contrastive version (from the data collected in experiment 1). For seven (out of ten) sentences, listeners had no clear preferences. This could be attributed to the read speech samples which are not as varied as natural speech (more natural data elicitation, however, often results in the use of contrastive or additive particles, see Dimroth, 2002). Another explanation is that listeners based their judgement not only on the realisation of contrast but evaluated different, possibly paralinguistic aspects. Since prosody is used to convey a plurality of functions, it is hard to imagine that subjects concentrated on the linguistic channel only. For three (out of ten) sentences, however, there was a significant preference for one of the two versions. These three sentence pairs are particularly significant because there appears to be a clear link between intonational form and semantic function.

To investigate whether the acoustic differences expressed by the speakers are distinct enough to be used in linguistic descriptions, a labelling experiment was carried out (experiment 3). A group of eight linguists trained in using GToBI for prosodic annotation, labelled the thematic accents for a subset of the data. In half of the cases the contrastive and non-contrastive versions were annotated with the same accent type. In the remaining cases, many different intonational contrasts were used, which makes a simple semantic distinction based on different AM accent types questionable.

For three utterance pairs (which are only partly the same as those well classified by non-linguists), the majority of annotators labelled the two versions with different accent types, which might indicate a categorical distinction for at least those sentence pairs. To make use of the intonational information, however, the labels would have to be differently defined so that consistent annotation becomes possible. Further phonetic analysis reveiled a correlation between the number of *different* labels and the f_0-excursion ratio between

contrastive and non-contrastive context. Peak position differences did not result in more *different* labels for the two versions of a sentence pair. This is highly surprising, given that different accent types are defined on the basis of the temporal association between tonal targets and segmental structure. GToBI (and ToBI) do not even offer a possibility to annotate a particularly high or low f_0-excursion in intonation-phrase initial accents. Maybe this restriction should be reconsidered.

Overall pitch range is highly speaker-specific and the heights of individual accents have to interpreted within that range. In that respect the perception of intonational phenomena resembles vowel perception, which is also dependent on the physiology of the speaker. The vowel space can be voluntarily increased by very careful articulation, either locally to mark out certain words (Lindblom, 1990) or globally, e.g. in infant-directed speech (Kuhl et al., 1997). The vowel category stays the same (if interpreted within that speaker's vowel frame, not if interpreted in isolation, see Ladefoged & Broadbent, 1957). In analogy, thematic contrast marking could be interpreted as some sort of salience operation that emphasises the prenuclear accent with respect to the nuclear one by delaying and/or raising the peak. Interpretation of thematic accents would have to take the intonational context into account.

Different peak heights have been reported to correspond to perceived prominence (Rietveld & Gussenhoven, 1995; Terken, 1991; Kohler & Gartenberg, 1991, among others), emphasis, surprise/exceptionality, and politeness (Ladd & Morton, 1997; Chen & Gussenhoven, 2002), all of which are gradual in nature. Abstracting away from the dichotomy contrast vs. non-contrast, sentences with contrastive and non-contrastive themes are part of hierarchical discourse structuring. This discourse structure is marked by prosodic means (Brown & Yule, 1983; Swerts, 1994; Wichmann et al., 2000; Shriberg et al., 2000). Shriberg et al. (2000), for instance, reported that peak height and f_0-excursion were the most important cues to detect topic changes. Wichmann et al. (2000) investigated the realisation of sentence-initial accents in paragraph-initial and -medial position and found that paragraph-initial sentences had a later and higher prenuclear peak than paragraph-medial sentences. Psycholinguistically, the marking of topic-changes and discourse-initiality can be linked to what Chafe termed *activation* (Chafe, 1994). It is more difficult for the listener to activate a new concept or a concept different from the previous one (contrastive theme). This could be facilitated by a stronger theme marking, i.e. by higher peaks and/or later peaks (Gussenhoven, 2002, 2004). Activation costs can be defined in a more fine-grained way

and could be related to an increased intonational marking without requiring categoricality. From that perspective, the dichotomy contrast vs. non-contrast has to be seen as a special case, one that is operationally definable and of semantic interest, but one that is not independent of overall discourse intonation.

Speakers encode semantic distinctions by fine prosodic detail. This demands a high level of interaction between the organsisation of speech sounds and voice source, which we only begin to understand; also there are individual differences in production. Listeners are more or less able to decode this interaction between intonation and segmental structure and relate it to different functions. It is thus not just blatant intonational contrasts, such as the presence or absence of a hat pattern or different accent positions that are communicatively relevant. Intonational description could not adequately mirror the semantic contrasts by intonational contrasts. As certain fine-grained intonational distinctions are found to be linguistically meaningful, intonational descriptions ultimately will have to adapt.

ENDNOTES

1. In this article, we will not use the term *topic* because it is associated with too many different criteria and definitions (Hockett, 1958; Strawson, 1964; Gundel, 1974; Dahl, 1974; Reinhart, 1981; Lambrecht, 1994). Following Steedman (2000), the information structural terms *theme* and *rheme* will be used in this article.

2. In English, a contrastive reading of the sentence 'I was at home' is achieved by accenting both 'I' and 'home'.

3. The findings for Dutch prenuclear rises are somewhat less clear-cut. Ladd et al. (2000) report different alignment patterns for the end of the rise depending on the syllable structure of the accented syllable. If there is a phonologically long vowel, the peak is found at the end of the vowel, but when there is a short vowel, the peak is found within the following onset consonant.

4. Problems with monotonous speakers have been reported in various production experiments e.g. Ladd et al. (1999), p. 1548, Ladd & Schepman (2003), p. 86, and Atterer & Ladd (2004), p. 189.

5. The Bonferroni correction only adjusts the significance level based on tests for independent measures, such as the 3 f_0-values for the different landmarks, all temporal variables (6), and the absolute alignment variables (5). The magnitude of the f_0-rise and the f_0-fall are calculated directly from the static f_0-measures. Similarly, the proportional alignment-measures and the slope are only derived variables. The significance level has to be adjusted for fourteen variables, resulting in $p = \frac{0.05}{14} = 0.004$.

6. Error values represent the standard error of the mean.

7. Phonetically, the overwhelming use of low rhematic accents might be responsible for the as yet unexplained duration distribution in the utterances found, namely that contrastive *thematic* elements were lengthened at the expense of contrastive *rhematic* elements.

8. We are well aware of the criticism of unnaturalness and the uncontrollable effects of visual presentation, but we believe that parallel constructions are syntactically and semantically sufficiently marked to constrain the degrees of interpretational and intonational freedom. In not presenting the first part of the parallel construction auditorily, we avoid the as yet rather unexplored territory of inter-clause peak height relations.

9. All χ^2-analyses for experiment 2 are based on an expected frequency of six (half of the subjects). Because of sporadic choices for 'no difference', this is a conservative statistic.

To compensate for the small sample, Yate's correction for discontinuity was applied.

References

Andreeva, B. & Barry, W. J. (1999), Intonation von *checks* in der Sofia Varietät des Bulgarischen. *Research Report Phonetics, University of the Saarland (PHONUS)*, **4**, 1–14.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998), Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, **26** (1), 3–25.

Atterer, M. & Ladd, D. R. (2004), On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics*, **32** (3), 177–197.

Beckman, M. & Ayers, G. M. (1997), Guidelines for ToBI Labelling. Tech. rep., Ohio State University.

Bolinger, D. L. (1989), *Intonation and its uses. Melody in Grammar and Discourse*. Stanford University Press, Stanford, California.

Braun, B. (2004), Answers to the perception of thematic contrast and questions concerning the perception of thematic "non-contrast". In *Proc. of the 2nd International Conference on Speech Prosody*, (pp. 685–688), Nara, Japan.

Brazil, D. (1985), *The Communicative Value of Intonation in English*. Discourse Analysis Monograph, no 2, Bleak House Books and English Language Research, Birmingham, UK.

Brown, G. & Yule, G. (1983), *Discourse Analysis*. Cambridge University Press, Cambridge, UK.

Bruce, G. (1977), *Swedish Word Accents in Sentence Perspective*. Travaux de l'Institut de linguistique de Lund 12, Gleerup, Lund.

Büring, D. (1997a), The great scope inversion conspiracy. *Linguistics and Philosophy*, **20**, 175–194.

Büring, D. (1997b), *The Meaning of Topic and Focus – the 59th Street Bridge Accent*. Routhledge, London.

Caspers, J. & van Heuven, V. J. (1993), Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, **50**, 161–171.

Chafe, W. L. (1994), *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. The University of Chicago Press, Chicago.

Chen, A. (2003), Effects of prosody on language comprehension: uniform or variable?, Talk given at the 6th NWCL International Conference on Prosody and Pragmatics, Preston, UK.

Chen, A. & Gussenhoven, C. (2002), Language-specific uses of the effort code. In *Proc. of the 1st International Conference on Speech Prosody*, (pp. 215–218).

Cohen, A. & t'Hart, J. (1967), On the anatomy of intonation. *Lingua*, **19**, 177–192.

Couper-Kuhlen, E. (1986), *An Introduction to English Prosody*. Edward Arnold, London.

Cruttenden, A. (1986), *Intonation*. Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge.

Crystal, D. (1969), *Prosodic Systems and Intonation in English*, vol. 1 of *Cambridge Studies in Linguistics*. Cambridge University Press, Cambridge.

Dahl, O. (1974), Topic-comment structure revisited. In *Topic and Comment, Contextual Boundedness and Focus*, Papers in Text Linguistics 6, (pp. 1–24), Buske, Hamburg.

Dimroth, C. (2002), Topics, assertions, and additive words: how L2 learners geht from information structure to target-language syntax. *Linguistics*, **40** (4), 891–923.

Féry, C. (1993), *German Intonational Patterns*, vol. 285 of *Linguistische Arbeiten*. Niemeyer, Tbingen.

Gilles, P. & Peters, J., eds. (2004), *Regional Variation in Intonation*. Linguistische Arbeiten, Niemeyer, Tübingen.

Grabe, E. (2004), Intonational variation in urban dialects of English spoken in the British Isles. In P. Gilles & J. Peters, eds., *Regional Variation in Intonation*, Linguistische Arbeiten, (pp. 9–31), Niemeyer, Tübingen.

Grice, M., Baumann, S., & Benzmüller, R. (2005), German intonation in autosegmental-metrical phonology. In S.-A. Jun, ed., *Prosodic Typology*, (pp. 55–83), Oxford University Press, Oxford.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996), Consistency in transcription and labelling of German intonation with GToBI. In *Proc. of the 4th International Conference on Spoken Language Processin (ICSLP)*, (pp. 1716–1719), Philadelphia, USA.

Gundel, J. K. (1974), *The Role of Topic and Comment in Linguistic Theory*. Ph.D. thesis, University of Texas, Austin.

Gussenhoven, C. (1984), *On the grammar and Semantics of Sentence Accents*. Foris Publications, Dordrecht.

Gussenhoven, C. (2002), Intonation and interpretation: Phonetics and phonology. In *Proc. of the 1st International Conference on Speech Prosody*, (pp. 47–58), Aix-en-Provence, France.

Gussenhoven, C. (2004), *The Phonology of Tone and Intonation*. Research Surveys in Linguistics, Cambridge University Press, Cambridge.

Hockett, C. F. (1958), *A Course in Modern Linguistics*. Macmillan, New York.

Jackendoff, R. (1972), *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, Massachusetts.

Jacobs, J. (1996), Bemerkungen zur I-Topikalisierung. *Sprache und Pragmatik*, **41**, 1–48.

Jacobs, J. (1997), I-Topikalisierung. *Linguistische Berichte*, **108**, 91–133.

Kohler, K. J. (1991a), A model of German intonation. *Arbeitsberichte des Instituts fr Phonetik der Universitt Kiel (AIPUK)*, **25**, 295–360.

Kohler, K. J. (1991b), Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, **25**, 115–185.

Kohler, K. J. & Gartenberg, R. (1991), The perception of accents: F0 peak height versus F0 peak position. *Arbeitsberichte des Instituts fr Phonetik der Universitt Kiel (AIPUK)*, **25**, 219–242.

König, W. (1994), *dtv-Atlas Deutsche Sprache*. Deutscher Taschenbuch Verlag, revised edition.

Krahmer, E. & Swerts, M. (1999), Contrastive accent in dialogue: Towards a presuppositional analysis. In *Proc. ESSLLI Workshop on Focus and Presupposition in Multi-Layer Dialogue*, Utrecht, Netherlands.

Krahmer, E. & 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** (1), 75–112.

Krifka, M. (1999), Additive particles under stress. In *Proceedings from Semantic and Linguistic Theory SALT 8*, CLC Publications, Cornell.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Rskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997), Cross-language analysis of phonetic units in language addressed to infants. *Science*, **277**, 684–686.

Ladd, D. R. (1980), *The Structure of Intonational Meaning*. Indiana University Press, Bloomington.

Ladd, D. R. (1983), Phonological features of intonational meaning. *Language*, **59**, 721–759.

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999), Constant "segmental anchoring" of f_0 movements under changes in speech rate. *Journal of the Acoustic Society of America*, **106** (3), 1543–1554.

Ladd, D. R., Mennen, I., & Schepman, A. (2000), Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustic Society of America*, **107** (5), 2685–2696.

Ladd, D. R. & Morton, R. (1997), The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics*, **25**, 313–342.

Ladd, D. R. & Schepman, A. (2003), "Sagging transitions" between high pitch accents in English; experimental evidence. *Journal of Phonetics*, **31**, 81–112.

Ladefoged, P. & Broadbent, D. E. (1957), Information conveyed by vowels. *Journal of the Acoustic Society of America*, **24**, 98–104.

Lambrecht, K. (1994), *Information Structure and Sentence Form. Topic, Focus and the Mental Representations of Discourse Referents*. Cambridge Studies in Linguistics, Cambridge University Press, Cambridge.

Lee, C. (1999), Contrastive topic: A locus of the interface. Evidence from Korean and English. In K. Turner, ed., *The Semantics/Pragmatics Interface from Different Points of View, Vol. 1*, (pp. 317–342), Elsevier Science Ltd, Oxford, UK.

Liberman, M. & Pierrehumbert, J. B. (1984), Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle, eds., *Language Sound Structure*, (pp. 157–233), MIT Press.

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

Lindblom, B. (1990), Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle & A. Marchal, eds., *Speech Production and Speech Modelling*, (pp. 403–439), Kluwer Academic Publisher, Dordrecht.

Maeda, J. (1976), *A Characterization of American English*. Ph.D. thesis, MIT, Cambridge, Massachusetts.

Mehlhorn, G. (2001), Produktion und Perzeption von Hutkonturen im Deutschen. *Linguistische Arbeitsberichte*, **77**, 31–57.

Nolan, F. (1995), The effect of emphasis on declination in English intonation. In J. W. Lewis, ed., *Studies in General and English Phonetics*, chap. 17, (pp. 241–254), Routledge, London.

O'Connor, J. D. & Arnold, G. F. (1961), *Intonation of Colloquial English*. Longman, London.

Peters, J. (2004), Regionale Variation der Intonation des Deutschen. Studien zu ausgewählten Regionalsprachen. Habilitationsschrift, Universität Potsdam/University of Nijmegen.

Peterson, G. E. & Lehiste, I. (1960), Duration of syllable nuclei in English. *Journal of the Acoustic Society of America*, **32**, 693–703.

Pierrehumbert, J. B. & Hirschberg, J. (1990), The meaning of intonation in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack, eds., *Intentions in Communication*, chap. 14, (pp. 271–311), MIT Press, Cambridge, Massachusetts.

Pitrelli, J., Beckman, M., & Hirschberg, J. (1994), Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proc. of the 3rd International Conference on Spoken Language Processing (ICSLP)*, vol. 2, (pp. 123–126), Yokohama, Japan.

Prevost, S. (1995), *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. thesis, University of Pennsylvania.

Prieto, P., van Santen, J., & Hirschberg, J. (1995), Tonal alignment patterns in Spanish. *Journal of Phonetics*, **23**, 429–451.

Reinhart, T. (1981), Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, **27** (1), 53–94.

Repp, B. H. (1982), Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, **92**, 81–110.

Rietveld, A. C. M. & Gussenhoven, C. (1985), On the relation between pitch excursion size and prominence. *Journal of Phonetics*, **13**, 299–308.

Rietveld, T. & Gussenhoven, C. (1995), Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics*, **23**, 375–385.

Schepman, A., Lickley, R., & Ladd, D. R. (submitted), Effects of vowel length and "right context" on the alignment of Dutch nuclear accents.

Shriberg, E., Stolcke, A., Hakkani-Tr, D., & Tr, G. (2000), Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, **32**, 127–154.

Silverman, K. E. A., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. B., & Hirschberg, J. (1992), ToBI: A standard for labeling English prosody. In *Proceedings of the 2nd International Conference on Spoken Language Processing*, vol. 2, (pp. 867–870), Banff, Canada.

Silverman, K. E. A. & Pierrehumbert, J. B. (1990), The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman, eds., *Papers in Laboratory Phonology I*, (pp. 72–106), Cambridge University Press, Cambridge, UK.

Sluijter, A. M. C. & Terken, J. M. B. (1993), Beyond sentence prosody: Paragraph intonaton in Dutch. *Phonetica*, **50**, 180–188.

Steedman, M. (2000), Information structure and the syntax-phonology interface. *Linguistic Inquiry*, **31** (4), 649–689.

Steele, S. A. (1986), Nuclear accent F_0 peak location: effects of rate, vowel, and number of following syllables. *Journal of the Acoustic Society of America*, **80**.

Steube, A. (2001), Grammatik und Pragmatik von Hutkonturen. *Linguistische Arbeitsberichte*, **77**, 7–29.

Strawson, P. F. (1964), Identifying reference and truth-values. In D. Steinberg & L. Jacobovits, eds., *Semantics*, (pp. 86–99), Cambridge University Press.

Swerts, M. (1994), *Prosodic Features of Discourse Units*. Ph.D. thesis, Eindhoven University of Technology.

Talkin, D. (1995), A robust algorithm for pitch tracking (rapt). In W. B. Klein & K. K. Palival, eds., *Speech Coding and Synthesis*, Elsevier.

Terken, J. (1991), Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustic Society of America*, **89** (4), 1768–1776.

Theune, M. (1999), Parallelism, coherence, and contrastive accent. In *Proc. of the 6th European Conference on Speech Communication and Technology*, vol. 1, (pp. 555–558), Budapest, Hungary.

Umbach, C. (2001), Relating contrast and contrastive topic. In I. Kruijff-Korbayova, ed., *Proc. ESSLLI Workshop on Information Structure, Discourse Structure, and Discouse Semantics*, Helsinki, Finland.

von Essen, O. (1964), *Grundzüge der hochdeutschen Satzintonation*. Henn, Ratingen, 2nd edn.

Weber, A., Braun, B., & Crocker, M. W. (2004), Adjectival modifiers and reference resolution: When prosodic focus matters. In *Poster presented at the 17th Annual CUNY Sentence Processing Conference*, CUNY, Boston, USA.

Weber, A. & Müller, K. (2004), Word order variation in German main clauses: A corpus analysis. In *Proc. of the 20th International Conference on Computational Linguistics*, Geneva, Switzerland.

Wichmann, A., House, J., & Rietveld, T. (2000), Discourse constraints on F_0 peak timing in English. In A. Botinis, ed., *Intonation: Analysis, Modelling and Technology*, Text, Speech, and Language Technology, vol. 15, chap. 8, (pp. 163–182), Kluwer, Dordrecht/Boston/London.

Wunderlich, D. (1991), Intonation and contrast. *Journal of Semantics*, **8**, 239–251.

Xu, Y. (2002), Articulatory constraints and tonal alignment. In *Proc. of the 1st International Conference on Speech Prosody*, (pp. 91–100), Aix-en-Provence, France.

APPENDIX

non-contrastive context	contrastive context
Many Europeans don't know much about Malaysia. The country consists of two islands. To ease the communications between the two parts, almost every household has a computer with Internet access. However, Malaysia is not a highly technological country. **The Malaysians live from agriculture.** They are neither especially poor nor rich.	Malaysia and Indonesia are neighbouring countries in the South China Sea. Despite their geographical adjacency, their living and working conditions differ tremendously. In Indonesia, tourism is very important and many people work in this sector. **The Malaysians live from agriculture.** They have mainly focused on the cultivation of rice.

Table 7: English translation of a sample non-contrastive and contrastive paragraph. The target sentences are highlighted in boldface.

Speaker	Origin	Region	Sex
BK	Lower Saxony	North	female
CS	Lower Saxony	North	female
UB	Lower Saxony	North	female
(SZ)	North Rhine-Westfalia	North	female
IB	Baden-Wuerttemberg	South	female
(CH)	Bavaria	South	female
PK	Bavaria	South	female
JL	Palatinate	South	male
CZ	Hesse	North	male
MB	Berlin	North	male
DS	North Rhine-Westfalia	North	male
(SV)	Schleswig-Holstein	North	male

Table 8: Information about the speakers. Speakers in brackets were excluded from analysis. The classification into northern and southern German is based on König (1994) and is the same classification as used by Atterer & Ladd (2004). Northern German speakers originate all from north of the bundle of isoglosses dividing broadly Northern dialects from Central and Southern ones.

Target sentence	long vowel	gramm. relation	pos. of stressed syllable
Ital**ie**ner sind sehr **gast**freundlich.	1	NP	3
'*Italians are very hospitable*'			penultimate
Die Ka**na**ren sind ein **Wan**derparadies.	1	NP	3
'*The Canaries are a paradise for hiking*'			penultimate
Die Ma**lay**en leben von der **Land**wirtschaft.	1	NP	3
'*The Malaysians live from agriculture*'			penultimate
Die **Rö**mer waren sehr organi**siert**.	1	NP	2
'*The Romans were very organised*'			penultimate
Die **Ma**ler arbeiten viel im **Frei**en.	1	NP	2
'*Painters often work outside*'			penultimate
Mar**le**ne spielt Kla**vier** und kann **sin**gen.	1	NP	2
'*Marlene plays the piano and can sing*'			penultimate
Die **Lä**mmer haben **Angst** vor Menschen.	0	NP	2
'*Lambs fear humans*'			penultimate
In Mi**la**no kann man gut **ein**kaufen.	1	PP	3
'In Milano, shopping is great'			penultimate
In Ar**me**nien schreibt man la**tei**nisch.	1	PP	3
'*In Armenia, the Latin alphabet is used*'			penultimate
In A**me**rika besitzen **vie**le eine **Waf**fe.	1	PP	3
'*In America, many people own a weapon*'			antepenultimate
In **Bay**ern beginnen die **Fe**rien Anfang **Ju**li.	1	PP	2
'*In Bavaria, holidays start early in July*'			penultimate
Im **Ja**nuar ist es **fros**tig.	0	PP	2
'*In January, it is frosty*'			antepenultimate

Table 9: Target sentences with information about the phonological vowel length of the thematic stress, the grammatical relation of the theme (subject-NP or PP), and the position of the stressed syllable in the theme. Stressed syllables are marked in bold face.

Sentence number	Speaker	Sentence	Rheme accent
1	BK	In Armenien schreibt man lateinisch.	H+L*
2	BK	In Milano kann man gut einkaufen.	H+L*
3	CH	Die Lämmer haben Angst vor Menschen.	H+L*
4	CS	In Bayern beginnen die Ferien Ende Juli.	H+L*
5	CS	In Milano kann man gut einkaufen.	H*
6	DS	Im Januar ist es frostig.	H+L*
7	DS	Die Römer waren sehr organisiert.	H*
8	IB	Italiener sind sehr gastfreundlich.	H+L*
9	JL	Die Lämmer haben Angst vor Menschen.	H+L*
10	UB	Italiener sind sehr gastfreundlich.	H*

Table 10: Stimulus sentences used in the labelling experiment (experiment 2). Translations are provided in Table 9.

Start of sentence	Target sentence
Fohlen haben sich gut an Menschen gewöhnt, aber '*Foils have become used to men*'	Die Lämmer haben Angst vor Menschen. (2)
Die Germanen waren ein wilder Haufen, aber '*The Germanic tribes were wild and uncivilised*'	Die Römer waren sehr organisiert.
Rom ist kulturell interessant, und '*Rome is culturally interesting*'	In Milano kann man gut einkaufen. (2)
Dänen sind lieber für sich allein, aber '*Danes prefer to be on their own*'	Italiener sind sehr gastfreundlich. (2)
Die Georgier haben eine eigene Schrift, und '*Georgians have their own writing system*'	In Armenien schreibt man lateinisch.
Die Saarländer starten früh in die Sommerferien, aber '*People from the Saarland have early summer holidays*'	In Bayern beginnen die Ferien Ende Juli.
Der Dezember ist oft vergleichsweise mild, aber '*In December it is often very mild*'	Im Januar ist es frostig.

Table 11: List of semantic parallelism used in experiment 3. Left column shows the visually presented sentence starts (and English translation), right column the respective target sentence (for translations see Table 9. Number in brackets shows that sentence was used twice for experiment 3.